(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115577798 A
(43)申请公布日 2023.01.06

(21)申请号 202211284555.4

(22)申请日 2022.10.17

(71)申请人 中国人民解放军国防科技大学
地址 410003 湖南省长沙市开福区德雅路109号

(72)发明人 罗来龙 胡煜晗 郭得科 唐国明 赵亚威 任棒棒 张千桢

(74)专利代理机构 北京风雅颂专利代理有限公司 11403
专利代理师 曾志鹏

(51)Int.Cl.
G06N 20/00(2019.01)
G06F 18/214(2023.01)

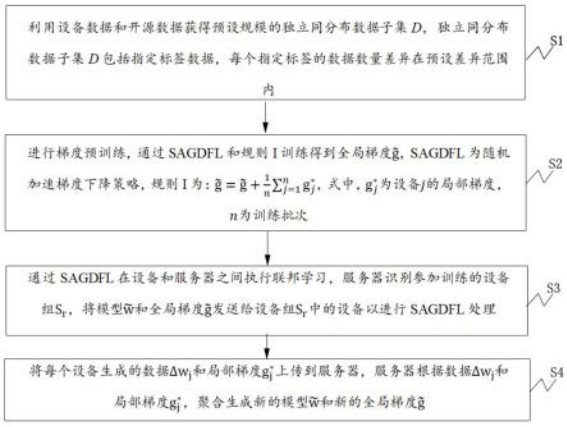
权利要求书2页 说明书13页 附图2页

(54)发明名称

基于随机加速梯度下降的半联邦学习方法及装置

(57)摘要

基于随机加速梯度下降的半联邦学习方法及装置，该方法利用设备数据和开源数据获得预设规模的独立同分布数据子集；进行梯度预训练，通过SAGDFL和规则I训练得到全局梯度；通过SAGDFL在设备和服务器之间执行联邦学习，服务器识别参加训练的设备组，将模型和全局梯度发送给设备组中的设备进行SAGDFL处理；将每个设备生成的数据和局部梯度上传到服务器，聚合生成新的模型和新的全局梯度。本申请通过随机加速梯度下降策略减少局部和全局梯度方差；可以大大缩短训练迭代，保证模型推理精度；确保隐私泄露风险可控；使额外的通信开销最小化，使模型参数变得稀疏，易于将传输的数据进行压缩，降低通信成本。



CN 115577798 A

1.基于随机加速梯度下降的半联邦学习方法,包括:

利用设备数据和开源数据获得预设规模的独立同分布数据子集D,独立同分布数据子集D包括指定标签数据,每个指定标签的数据数量差异在预设差异范围内;

进行梯度预训练,通过SAGDFL和规则I训练得到全局梯度$\tilde{\mathbf{g}}$,SAGDFL为随机加速梯度下降策略,规则I为:$\tilde{\mathbf{g}}=\tilde{\mathbf{g}}+\frac{1}{n}\sum_{j=1}^{n}\mathbf{g}_j^*$,式中,$\mathbf{g}_j^*$为设备j的局部梯度,n为训练批次;

通过SAGDFL在设备和服务器之间执行联邦学习,服务器识别参加训练的设备组S_r,将模型$\tilde{\mathbf{w}}$和全局梯度$\tilde{\mathbf{g}}$发送给设备组S_r中的设备以进行SAGDFL处理;

将每个设备生成的数据Δw_j和局部梯度$\mathbf{g}_j^*$上传到服务器,服务器根据数据Δw_j和局部梯度$\mathbf{g}_j^*$,聚合生成新的模型$\tilde{\mathbf{w}}$和新的全局梯度$\tilde{\mathbf{g}}$。

2.根据权利要求1所述的基于随机加速梯度下降的半联邦学习方法,其中,将利用独立同分布数据子集D计算的初始值$\tilde{\mathbf{g}}$作为全局梯度的无偏估计;独立同分布数据子集D来源于设备的本地数据和互联网的开源数据;

对采集的设备的本地数据通过隐私测量仪进行敏感性评估,隐私测量仪采用ML privacy meter;

对采集的互联网的开源数据,通过翻转、平移或旋转的方式进行数据样本扩充。

3.根据权利要求1所述的基于随机加速梯度下降的半联邦学习方法,其中,对需要上传的数据Δw_j进行稀疏化处理,稀疏化处理过程保留关键参数的值,不相关参数的值置零;

通过配置阈值区分局部模型中的关键参数和无关参数。

4.根据权利要求3所述的基于随机加速梯度下降的半联邦学习方法,其中,通过联合优化使上传的数据Δw_j稀疏化,联合优化过程在目标函数中加入一阶惩罚函数;

获得稀疏化处理的数据后,对数据进行无损压缩。

5.根据权利要求1所述的基于随机加速梯度下降的半联邦学习方法,其中,联邦学习过程,采用规则II更新全局梯度$\tilde{\mathbf{g}}$,规则II为:

$$\tilde{\mathbf{g}}=\frac{1}{n}\sum_{j=1}^{n}\mathbf{g}_j^*$$

式中,$\mathbf{g}_j^*$为设备j的局部梯度,n为训练批次。

6.基于随机加速梯度下降的半联邦学习装置,其中,包括:

数据子集构建模块,用于利用设备数据和开源数据获得预设规模的独立同分布数据子集D,独立同分布数据子集D包括指定标签数据,每个指定标签的数据数量差异在预设差异范围内;

梯度预训练模块,用于进行梯度预训练,通过SAGDFL和规则I训练得到全局梯度$\tilde{\mathbf{g}}$,SAGDFL为随机加速梯度下降策略,规则I为:$\tilde{\mathbf{g}}=\tilde{\mathbf{g}}+\frac{1}{n}\sum_{j=1}^{n}\mathbf{g}_j^*$,式中,$\mathbf{g}_j^*$为设备j的局部梯度,n为训练批次;

联邦学习模块，用于通过SAGDFL在设备和服务器之间执行联邦学习，服务器识别参加训练的设备组S_r，将模型$\tilde{w}$和全局梯度$\tilde{g}$发送给设备组S_r中的设备以进行SAGDFL处理；

训练结果处理模块，用于将每个设备生成的数据Δw_j和局部梯度g_j^*上传到服务器，服务器根据数据Δw_j和局部梯度g_j^*，聚合生成新的模型$\tilde{w}$和新的全局梯度$\tilde{g}$。

7.根据权利要求6所述的基于随机加速梯度下降的半联邦学习装置，其中，数据子集构建模块中，将利用独立同分布数据子集D计算的初始值$\tilde{g}$作为全局梯度的无偏估计；独立同分布数据子集D来源于设备的本地数据和互联网的开源数据；

数据子集构建模块包括：

敏感性评估子模块，用于对采集的设备的本地数据通过隐私测量仪进行敏感性评估，隐私测量仪采用ML privacy meter；

样本扩充子模块，用于对采集的互联网的开源数据，通过翻转、平移或旋转的方式进行数据样本扩充。

8.根据权利要求6所述的基于随机加速梯度下降的半联邦学习装置，其中，还包括稀疏化处理模块，用于对需要上传的数据Δw_j进行稀疏化处理，稀疏化处理过程保留关键参数的值，不相关参数的值置零；

稀疏化处理模块中，通过配置阈值区分局部模型中的关键参数和无关参数。

9.根据权利要求8所述的基于随机加速梯度下降的半联邦学习装置，其中，还包括联合优化模块，用于通过联合优化使上传的数据Δw_j稀疏化，联合优化过程在目标函数中加入一阶惩罚函数；

