(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112990488 B
(45)授权公告日 2024.03.26

(21)申请号 202110279647.2

(22)申请日 2021.03.16

(65)同一申请的已公布的文献号
申请公布号 CN 112990488 A

(43)申请公布日 2021.06.18

(73)专利权人 香港理工大学深圳研究院
地址 518057 广东省深圳市南山区粤海街道高新技术产业园南区粤兴一道18号香港理工大学产学研大楼205室

(72)发明人 郭嵩 吴非杰 王号召

(74)专利代理机构 深圳市君胜知识产权代理事务所(普通合伙) 44268
专利代理师 朱阳波 王永文

(51)Int.Cl.
G06N 20/20(2019.01)

(56)对比文件
CN 110490738 A,2019.11.22
CN 111931950 A,2020.11.13
CN 112052958 A,2020.12.08
CN 112132277 A,2020.12.25
Sai Praneeth Karimireddy et al..SCAFFOLD: Stochastic Controlled Averaging for Federated Learning.《Proceedings of the 37th International Conference on Machine Learning》.2020,第1-12页.

审查员 任丽娜

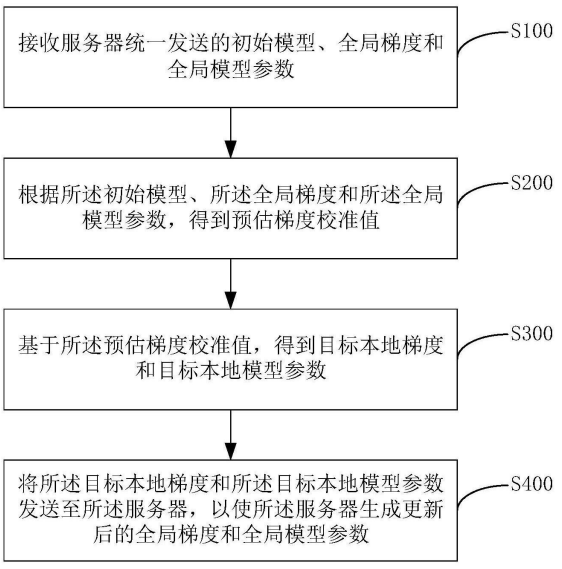
权利要求书2页 说明书9页 附图2页

(54)发明名称

一种基于机器异构性的联邦学习方法

(57)摘要

本发明公开了一种基于机器异构性的联邦学习方法,方法包括:接收服务器统一发送的初始模型、全局梯度和全局模型参数;根据初始模型、全局梯度和全局模型参数,得到预估梯度校准值;其中,预估梯度校准值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差以及各边缘设备因本地更新次数不同而产生的偏差;基于预估梯度校准值,得到目标本地梯度和目标本地模型参数;将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数。本发明实施例通过对各边缘设备的预估梯度校准技术来实现移除各边缘设备与服务器的偏差,同时补偿本地更新次数不同导致的偏差,从而提高联邦学习的训练效率。



CN 112990488 B

1.一种基于机器异构性的联邦学习方法,联邦学习方法用于不同地区的银行样本的联合,其特征在于,所述方法包括:

接收服务器统一发送的初始模型、全局梯度和全局模型参数,将所述全局模型参数作为初始本地模型参数;

根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值;其中,所述预估梯度校准值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差以及各边缘设备因本地更新次数不同而产生的偏差;

基于所述预估梯度校准值,得到目标本地梯度和目标本地模型参数;

将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数;

所述根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值包括:

基于预设的本地数据和所述初始模型,得到本地梯度;

基于所述全局梯度和所述本地梯度,得到第一偏差值,其中,所述第一偏差值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差;

基于所述全局模型参数和所述初始本地模型参数,得到第二偏差值,其中,所述第二偏差值用于表征各边缘设备因本地更新次数不同而产生的偏差;

获取本地更新次数;

根据所述本地更新次数和所述第二偏差值,得到中间第二偏差值;

将所述第一偏差值加上所述中间第二偏差值,得到预估梯度校准值c:

$$c = \tilde{v} - v^{i} + \frac{\lambda}{\eta K_i}(\widetilde{x_{t+1}} - x_{t;K_i}^{i})$$