还包括压缩处理模块，用于获得稀疏化处理的数据后，对数据进行无损压缩。

10.根据权利要求6所述的基于随机加速梯度下降的半联邦学习装置，其中，联邦学习模块中，采用规则II更新全局梯度$\tilde{g}$，规则II为：

$$\tilde{g} = \frac{1}{n}\sum_{j=1}^{n} g_j^*$$

式中，g_j^*为设备j的局部梯度，n为训练批次。

基于随机加速梯度下降的半联邦学习方法及装置

技术领域

[0001] 本申请涉及人工智能数据处理技术领域，尤其涉及一种基于随机加速梯度下降的半联邦学习方法及装。

背景技术

[0002] 随着物联网设备和网络终端设备中数据的爆炸式增长，跨数据集分析处理的需求正在不断增加。然而，随着隐私保护法案的不断出台，数据世界又面临了新的难题“数据孤岛”。法律的约束和社会对数据安全的关注使得传统的基于数据收集的集中式或分布式的训练不再可行。在数据无法共享的情况下，数据只能在本地进行使用，在这种情况下通过聚合从每个设备本地计算的参数或梯度来训练全局模型的联邦学习(FL)应运而生。

[0003] 在联邦学习中，由于数据被保留在本地而不进行移动，因此可以对本地数据起到天然的保护作用。同时，这种新的学习方式也能帮助打破了“数据孤岛”困境。然而，由于缺乏了数据中心等平台收集数据进行统一的处理，不同设备之间的本地数据将很难保证独立同分布条件，导致传统机器学习在联邦学习训练中所面临的梯度方差问题比在集中式和分布式场景下都更加严重。

[0004] 联邦学习往往需要进行更多次的模型训练才能达到较好的收敛。然而更多的迭代，也意味着更高的通信成本，以及更高的隐私泄露和恶意攻击的风险，这将使模型的实际训练和部署更加困难。其中的根本原因是通常假设每个设备上的数据是独立同分布的(IID)，但是，在大多数情况下，设备上的数据是非独立同分布(Non-IID)的，而这也将使得上述问题更加严重。

发明内容

[0005] 有鉴于此，本申请的目的在于提出一种基于随机加速梯度下降的半联邦学习方法及装置，用以解决或部分解决上述技术问题。

[0006] 基于上述目的，本申请的第一方面提供了一种基于随机加速梯度下降的半联邦学习方法，包括：

[0007] 利用设备数据和开源数据获得预设规模的独立同分布数据子集D，独立同分布数据子集D包括指定标签数据，每个指定标签的数据数量差异在预设差异范围内；

[0008] 进行梯度预训练，通过SAGDFL和规则I训练得到全局梯度$\tilde{\mathbf{g}}$，SAGDFL为随机加速梯度下降策略，规则I为：$\tilde{\mathbf{g}} = \tilde{\mathbf{g}} + \frac{1}{n}\sum_{j=1}^{n} \mathbf{g}_j^*$，式中，$\mathbf{g}_j^*$为设备j的局部梯度，n为训练批次；

[0009] 通过SAGDFL在设备和服务器之间执行联邦学习，服务器识别参加训练的设备组S_r，将模型$\tilde{\mathbf{w}}$和全局梯度$\tilde{\mathbf{g}}$发送给设备组S_r中的设备以进行SAGDFL处理；

[0010] 将每个设备生成的数据Δw_j和局部梯度$\mathbf{g}_j^*$上传到服务器，服务器根据数据Δw_j和局部梯度$\mathbf{g}_j^*$，聚合生成新的模型$\tilde{\mathbf{w}}$和新的全局梯度$\tilde{\mathbf{g}}$。

[0011] 作为基于随机加速梯度下降的半联邦学习方法优选方案，将利用独立同分布数据子集D计算的初始值$\tilde{\mathbf{g}}$作为全局梯度的无偏估计；独立同分布数据子集D来源于设备的本地数据和互联网的开源数据；

[0012] 对采集的设备的本地数据通过隐私测量仪进行敏感性评估，隐私测量仪采用ML privacy meter；

[0013] 对采集的互联网的开源数据，通过翻转、平移或旋转的方式进行数据样本扩充。

[0014] 作为基于随机加速梯度下降的半联邦学习方法优选方案，对需要上传的数据Δw_j进行稀疏化处理，稀疏化处理过程保留关键参数的值，不相关参数的值置零；

[0015] 通过配置阈值区分局部模型中的关键参数和无关参数。

[0016] 作为基于随机加速梯度下降的半联邦学习方法优选方案，通过联合优化使上传的数据Δw_j稀疏化，联合优化过程在目标函数中加入一阶惩罚函数；

[0017] 获得稀疏化处理的数据后，对数据进行无损压缩。

[0018] 作为基于随机加速梯度下降的半联邦学习方法优选方案，联邦学习过程，采用规则II更新全局梯度$\tilde{\mathbf{g}}$，规则II为：

[0019] $$\tilde{\mathbf{g}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{g}_j^*$$

[0020] 式中，$\mathbf{g}_j^*$为设备j的局部梯度，n为训练批次。

[0021] 本申请的第二方面提供了一种基于随机加速梯度下降的半联邦学习装置，包括：

[0022] 数据子集构建模块，用于利用设备数据和开源数据获得预设规模的独立同分布数据子集D，独立同分布数据子集D包括指定标签数据，每个指定标签的数据数量差异在预设差异范围内；

[0023] 梯度预训练模块，用于进行梯度预训练，通过SAGDFL和规则I训练得到全局梯度$\tilde{\mathbf{g}}$，SAGDFL为随机加速梯度下降策略，规则I为：$\tilde{\mathbf{g}} = \tilde{\mathbf{g}} + \frac{1}{n}\sum_{j=1}^{n} \mathbf{g}_j^*$，式中，$\mathbf{g}_j^*$为设备j的局部梯度，n为训练批次；

[0024] 联邦学习模块，用于通过SAGDFL在设备和服务器之间执行联邦学习，服务器识别参加训练的设备组S_r，将模型$\tilde{\mathbf{w}}$和全局梯度$\tilde{\mathbf{g}}$发送给设备组S_r中的设备以进行SAGDFL处理；

[0025] 训练结果处理模块，用于将每个设备生成的数据Δw_j和局部梯度$\mathbf{g}_j^*$上传到服务器，服务器根据数据Δw_j和局部梯度$\mathbf{g}_j^*$，聚合生成新的模型$\tilde{\mathbf{w}}$和新的全局梯度$\tilde{\mathbf{g}}$。

[0026] 作为基于随机加速梯度下降的半联邦学习装置优选方案，数据子集构建模块中，将利用独立同分布数据子集D计算的初始值$\tilde{\mathbf{g}}$作为全局梯度的无偏估计；独立同分布数据子集D来源于设备的本地数据和互联网的开源数据；

[0027] 数据子集构建模块包括：

[0028] 敏感性评估子模块，用于对采集的设备的本地数据通过隐私测量仪进行敏感性评估，隐私测量仪采用ML privacy meter；

[0029] 样本扩充子模块，用于对采集的互联网的开源数据，通过翻转、平移或旋转的方式

进行数据样本扩充。

[0030] 作为基于随机加速梯度下降的半联邦学习装置优选方案，还包括稀疏化处理模块，用于对需要上传的数据Δw_j进行稀疏化处理，稀疏化处理过程保留关键参数的值，不相关参数的值置零；

[0031] 稀疏化处理模块中，通过配置阈值区分局部模型中的关键参数和无关参数。

[0032] 作为基于随机加速梯度下降的半联邦学习装置优选方案，还包括联合优化模块，用于通过联合优化使上传的数据Δw_j稀疏化，联合优化过程在目标函数中加入一阶惩罚函数；

[0033] 还包括压缩处理模块，用于获得稀疏化处理的数据后，对数据进行无损压缩。

[0034] 作为基于随机加速梯度下降的半联邦学习装置优选方案，联邦学习模块中，采用规则II更新全局梯度$\tilde{\mathbf{g}}$，规则II为：

[0035] $$\tilde{\mathbf{g}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{g}_j^*$$

[0036] 式中，$\mathbf{g}_j^*$为设备j的局部梯度，n为训练批次。

[0037] 本申请的第三方面提出了一种电子设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现第一方面所述的基于随机加速梯度下降的半联邦学习方法。

[0038] 本申请的第四方面提出了一种非暂态计算机可读存储介质，所述非暂态计算机可读存储介质存储计算机指令，所述计算机指令用于使计算机执行实现第一方面所述的基于随机加速梯度下降的半联邦学习方法。

[0039] 从上面所述可以看出，本申请提供的技术方案，利用设备数据和开源数据获得预设规模的独立同分布数据子集D，独立同分布数据子集D包括指定标签数据，每个指定标签的数据数量差异在预设差异范围内；进行梯度预训练，通过SAGDFL和规则I训练得到全局梯度$\tilde{\mathbf{g}}$，SAGDFL为随机加速梯度下降策略，规则I为：$\tilde{\mathbf{g}} = \tilde{\mathbf{g}} + \frac{1}{n}\sum_{j=1}^{n} \mathbf{g}_j^*$，式中，$\mathbf{g}_j^*$为设备j的局部梯度，n为训练批次；通过SAGDFL在设备和服务器之间执行联邦学习，服务器识别参加训练的设备组S_r，将模型$\tilde{\mathbf{w}}$和全局梯度$\tilde{\mathbf{g}}$发送给设备组S_r中的设备以进行SAGDFL处理；将每个设备生成的数据Δw_j和局部梯度$\mathbf{g}_j^*$上传到服务器，服务器根据数据Δw_j和局部梯度$\mathbf{g}_j^*$，聚合生成新的模型$\tilde{\mathbf{w}}$和新的全局梯度$\tilde{\mathbf{g}}$。本申请通过随机加速梯度下降策略SAGDFL，以减少局部和全局梯度方差；可以大大缩短训练迭代，同时保证模型的推理精度；可以确保隐私泄露风险可控；使额外的通信开销最小化，使模型参数(如梯度和权值)变得稀疏，易于将传输的数据进行压缩，降低通信成本。

附图说明

[0040] 为了更清楚地说明本申请或相关技术中的技术方案，下面将对实施例或相关技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附

图获得其他的附图。

[0041] 图1为本申请实施例的基于随机加速梯度下降的半联邦学习方法流程示意图；

[0042] 图2为本申请实施例的随机加速梯度下降策略SAGDFL形象化示意图；

[0043] 图3为本申请实施例的基于随机加速梯度下降的半联邦学习装置示意图；

[0044] 图4为本申请实施例的电子设备的结构示意图。

具体实施方式

[0045] 为使本申请的目的、技术方案和优点更加清楚明白，以下结合具体实施例，并参照附图，对本申请进一步详细说明。

[0046] 需要说明的是，除非另外定义，本申请实施例使用的技术术语或者科学术语应当为本申请所属领域内具有一般技能的人士所理解的通常意义。本申请实施例中使用的“包括”或者“包含”等类似的词语意指出现该词前面的元件或者物件涵盖出现在该词后面列举的元件或者物件及其等同，而不排除其他元件或者物件。

[0047] 机器学习的持续发展极大地促进了诸如人脸识别和语音识别等在内的人工智能(AI)领域的兴起。目前，大多数人工智能应用都是基于模型训练和推理实现的。而随机梯度下降法(SGD)作为求解大规模优化问题的一种普遍方法，在人工智能应用的模型训练中得到了广泛的应用。但是由于在每次迭代计算梯度时，为避免输入数据量过大，往往会随机抽样部分数据来替代全局数据计算梯度并更新模型参数，这一过程中SGD不可避免地会对梯度引入方差。这种方差不仅不利于模型训练，而且不利于模型推理。主要表现在训练收敛速度慢，推理精度低；无法在短时间内训练高精度模型，这也是在一些时间敏感的场景下，人工智能应用程序很难发展的一个主要原因。

[0048] 随着物联网设备和网络终端设备中数据的爆炸式增长，跨数据集分析处理的需求正在不断增加。然而，随着隐私保护法案的不断出台，数据世界又面临了新的难题“数据孤岛”。法律的约束和社会对数据安全的关注使得传统的基于数据收集的集中式或分布式的训练不再可行。在数据无法共享的情况下，数据只能在本地进行使用，在这种情况下通过聚合从每个设备本地计算的参数或梯度来训练全局模型的联邦学习(FL)应运而生。在联邦学习中，由于数据被保留在本地而不进行移动，因此可以对本地数据起到天然的保护作用。同时，这种新的学习方式也能帮助打破了“数据孤岛”困境。然而，由于缺乏了数据中心等平台收集数据进行统一的处理，不同设备之间的本地数据将很难保证独立同分布条件，导致传统机器学习在联邦学习训练中所面临的梯度方差问题比在集中式和分布式场景下都更加严重。

[0049] 相关技术中，联邦学习最初用来联合大量的移动设备来训练一个全局模型，其中每台设备都拥有近似特征的数据。然后每个设备利用随机梯度下降法SGD训练一个局部模型，训练过程中数据会被分成几个批次，每次只选择一批数据来代表局部数据计算梯度，并利用这一梯度对模型参数进行更新，这一方式也即是所谓的小批量梯度下降法mini-batch SGD。

[0050] 但是在此过程中，每个局部模型都会受到局部方差的影响。而当局部模型聚合之后，每个局部模型所携带的局部方差之间会相互作用，对聚合所得的全局模型产生不可预测的影响。因此，联邦学习往往需要进行更多次的模型训练才能达到较好的收敛。然而更多

的迭代，也意味着更高的通信成本，以及更高的隐私泄露和恶意攻击的风险。这将使模型的实际训练和部署更加困难。其中的根本原因是在研究过程中，通常假设每个设备上的数据是独立同分布的(IID)，但是，在大多数情况下，设备上的数据是非独立同分布Non-IID的，而这也将使得上述问题更加严重。

[0051] 非独立同分布数据在分布式学习中是一个不容忽视的问题，尤其是在联邦学习中。本地数据会受到设备所处位置、户主偏好等客观因素的影响，呈现出特定的分布。一旦本地数据发生了倾斜，设备之间的梯度方差就会很大，这将极大降低通信效率和推理准确性。

[0052] 相关技术中，尽管已经提出了诸如部分数据共享、蒸馏、聚类、局部批量归一化等方法来处理非独立同分布问题，但如何在高度倾斜的数据下尽快训练出高质量的模型仍然是联邦学习面临的巨大挑战。相关实验研究表明，即使是现有的最先进的联邦学习算法也不能在非独立同分布数据的所有场景下都是最优的。

[0053] 有鉴于此，本发明实施例提供一种基于随机加速梯度下降的半联邦学习方法及装置，通过建立一个小规模数据集，在服务器端训练一个近似的全局梯度，通过这一梯度来减少局部和全局的梯度方差，而不需要额外的局部存储，以下为本发明实施例的具体内容。

[0054] 参见图1，本实施例提供一种基于随机加速梯度下降的半联邦学习方法，包括以下步骤：

[0055] S1、利用设备数据和开源数据获得预设规模的独立同分布数据子集D，独立同分布数据子集D包括指定标签数据，每个指定标签的数据数量差异在预设差异范围内；

[0056] S2、进行梯度预训练，通过SAGDFL和规则I训练得到全局梯度$\tilde{\mathbf{g}}$，SAGDFL为随机加速梯度下降策略，规则I为：$\tilde{\mathbf{g}} = \tilde{\mathbf{g}} + \frac{1}{n}\sum_{j=1}^{n} \mathbf{g}_j^*$，式中，$\mathbf{g}_j^*$为设备j的局部梯度，n为训练批次；