$\tilde{v}$为全局梯度,v^{i}为本地梯度,λ为调整因子,η为学习率的步长,K_i为本地更新次数,$\widetilde{x_{t+1}}$为全局模型参数,$x_{t;K_i}^{i}$为目标本地模型参数。

2.根据权利要求1所述的基于机器异构性的联邦学习方法,其特征在于,所述基于所述预估梯度校准值,得到目标本地梯度和目标本地模型参数包括:

获取数据样本;其中,所述数据样本是从边缘设备的样本中获取得到;

根据所述数据样本和所述初始本地模型参数,得到损失函数偏导数;

根据所述损失函数偏导数,得到目标本地梯度;

根据所述损失函数偏导数、所述初始本地模型参数和所述预估梯度校准值,得到目标本地模型参数。

3.根据权利要求2所述的基于机器异构性的联邦学习方法,其特征在于,所述根据所述数据样本和所述初始本地模型参数,得到损失函数偏导数包括:

根据所述数据样本和所述初始本地模型参数,得到损失函数;

对所述损失函数进行求偏导,得到损失函数偏导数。

4.根据权利要求3所述的基于机器异构性的联邦学习方法,其特征在于,所述将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数之后包括:

接收服务器统一发送的更新后的全局梯度和全局模型参数,并重复执行根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值的步骤。

5.一种智能终端,其特征在于,包括有存储器,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于执行如权利要求1-4中任意一项所述的方法。

6.一种非临时性计算机可读存储介质,其特征在于,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行如权利要求1-4中任意一项所述的方法。

一种基于机器异构性的联邦学习方法

技术领域

[0001] 本发明涉及人工智能技术领域,尤其涉及的是一种基于机器异构性的联邦学习方法。

背景技术

[0002] 当前针对联邦学习的算法主要基于假设其能够在同样时间内在本地训练一定次数。这个方法在同构环境下能够提高训练的效率以及通讯开销,然而在绝大部分的分布式场景都属于异构环境,因此,这种计算方法并不具备实用性。如果以异步的方式解决这一问题的话,会存在一些数据无法充分利用(例如:如果一个边缘设备过久未与服务器更新的话,服务器中的异步算法可能会舍弃边缘设备提交的信息)。当各边缘设备的本地更新次数是相同的情况下,利用传统的随机梯度下降法SGD进行联邦学习的效果很好,而当各边缘设备本地更新次数不同的情况下,利用传统的随机梯度下降法SGD进行联邦学习,则出现需要优化的目标函数和实际优化的目标函数不一致的情况。

[0003] 因此,现有技术还有待改进和发展。

发明内容

[0004] 本发明要解决的技术问题在于,针对现有技术的上述缺陷,提供一种基于机器异构性的联邦学习方法,旨在解决现有技术中联邦学习中的异构网络在进行模型训练时训练效率低的问题。

[0005] 本发明解决问题所采用的技术方案如下:

[0006] 第一方面,本发明实施例提供一种基于机器异构性的联邦学习方法,其中,所述方法包括:

[0007] 接收服务器统一发送的初始模型、全局梯度和全局模型参数;

[0008] 根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值;其中,所述预估梯度校准值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差以及各边缘设备因本地更新次数不同而产生的偏差;

[0009] 基于所述预估梯度校准值,得到目标本地梯度和目标本地模型参数;

[0010] 将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数。

[0011] 在一种实现方式中,其中,所述接收服务器统一发送的初始模型、全局梯度和全局模型参数之后包括:

[0012] 将所述全局模型参数作为初始本地模型参数。

[0013] 在一种实现方式中,其中,所述根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值包括:

[0014] 基于预设的本地数据和所述初始模型,得到本地梯度;

[0015] 基于所述全局梯度、所述全局模型参数和所述本地梯度,得到预估梯度校准值。

[0016] 在一种实现方式中,其中,所述基于所述全局梯度、所述全局模型参数和所述本地梯度,得到预估梯度校准值包括:

[0017] 基于所述全局梯度和所述本地梯度,得到第一偏差值,其中,所述第一偏差值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差;