[0057] S3、通过SAGDFL在设备和服务器之间执行联邦学习，服务器识别参加训练的设备组S_r，将模型$\tilde{\mathbf{w}}$和全局梯度$\tilde{\mathbf{g}}$发送给设备组S_r中的设备以进行SAGDFL处理；

[0058] S4、将每个设备生成的数据Δw_j和局部梯度$\mathbf{g}_j^*$上传到服务器，服务器根据数据Δw_j和局部梯度$\mathbf{g}_j^*$，聚合生成新的模型$\tilde{\mathbf{w}}$和新的全局梯度$\tilde{\mathbf{g}}$。

[0059] 本实施例中，为了减小局部和全局方差，设计了随机加速梯度下降策略SAGDFL。SAGDFL实现的核心是对全局梯度进行无偏估计。为此，通过半联邦学习框架来帮助计算全局梯度的无偏估计。同时，半联邦学习框架中引入的额外通信成本也是一个不容忽视的问题。进一步通过联合优化，在确保不对训练精度产生影响的情况下使所需通信的数据稀疏化，从而实现无损压缩传输。

[0060] 参见图2，本实施例中，描述了一种新的梯度下降策略，即随机加速梯度下降策略SAGDFL，可以帮助控制训练中的局部和全局方差。与经典方法SVRG一开始估计模型参数w不同，本发明在每一轮开始时计算或更新全局梯度$\tilde{\mathbf{g}}$的无偏估计。

[0061] 具体的，在每个设备接收到当前的全局模型$\tilde{\mathbf{w}}$和梯度$\tilde{\mathbf{g}}$后，本地训练将按照图2中的(a)部分所示的更新模式进行。在设备j中，第t-1轮的初始模型记作$\tilde{\mathbf{w}}$。如果通过传统的SGD进行训练，则模型$w_j^{(t-1)}$通过训练将得到$w_j^{(t),*}$。通过每一批数据计算得到的梯度($w_j^{(t),*}$对

应的虚线)在方向上同本地的梯度$\mathbf{g}_j^*$是接近的。这两个梯度之间的差异($\mathbf{w}_i^{(t),*}$对应的虚线)是对本地模型的关键更新。因此,为了确保本地更新方向的正确性,即能够指向全局梯度($\mathbf{w}_j^{(t)}$和$\tilde{\mathbf{g}}$对应虚线)的方向,本发明通过在全局梯度中添加差值($w_j^{(t),*}$对应的虚线)来计算更新本地模型的实际梯度。在完成本地训练,得到本地模型($\mathbf{w}_j^{(t)}$方块和$\mathbf{w}_i^{(t)}$方块)后,还需要对模型进行聚合,如图2中的(b)部分所示。通过SAGDFL最终聚合而成的模型($w^{(t)}$方块)相较于利用$\mathbf{w}_j^{(t),*}$和$\mathbf{w}_i^{(t),*}$直接聚合而成的模型要明显更接近于全局梯度方向,换而言之前者的全局方差要更小。

[0062] 本实施例中,假设每个设备中的数据在被分成若干批数据之前进行了随机洗牌,从而可以认为每批数据在本地都是独立同分布的。因此,在本地训练中,设备j的梯度g_j可以通过本地数据中的任意一批进行计算:

[0063] $$\mathbf{g}_j^* = \nabla f\left(\mathbf{x}_j^*, \tilde{w}\right)$$

[0064] 设备j的本地训练可以按照以下更新规则进行:

[0065] $$\mathbf{w}_j^{(t)} = \mathbf{w}_j^{(t-1)} - \eta_t \mathbf{g}$$

[0066] 其中:

[0067] $$\mathbf{g} = \nabla f_j\left(\mathbf{x}_j^i, \mathbf{w}_j^{(t-1)}\right) - \mathbf{g}_j^* + \tilde{\mathbf{g}}$$ 。

[0068] 与传统SGD相比,本实施例引入了两个额外的项来控制更新时的梯度方差,即随机局部梯度$\mathbf{g}_j^*$和无偏全局梯度$\tilde{\mathbf{g}}$。在这个更新中,$\nabla f_j\left(\mathbf{x}_j^i, \mathbf{w}_j^{(t-1)}\right)$和$\mathbf{g}_j^*$的值会是相似的,因为他们是用同一来源的数据计算的。这意味着全局梯度$\tilde{\mathbf{g}}$将是每个梯度g的主体。因此,当所有梯度都接近全局梯度时,局部方差和全局方差都会很小。

[0069] 事实上,全局梯度$\tilde{\mathbf{g}}$不是固定的,还需要随着产品,即每个设备在每一轮中得到的局部梯度$\mathbf{g}_j^*$,进行更新。全局梯度的更新可分为两个阶段,即梯度预训练阶段和联邦学习阶段。在这两个阶段,应用不同的规则来更新全局梯度。本实施例给出两个更新全局梯度$\tilde{\mathbf{g}}$的可选规则:Rule I:$\tilde{\mathbf{g}} = \tilde{\mathbf{g}} + \frac{1}{n}\sum_{j=1}^{n} \mathbf{g}_j^*$;Rule II:$\tilde{\mathbf{g}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{g}_j^*$。

[0070] 具体的,规则I将局部梯度的加权平均直接加到旧梯度上更新全局梯度,而规则II则将局部梯度的加权平均直接作为新的全局梯度。与在规则II下的训练相比,模型在规则I下训练得更快,因为在规则I下得到的新全局梯度更大。此外,由于每一步的梯度都会被记录和累积,初始模型可以使用规则I训练的最终全局梯度快速接近最优模型。但是,如果在联邦学习阶段仍然使用规则I训练模型,那么模型很容易过拟合。相反,规则II下的模型训练性能更稳定。因此,为了加快梯度下降速度,同时保证训练效果,设计了一种基于规则I和规则II的训练策略。

[0071] 具体的,SAGDFL中,联邦学习阶段的随机加速梯度下降算法如下:

算法 1 SAGDFL：FL 的随机加速梯度下降

1:**输入：**客户端的学习率为η_t，服务器的学习率为α_t，模型$\tilde{\mathbf{w}}$和全局梯度$\tilde{\mathbf{g}}$

2:对每一轮迭代$t = 1,2,\cdots,T$执行

3: $S_t \leftarrow$（随机采样一个设备组）

4: 对每个设备$j \in S_t$并行执行

5: 随机选取$\mathbf{x}_j^* \in \mathbf{X}_j$，并计算本地梯度$\mathbf{g}_j^*$

6: $\mathbf{g}_j^* = \nabla f(\mathbf{x}_j^*, \tilde{\mathbf{w}})$

7: $\mathbf{w}_{j,1} = \tilde{\mathbf{w}}$

[0072]

8: 对本地迭代$r = 1,2,\cdots,R$执行

9: $\mathbf{w}_{j,r+1} \leftarrow \mathbf{w}_{j,r} - \eta_t(\nabla f(x_r, w_{j,r}) - \mathbf{g}_j^* + \tilde{\mathbf{g}})$

10: $\Delta\mathbf{w}_j = \mathbf{w}_{j,R} - \tilde{\mathbf{w}}$

11: 服务器聚合：

12: $\tilde{\mathbf{w}} \leftarrow \tilde{\mathbf{w}} + \alpha_t \frac{1}{|S_t|}\sum_{j\in S_t}\Delta\mathbf{w}_j$

13: Rule I：$\tilde{\mathbf{g}} = \tilde{\mathbf{g}} + \frac{1}{n}\sum_{j=1}^{n}\mathbf{g}_j^*$

14: Rule II：$\tilde{\mathbf{g}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{g}_j^*$

[0073] SAGDFL策略的基本思想如下。在梯度预训练阶段，想要获得一个可以帮助快速从初始模型接近最优模型的最终梯度，所以应用规则I来更新全局梯度$\tilde{\mathbf{g}}$。在联邦学习阶段，为确保训练的稳定性，选择规则II来更新全局梯度。

[0074] 本实施例中，SAGDFL的设计有助于降低联邦学习的局部和全局方差，但全局梯度的无偏估计对于快速训练出高质量的模型具有重要意义。在不泄露隐私的情况下，从设备中收集一些低敏感数据是可行的。因此，设计半联邦学习框架，帮助SAGDFL通过使用部分全局数据或开源数据构建的小数据集进行训练，获得全局梯度$\tilde{\mathbf{g}}$。半联邦学习框架可细分为三个阶段，即数据收集阶段、梯度预训练阶段和模型训练阶段，实现算法如下：

算法 2 基于 SAGDFL 的半联邦学习

1:**数据收集：** 利用设备数据和开源数据获得 IID 的数据子集D

2: $D \leftarrow D + D_{augmentation}$

3:**初始化：**$\tilde{\mathbf{w}}$和$\tilde{\mathbf{g}}$

[0075]

4:**梯度预训练：**将D分为 n 批训练初始全局梯度：

5: $\tilde{\mathbf{g}} \leftarrow SAGDFL(\eta_t, \alpha_t, \tilde{\mathbf{w}}, \tilde{\mathbf{g}})$ with the **Rule I**

6:**联邦学习：**初始化的$\tilde{\mathbf{w}}$和预训练的$\tilde{\mathbf{g}}$

7: $(\tilde{\mathbf{w}}, \tilde{\mathbf{g}}) \leftarrow SAGDFL(\eta_t, \alpha_t, \tilde{\mathbf{w}}, \tilde{\mathbf{g}})$ with the **Rule II**

[0076] 首先,收集一个预设规模(规模较小)的独立同分布数据子集D。独立同分布数据子集D应该包含所有标签的数据,每个标签的数据数量应该相近,从而,用这个独立同分布数据子集D计算的初始$\tilde{\mathbf{g}}$可以被视为全局梯度的无偏估计。

[0077] 具体的,构建同分布数据子集D有两个主要的数据来源,分别是来自所有设备的本地数据和来自互联网的开源数据。在从设备采集数据之前,需要对本地数据进行敏感性评估。考虑到隐私和安全问题,服务器只会收集低敏感的数据,但是,在这种情况下很难确保收集到的数据集是IID的。因此,通过翻转、平移、旋转等数据扩充方法来扩充不足的数据样本。对于服务器来说,在权衡数据收集开销和训练性能提升之后,收集一个小型IID数据集是可以接受的。在一定的激励机制之下,设备也可以通过对数据收集的贡献得到奖励。

[0078] 其中,梯度预训练阶段,在服务器上计算初始全局梯度$\tilde{\mathbf{g}}$。为了避免在$\tilde{\mathbf{g}}$中引入方差,$\tilde{\mathbf{g}}$在SAGDFL与规则I下训练得到。在这个过程中,同分布数据子集D的每一批数据可以视为是联邦学习中的设备。每一批的数据可以通过挑选来模拟全局数据的分布。在每一轮SAGDFL中使用所有批次数据来训练全局梯度$\tilde{\mathbf{g}}$,训练过程持续进行,直到损失不再减少。

[0079] 具体的,通过SAGDFL在设备和服务器之间执行联邦学习。在每一轮中,服务器首先识别参加训练的设备组S_r。然后服务器将最新的模型$\widetilde{\mathbf{w}}$和全局梯度$\tilde{\mathbf{g}}$发送给这些设备,以便进行SAGDFL。在SAGDFL中,每个设备生成的Δw_j和$\mathbf{g}_j^*$会被上传到服务器,以便服务器可以聚合一个新的模型$\widetilde{\mathbf{w}}$和一个新的全局梯度$\tilde{\mathbf{g}}$。

[0080] 本实施例中,一对服务器和设备之间的数据传输量是FedAvg中相同情况下的两倍。如果在数据传输过程中考虑加密来保护数据的私密性,则会进一步增加通信成本。这也将大大增加传输时延。因此,降低SAGDFL的通信成本至关重要。传统做法是在传输数据之前压缩数据。然而,这种压缩往往会对模型更新产生不可预见的影响,甚至会给联邦学习引入新的方差。同时,在利用高度倾斜数据进行的局部训练中,经过几轮训练后,梯度很容易变得稀疏。因此,本发明不直接压缩数据,而是考虑先对需要上传的数据Δw_j进行稀疏化处理,即只保留关键参数的值,不相关参数的值置零。这一过程可以通过设置一个阈值来区分局部模型中的关键参数和无关参数。但这种方式会引发计算量增加,模型性能下降等不稳定因素。因此,考虑通过联合优化,在保证最终模型的性能等同时使Δw_j稀疏化。

[0081] 具体的,解决联合优化问题时,在目标函数中加入了一阶惩罚函数。即,本发明没有让w_j变得稀疏,而是希望Δw_j是稀疏的,这样梯度也会是稀疏的。联合优化的目标函数提出如下式:

[0082] $$w_j = \arg\min_{w\in\mathcal{W}} \langle g_t, w\rangle + \frac{1}{2\eta_t}\|w - w_j\|_2^2 + \frac{1}{2\eta_t}\|w - w_j\|_1^2$$

[0083] 其中,$g_t := \nabla f_j(w_j) - \nabla f_j(\widetilde{w}) + \tilde{g}$。

[0084] 在得到稀疏通信数据后,可以使用现有的任意技术对数据进行无损压缩,如稀疏三值压缩(STC),稀疏抖动(SD)。联合优化获得稀疏Δw_j和g_j的优势在于,既能保证Δw_j的质量,又能实现压缩目的。

[0085] 通过实验评估在本发明帮助下逻辑回归(LR)和卷积神经网络(CNN)模型的效果表

现。在LR模型中,SAGDFL在所有参数状态下都显著优于FedAvg和SCAFFOLD;在CNN模型的训练中,SAGDFL仍然获得了良好的加速效果。

[0086] 实验过程中,包括两个数据集上的两个模型族。不同实验的基本设置是一致的。将训练集按照一定的规则排序,并划分给100个设备,每次迭代随机选择10个设备。

[0087] 第一个实验是使在逻辑回归上进行MNIST数字识别任务。另一个实验是在CNN上进行FEMNIST分类任务,CNN设置如下:有两个5×5卷积层(第一个有32个通道,第二个有64个,每个都有2×2最大池化),一个有512个单元和ReLu激活器的全连接层,以及一个softmax输出层(总共1663370个参数)。在两个实验中,研究不同算法在三种数据分布下的性能。对于IID设置,数据被随机打乱,然后均匀地划分到100个设备中。对于Non-IID(1)设置,数据按标签排序,每个设备分配一类数据。对于Non-IID(2)设置,数据按标签排序,每个设备分配两类数据。

[0088] 小规模的IID数据集是通过从训练集中随机选取1%的数据来收集的。因此,它是一个IID数据集。每个实验中局部训练的批大小B设为100,局部迭代E设为1。在局部训练中,输入大小为B的批量数据而非所有的局部数据来计算梯度,并进行一次模型更新。此外,局部迭代表明,在FL的每次迭代中,局部的每批数据将使用E次来训练模型。本发明用相同的初始模型实现了FedAvg、SCAFFOLD和我们的SAGDFL,并比较了它们的准确性和损失。