[0018] 基于所述全局模型参数和所述初始本地模型参数,得到第二偏差值,其中,所述第二偏差值用于表征各边缘设备因本地更新次数不同而产生的偏差;

[0019] 基于所述第一偏差值和所述第二偏差值,得到预估梯度校准值。

[0020] 在一种实现方式中,其中,所述基于所述第一偏差值和所述第二偏差值,得到预估梯度校准值包括:

[0021] 获取本地更新次数;

[0022] 根据所述本地更新次数和所述第二偏差值,得到中间第二偏差值;

[0023] 将所述第一偏差值加上所述中间第二偏差值,得到预估梯度校准值。

[0024] 在一种实现方式中,其中,所述基于所述预估梯度校准值,得到目标本地梯度和目标本地模型参数包括:

[0025] 获取数据样本;其中,所述数据样本是从边缘设备的样本中获取得到;

[0026] 根据所述数据样本和所述初始本地模型参数,得到损失函数偏导数;

[0027] 根据所述损失函数偏导数,得到目标本地梯度;

[0028] 根据所述损失函数偏导数、所述初始本地模型参数和所述预估梯度校准值,得到目标本地模型参数。

[0029] 在一种实现方式中,其中,所述根据所述数据样本和所述初始本地模型参数,得到损失函数偏导数包括:

[0030] 根据所述数据样本和所述初始本地模型参数,得到损失函数;

[0031] 对所述损失函数进行求偏导,得到损失函数偏导数。

[0032] 在一种实现方式中,其中,所述将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数之后包括:

[0033] 接收服务器统一发送的更新后的全局梯度和全局模型参数,并重复执行根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值的步骤。

[0034] 第二方面,本发明实施例还提供一种基于机器异构性的联邦学习装置,其中,所述装置包括:

[0035] 服务器的数据接收单元,用于接收服务器统一发送的初始模型、全局梯度和全局模型参数;

[0036] 预估梯度校准值获取单元,用于根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值;其中,所述预估梯度校准值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差以及各边缘设备因本地更新次数不同而产生的偏差;

[0037] 目标参数获取单元,用于基于所述预估梯度校准值,得到目标本地梯度和目标本地模型参数;

[0038] 目标参数发送单元,用于将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数。

[0039] 第三方面,本发明实施例还提供一种智能终端,包括有存储器,以及一个或者一个

以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于执行如上述任意一项所述的基于机器异构性的联邦学习方法。

[0040] 第四方面,本发明实施例还提供一种非临时性计算机可读存储介质,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行如上述中任意一项所述的基于机器异构性的联邦学习方法。

[0041] 本发明的有益效果:本发明公开了一种基于机器异构性的联邦学习方法,方法包括:接收服务器统一发送的初始模型、全局梯度和全局模型参数;根据初始模型、全局梯度和全局模型参数,得到预估梯度校准值;其中,预估梯度校准值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差以及各边缘设备因本地更新次数不同而产生的偏差;基于预估梯度校准值,得到目标本地梯度和目标本地模型参数;将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数。本发明实施例通过对各边缘设备的预估梯度校准技术来实现移除各边缘设备与服务器的偏差,同时补偿本地更新次数不同导致的偏差,从而使得每次本地更新尽可能接近于全局更新,并且,模型更新的效果不受各边缘设备更新次数的限制。

附图说明

[0042] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有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发明中记载的一些实施例,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。

[0043] 图1为本发明实施例提供的基于机器异构性的联邦学习方法流程示意图。

[0044] 图2为本发明实施例提供的基于机器异构性的联邦学习的算法效果图。

[0045] 图3为本发明实施例提供的基于机器异构性的联邦学习装置的原理框图。

[0046] 图4为本发明实施例提供的智能终端的内部结构原理框图。

具体实施方式

[0047] 本发明公开了基于机器异构性的联邦学习方法、智能终端、存储介质,为使本发明的目的、技术方案及效果更加清楚、明确,以下参照附图并举实施例对本发明进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。

[0048] 本技术领域技术人员可以理解,除非特意声明,这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是,本发明的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件,但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。应该理解,当我们称元件被“连接”或“耦接”到另一元件时,它可以直接连接或耦接到其他元件,或者也可以存在中间元件。此外,这里使用的“连接”或“耦接”可以包括无线连接或无线耦接。这里使用的措辞“和/或”包括一个或更多个相关联的列出项的全部或任一单元和全部组合。

[0049] 本技术领域技术人员可以理解,除非另外定义,这里使用的所有术语(包括技术术语和科学术语),具有与本发明所属领域中的普通技术人员的一般理解相同的意义。还应该

理解的是,诸如通用字典中定义的那些术语,应该被理解为具有与现有技术的上下文中的意义一致的意义,并且除非像这里一样被特定定义,否则不会用理想化或过于正式的含义来解释。

[0050] 由于现有技术中,多数联邦学习训练方法是针对同构环境的,对于异构环境并不实用性。如果以异步的方式解决这一问题的话,会存在一些数据无法充分利用,若单纯通过各边缘设备训练不等次数以提高利用率的话,则会出现需要优化的目标函数和实际优化的目标函数不一致的情况。

[0051] 为了解决现有技术的问题,本实施例提供了一种基于机器异构性的联邦学习方法,通过上述方法对各边缘设备的预估梯度校准技术来实现移除各边缘设备与服务器的偏差,同时补偿本地更新次数不同导致的偏差,从而使得每次本地更新尽可能接近于全局更新,并且,模型更新的效果不受各边缘设备更新次数的限制。具体实施时,先接收服务器统一发送的初始模型、全局梯度和全局模型参数;然后根据初始模型、全局梯度和全局模型参数,得到预估梯度校准值;其中,预估梯度校准值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差以及各边缘设备因本地更新次数不同而产生的偏差;接着基于预估梯度校准值,得到目标本地梯度和目标本地模型参数;最后将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数。

[0052] 示例性方法

[0053] 本实施例提供一种基于机器异构性的联邦学习方法,该方法可以应用于人工智能的智能终端。具体如图1所示,所述方法包括:

[0054] 步骤S100、接收服务器统一发送的初始模型、全局梯度和全局模型参数;

[0055] 联邦学习本质上是一种分布式机器学习技术,或机器学习框架,其目标是在保证数据隐私安全及合法合规的基础上,实现共同建模,提升AI模型的效果。联邦学习分为横向联邦学习、纵向联邦学习和联邦迁移学习。在本发明实施例中,联邦学习采用的是横向联邦学习,横向联邦学习的本质是样本的联合,适用于参与者间业态相同但触达客户不同,即特征重叠多,用户重叠少时的场景,比如不同地区的银行间,他们的业务相似(特征相似),但用户不同(样本不同)。在横向联邦学习过程中,所有的边缘设备都是从服务器中获取同等的资源,如服务器统一发送的初始模型、全局梯度和全局模型参数,以便各边缘设备在进行本地训练时训练效率更高。在本实施例中,各边缘设备与服务器的交互过程是循环进行的,其中,所述循环的全局次数为T,T可以设置为200次,在每一次的全局循环中,在接收服务器统一发送的初始模型、全局梯度和全局模型参数之后需要将所述全局模型参数作为初始本地模型参数。

[0056] 得到服务器统一发送的初始模型、全局梯度和全局模型参数后,就可以执行如图1所示的如下步骤:步骤S200、根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值;其中,所述预估梯度校准值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差以及各边缘设备因本地更新次数不同而产生的偏差;

[0057] 具体地,由于现有技术中在处理异构联邦学习网络时,会由于不同边缘设备的算力的不同而导致联邦学习的训练效率不高的问题。如:假设有A和B两个不同的设备,A进行100次本地更新需要1小时,而B进行100次本地更新需要24小时。服务器设定的前提分为同步算法和异步算法两种情况,在现有同步算法下,为了使A设备与B设备的更新进度同步,服