[0089] 在训练线性模型时,如LR,SAGDFL在收敛性和测试精度方面的优势是非常明显的,特别是当数据分布是非IID的时候。在CNN训练中,SAGDFL与SCAFFOLD相比,测试精度的提高幅度虽然不大,但速度仍然比SCAFFOLD和FedAvg都快,特别是在训练开始时,可以观察到更多训练细节,具体如下:

[0090] SAGDFL可以减小FL的方差,并提高其性能。无论在MNIST LR场景还是FEMNIST CNN场景中,SAGDFL都可以在Non-IID数据下得到接近于IID数据下的结果。这证实了SAGDFL在减少异构数据引起的方差方面是有效的。同时,模型测试精度的提高表明,方差是导致模型精度下降的原因。

[0091] SAGDFL总是最快收敛的。在所有设置中,SAGDFL的训练收敛速度总是比SCAFFOLD和FedAvg快。这种速度对提升得益于规则I的全局梯度积累,可以帮助模型快速接近最优区域。规则II在训练后期可以进一步提高模型的准确性。

[0092] SAGDFL在LR中具有鲁棒性。对于两种倾斜数据场景,FedAvg的收敛是不稳定的,因为每一轮选择的设备都是随机的。这意味着在每一轮训练中,每个标签数据的数量不相等。换句话说,不同数据在每一轮中对模型的贡献度都不同,而这也是导致全局方差对原因之一。但是,结果表明在相同场景下我们的SAGDFL在训练LR上是稳健的,并没有受到因设备的随机选择所带来的数据分布差异的影响。

[0093] 此外,本发明在非IID(1)中验证了与FEMNIST CNN联合优化的有效性,通过联合优化,通信数据Δw_j会变得稀疏,特别是在训练的早期。同时,通过联合优化训练得到的模型性能与原SAGDFL没有太大差别。这表明,在现有压缩技术的加持下,联合优化不仅有助于降低SAGDFL的额外通信成本,还能保证最终模型与SAGDFL直接训练得到模型保有相近的准确性。

[0094] 综上所述,本申请通过利用设备数据和开源数据获得预设规模的独立同分布数据子集D,独立同分布数据子集D包括指定标签数据,每个指定标签的数据数量差异在预设差

异范围内;进行梯度预训练,通过SAGDFL和规则I训练得到全局梯度$\tilde{\mathbf{g}}$,SAGDFL为随机加速梯度下降策略,规则I为:$\tilde{\mathbf{g}} = \tilde{\mathbf{g}} + \frac{1}{n}\sum_{j=1}^{n} \mathbf{g}_j^*$,式中,$\mathbf{g}_j^*$为设备j的局部梯度,n为训练批次;通过SAGDFL在设备和服务器之间执行联邦学习,服务器识别参加训练的设备组S_r,将模型$\tilde{\mathbf{w}}$和全局梯度$\tilde{\mathbf{g}}$发送给设备组S_r中的设备以进行SAGDFL处理;将每个设备生成的数据Δw_j和局部梯度$\mathbf{g}_j^*$上传到服务器,服务器根据数据Δw_j和局部梯度$\mathbf{g}_j^*$,聚合生成新的模型$\tilde{\mathbf{w}}$和新的全局梯度$\tilde{\mathbf{g}}$。本申请通过随机加速梯度下降策略SAGDFL,以减少局部和全局梯度方差;可以大大缩短训练迭代,同时保证模型的推理精度;可以确保隐私泄露风险可控;使额外的通信开销最小化,使模型参数(如梯度和权值)变得稀疏,易于将传输的数据进行压缩,降低通信成本。实验表明,SAGDFL能够有效地降低由于数据异构而引起的局部和全局梯度方差,使其精度从FedAvg的74.8%、SCAFFOLD的82.5%提高到SAGDFL的89.1%。此外,通过联合优化,在不降低最终模型精度的前提下,使通信数据Δw_j稀疏化,从而实现Δw_j的无损压缩,降低通信成本。尤其是在CNN的训练中,基于SAGDFL的半联邦学习的通信效率和性能还可以进一步提高。

[0095] 需要说明的是,本申请实施例的方法可以由单个设备执行,例如一台计算机或服务器等。本实施例的方法也可以应用于分布式场景下,由多台设备相互配合来完成。在这种分布式场景的情况下,这多台设备中的一台设备可以只执行本申请实施例的方法中的某一个或多个步骤,这多台设备相互之间会进行交互以完成所述的方法。

[0096] 需要说明的是,上述对本申请的一些实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下,在权利要求书中记载的动作或步骤可以按照不同于上述实施例中的顺序来执行并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中,多任务处理和并行处理也是可以的或者可能是有利的。

[0097] 参见图3,基于同一发明构思,与上述任意实施例方法相对应的,本申请还提供了一种基于随机加速梯度下降的半联邦学习装置,包括:

[0098] 数据子集构建模块11,用于利用设备数据和开源数据获得预设规模的独立同分布数据子集D,独立同分布数据子集D包括指定标签数据,每个指定标签的数据数量差异在预设差异范围内;

[0099] 梯度预训练模块12,用于进行梯度预训练,通过SAGDFL和规则I训练得到全局梯度$\tilde{\mathbf{g}}$,SAGDFL为随机加速梯度下降策略,规则I为:$\tilde{\mathbf{g}} = \tilde{\mathbf{g}} + \frac{1}{n}\sum_{j=1}^{n} \mathbf{g}_j^*$,式中,$\mathbf{g}_j^*$为设备j的局部梯度,n为训练批次;

[0100] 联邦学习模块13,用于通过SAGDFL在设备和服务器之间执行联邦学习,服务器识别参加训练的设备组S_r,将模型$\tilde{\mathbf{w}}$和全局梯度$\tilde{\mathbf{g}}$发送给设备组S_r中的设备以进行SAGDFL处理;

[0101] 训练结果处理模块14,用于将每个设备生成的数据Δw_j和局部梯度$\mathbf{g}_j^*$上传到服务器,服务器根据数据Δw_j和局部梯度$\mathbf{g}_j^*$,聚合生成新的模型$\tilde{\mathbf{w}}$和新的全局梯度$\tilde{\mathbf{g}}$。

[0102] 本实施例中，数据子集构建模块11中，将利用独立同分布数据子集D计算的初始值$\tilde{\mathbf{g}}$作为全局梯度的无偏估计；独立同分布数据子集D来源于设备的本地数据和互联网的开源数据；

[0103] 数据子集构建模块11包括：

[0104] 敏感性评估子模块111，用于对采集的设备的本地数据通过隐私测量仪进行敏感性评估，隐私测量仪采用ML privacy meter；

[0105] 样本扩充子模块112，用于对采集的互联网的开源数据，通过翻转、平移或旋转的方式进行数据样本扩充。

[0106] 本实施例中，还包括稀疏化处理模块15，用于对需要上传的数据Δw_j进行稀疏化处理，稀疏化处理过程保留关键参数的值，不相关参数的值置零；

[0107] 稀疏化处理模块15中，通过配置阈值区分局部模型中的关键参数和无关参数。

[0108] 本实施例中，还包括联合优化模块16，用于通过联合优化使上传的数据Δw_j稀疏化，联合优化过程在目标函数中加入一阶惩罚函数；

[0109] 还包括压缩处理模块17，用于获得稀疏化处理的数据后，对数据进行无损压缩。

[0110] 本实施例中，联邦学习模块13中，采用规则II更新全局梯度$\tilde{\mathbf{g}}$，规则II为：

[0111] $$\tilde{\mathrm{g}}=\frac{1}{n}\sum_{j=1}^{n}\mathrm{g}_j^*$$

[0112] 式中，$\mathbf{g}_j^*$为设备j的局部梯度，n为训练批次。

[0113] 为了描述得方便，描述以上系统时以功能分为各种模块分别描述。当然，在实施本申请时可以把各模块的功能在同一个或多个软件和/或硬件中实现。

[0114] 上述实施例的装置用于实现前述任一实施例中相应地基于随机加速梯度下降的半联邦学习方法，并且具有相应的方法实施例的有益效果，在此不再赘述。

[0115] 基于同一发明构思，与上述任意实施例方法相对应的，本申请还提供了一种电子设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现上任意一实施例所述的基于随机加速梯度下降的半联邦学习方法。

[0116] 图4示出了本实施例所提供的一种更为具体的电子设备硬件结构示意图，该设备可以包括：处理器1010、存储器1020、输入/输出接口1030、通信接口1040和总线1050。其中处理器1010、存储器1020、输入/输出接口1030和通信接口1040通过总线1050实现彼此之间在设备内部的通信连接。

[0117] 处理器1010可以采用通用的CPU(Central Processing Unit，中央处理器)、微处理器、应用专用集成电路(Application Specific Integrated Circuit，ASIC)、或者一个或多个集成电路等方式实现，用于执行相关程序，以实现本说明书实施例所提供的技术方案。

[0118] 存储器1020可以采用ROM(Read Only Memory，只读存储器)、RAM(Random Access Memory，随机存取存储器)、静态存储设备，动态存储设备等形式实现。存储器1020可以存储操作系统和其他应用程序，在通过软件或者固件来实现本说明书实施例所提供的技术方案时，相关的程序代码保存在存储器1020中，并由处理器1010来调用执行。

[0119] 输入/输出接口1030用于连接输入/输出模块，以实现信息输入及输出。输入输出/模块可以作为组件配置在设备中(图中未示出)，也可以外接于设备以提供相应功能。其中输入设备可以包括键盘、鼠标、触摸屏、麦克风、各类传感器等，输出设备可以包括显示器、扬声器、振动器、指示灯等。

[0120] 通信接口1040用于连接通信模块(图中未示出)，以实现本设备与其他设备的通信交互。其中通信模块可以通过有线方式(例如USB、网线等)实现通信，也可以通过无线方式(例如移动网络、WIFI、蓝牙等)实现通信。

[0121] 总线1050包括一通路，在设备的各个组件(例如处理器1010、存储器1020、输入/输出接口1030和通信接口1040)之间传输信息。

[0122] 需要说明的是，尽管上述设备仅示出了处理器1010、存储器1020、输入/输出接口1030、通信接口1040以及总线1050，但是在具体实施过程中，该设备还可以包括实现正常运行所必需的其他组件。此外，本领域的技术人员可以理解的是，上述设备中也可以仅包含实现本说明书实施例方案所必需的组件，而不必包含图中所示的全部组件。

[0123] 上述实施例的电子设备用于实现前述任一实施例中相应地基于随机加速梯度下降的半联邦学习方法，并且具有相应的方法实施例的有益效果，在此不再赘述。

[0124] 基于同一发明构思，与上述任意实施例方法相对应的，本申请还提供了一种非暂态计算机可读存储介质，所述非暂态计算机可读存储介质存储计算机指令，所述计算机指令用于使所述计算机执行如上任一实施例所述的基于随机加速梯度下降的半联邦学习方法。

[0125] 本实施例的计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。

[0126] 上述实施例的存储介质存储的计算机指令用于使所述计算机执行如上任一实施例所述的基于随机加速梯度下降的半联邦学习方法，并且具有相应的方法实施例的有益效果，在此不再赘述。

[0127] 所属领域的普通技术人员应当理解：以上任何实施例的讨论仅为示例性的，并非旨在暗示本申请的范围(包括权利要求)被限于这些例子；在本申请的思路下，以上实施例或者不同实施例中的技术特征之间也可以进行组合，步骤可以以任意顺序实现，并存在如上所述的本申请实施例的不同方面的许多其它变化，为了简明它们没有在细节中提供。

[0128] 另外，为简化说明和讨论，并且为了不会使本申请实施例难以理解，在所提供的附图中可以示出或可以不示出与集成电路(IC)芯片和其它部件的公知的电源/接地连接。此外，可以以框图的形式示出装置，以便避免使本申请实施例难以理解，并且这也考虑了以下事实，即关于这些框图装置的实施方式的细节是高度取决于将要实施本申请实施例的平台的(即，这些细节应当完全处于本领域技术人员的理解范围内)。在阐述了具体细节(例如，电路)以描述本申请的示例性实施例的情况下，对本领域技术人员来说显而易见的是，可以

在没有这些具体细节的情况下或者这些具体细节有变化的情况下实施本申请实施例。因此,这些描述应被认为是说明性的而不是限制性的。

[0129] 尽管已经结合了本申请的具体实施例对本申请进行了描述,但是根据前面的描述,这些实施例的很多替换、修改和变型对本领域普通技术人员来说将是显而易见的。例如,其它存储器架构(例如,动态RAM(DRAM))可以使用所讨论的实施例。

[0130] 本申请实施例旨在涵盖落入所附权利要求的宽泛范围之内的所有这样的替换、修改和变型。因此,凡在本申请实施例的精神和原则之内,所做的任何省略、修改、等同替换、改进等,均应包含在本申请的保护范围之内。

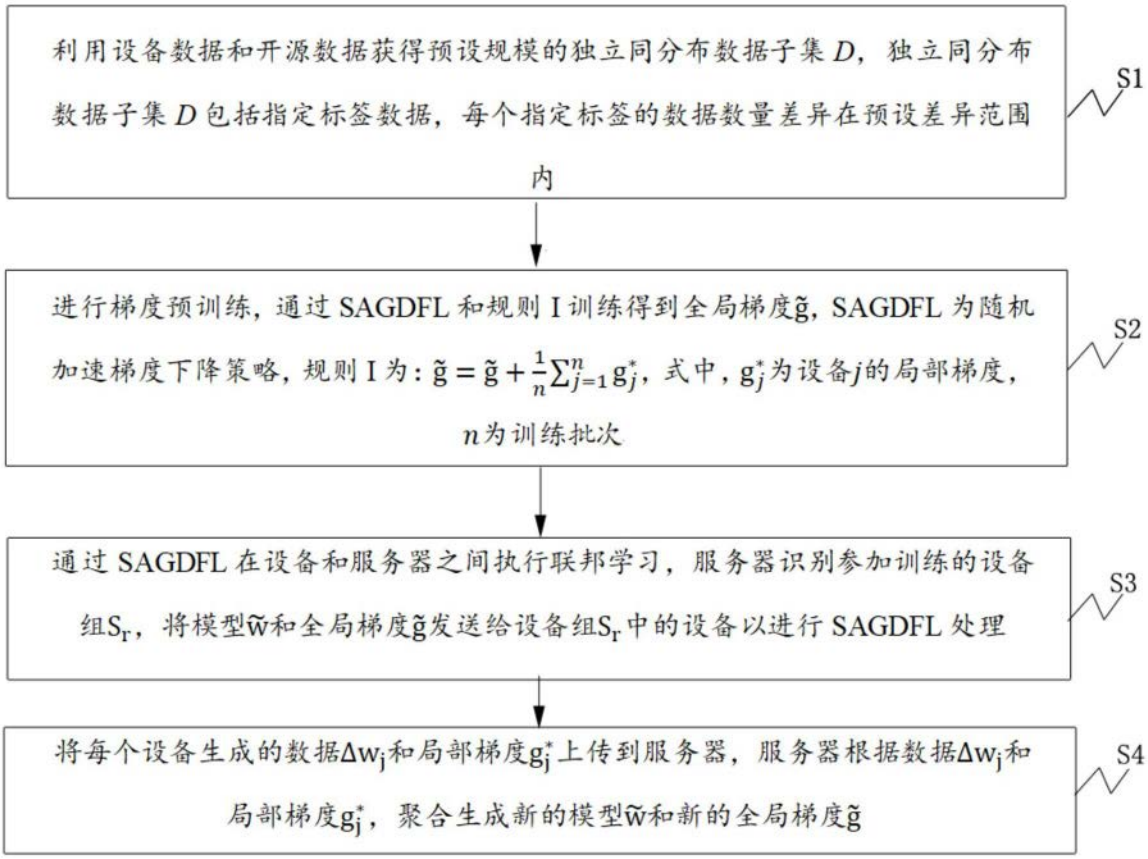

利用设备数据和开源数据获得预设规模的独立同分布数据子集D，独立同分布数据子集D包括指定标签数据，每个指定标签的数据数量差异在预设差异范围内
S1
进行梯度预训练，通过SAGDFL和规则I训练得到全局梯度$\tilde{g}$，SAGDFL为随机加速梯度下降策略，规则I为：$\tilde{g}=\tilde{g}+\frac{1}{n}\sum_{j=1}^{n}g_j^*$，式中，$g_j^*$为设备$j$的局部梯度，$n$为训练批次
S2
通过SAGDFL在设备和服务器之间执行联邦学习，服务器识别参加训练的设备组S_r，将模型$\tilde{w}$和全局梯度$\tilde{g}$发送给设备组S_r中的设备以进行SAGDFL处理
S3
将每个设备生成的数据Δw_j和局部梯度g_j^*上传到服务器，服务器根据数据Δw_j和局部梯度g_j^*，聚合生成新的模型$\tilde{w}$和新的全局梯度$\tilde{g}$
S4

图1

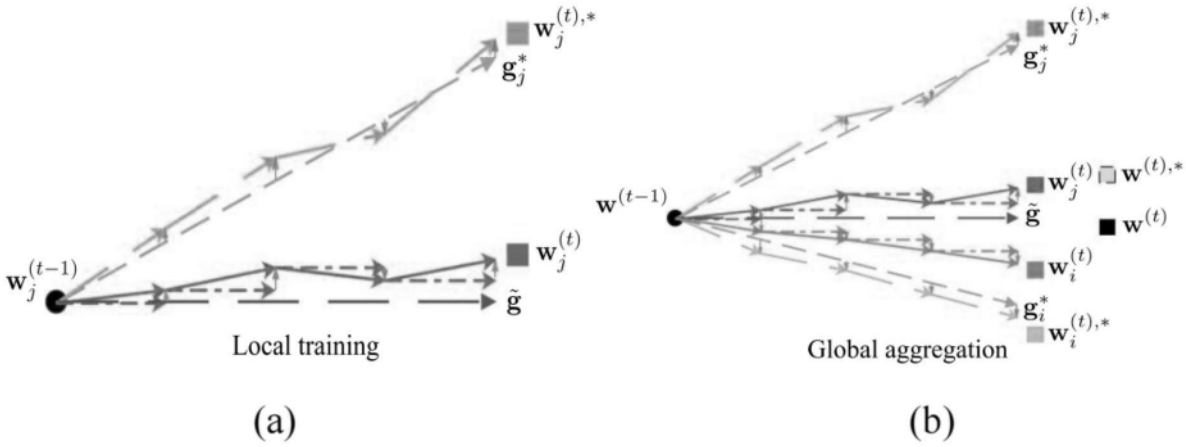

$w_j^{(t),*}$
g_j^*
$w_j^{(t-1)}$
$w_j^{(t)}$
$\tilde{g}$
Local training
(a)
$w^{(t-1)}$
$w^{(t)}$
$w_i^{(t)}$
g_i^*
$w_i^{(t),*}$
Global aggregation
(b)

图2

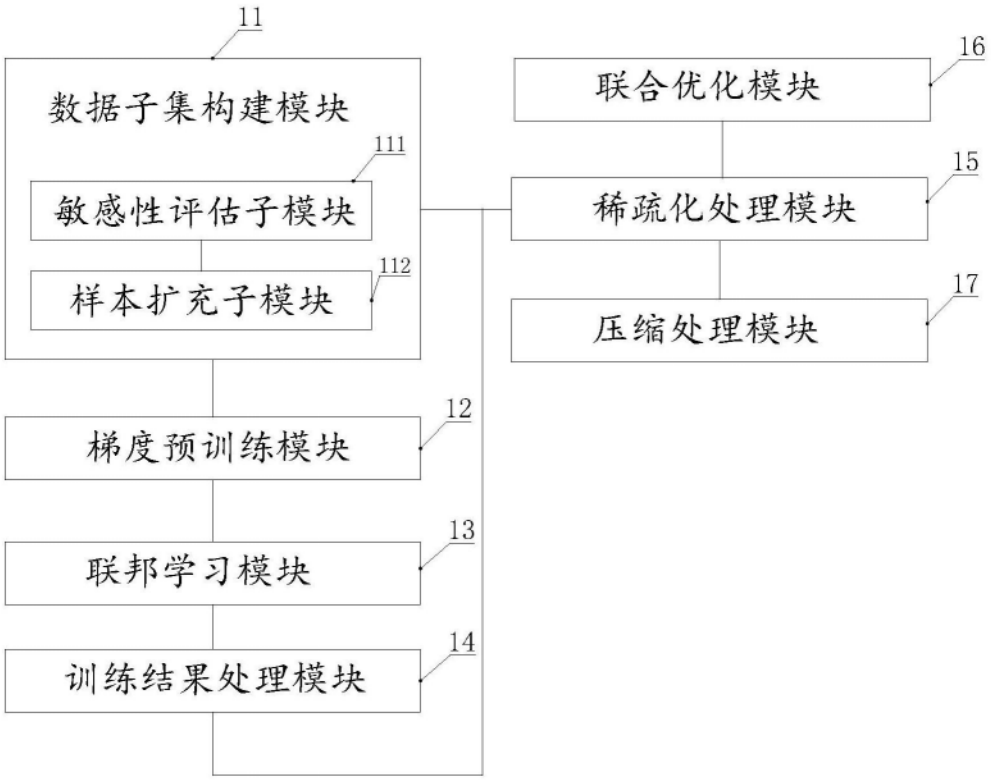

11
数据子集构建模块
111
敏感性评估子模块
112
样本扩充子模块
16
联合优化模块
15
稀疏化处理模块
17
压缩处理模块
12
梯度预训练模块
13
联邦学习模块
14
训练结果处理模块

图3

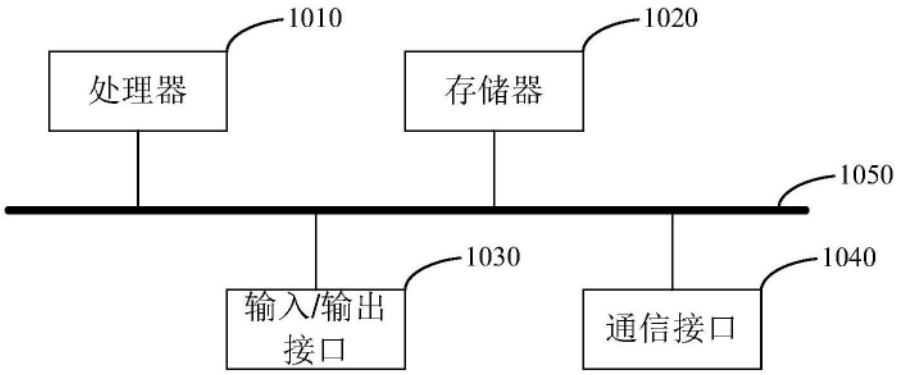

1010
处理器
1020
存储器
1050
1030
输入/输出
接口
1040
通信接口

图4