务器每完成一次全局更新需要24小时,此时,A设备和B设备都完成了100次的本地更新,但是,对于A设备而言,A设备在1小时内已经完成了100次更新,就会产生23小时的空余时间,这对A设备造成了时间资源浪费。在异步算法下,由于服务器设定每个设备需进行100次本地更新后方可与服务器的参数进行交互,那么,当B设备完成这样一整轮的本地更新后,A设备已经与服务器进行了24次交互,服务器通常不会利用B设备的参数调整全局模型的参数,显然,这样会导致整个联邦学习的训练效率低下的问题。为此,本发明实施例根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值,通过调整各边缘设备中的预估梯度校准值,在同步模式下,该算法利用预估梯度校准的方式使得优化的目标函数和实际优化的目标函数一致,同时支持各边缘设备的本地模型训练更新次数不一致。相应的,所述根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值包括如下步骤:基于预设的本地数据和所述初始模型,得到本地梯度;基于所述全局梯度、所述全局模型参数和所述本地梯度,得到预估梯度校准值。

[0058] 具体地,预设的本地数据为各边缘设备根据本地用户的实际设置的训练数据集,将所述本地数据输入到所述初始模型,并对所述初始模型进行训练,得到本地梯度;然后基于得到的所述本地梯度,再加上所述全局梯度和所述全局模型参数就可以得到预估梯度校准值。相应的,所述基于所述全局梯度、所述本地梯度和所述初始本地模型参数,得到预估梯度校准值包括如下步骤:基于所述全局梯度和所述本地梯度,得到第一偏差值,其中,所述第一偏差值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差;基于所述全局模型参数和所述初始本地模型参数,得到第二偏差值,其中,所述第二偏差值用于表征各边缘设备因本地更新次数不同而产生的偏差;基于所述第一偏差值和所述第二偏差值,得到预估梯度校准值。

[0059] 具体地,在每一次的全局交互过程中,将所述全局梯度$\tilde{\mathbf{v}}$减去所述本地梯度v^i,得到第一偏差值;其中,所述第一偏差值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差;根据所述初始本地模型参数$x_{t;0}^i$,得到前一次全局交互时的目标本地模型参数$x_{t-1;K_i}^i$,将所述全局模型参数$\widetilde{x_{t+1}}$减去前一次全局交互时的所述目标本地模型参数$x_{t-1;K_i}^i$,得到第二偏差值,其中,所述第二偏差值用于表征各边缘设备因本地更新次数不同而产生的偏差;最后基于所述第一偏差值和所述第二偏差值,得到预估梯度校准值。相应的,为了得到预估梯度校准值,所述基于所述第一偏差值和所述第二偏差值,得到预估梯度校准值包括如下步骤:获取所述边缘设备的本地更新次数;根据所述本地更新次数和所述第二偏差值,得到中间第二偏差值;将所述第一偏差值加上所述中间第二偏差值,得到预估梯度校准值。

[0060] 具体地,对于每一个并行处理的边缘设备而言,先要获取各个边缘设备的本地更新次数K_i;同时要获取学习率的步长η,η取值可以为0.01,获取调整因子λ,其中调整因子λ的取值可以为0.1,将学习率的步长η乘以各个边缘设备的本地更新次数K_i,得到第一乘积结果,将所述第一乘积结果乘以所述第二偏差值,得到第二乘积结果,再将调整因子λ除以第二乘积结果,得到中间第二偏差值,在将所述第一偏差值加上所述中间第二偏差值,得到

得到预估梯度校准值c。例如：$c = \tilde{v} - v^i + \frac{\lambda}{\eta K_i}(\widetilde{x_{t+1}} - x_{t;K_i}^i)$。如图2左一所示，传统的算法在本地进行朴素更新，在本地更新次数一致的情况下，这种做法具备较好的更新效率；如图2左二所示，当更新不一致时，由于实际的优化目标函数与目标不一致，因而稳定点偏离了全局最优的位置，该训练方式破坏了模型的可用性，使得全局模型的可用性大打折扣。如图2左三所示，本发明实施例可以对每次本地更新的梯度进行校准，避免过分偏离全局更新的方向。

[0061] 得到预估梯度校准值之后，就可以执行如图1所示的如下步骤：步骤S300、基于所述预估梯度校准值，得到目标本地梯度和目标本地模型参数；

[0062] 由于在联邦学习的异构网络中，所述联邦学习训练效率的低下主要是由于各边缘设备的算力不同，导致在相同的时间内各边缘设备的本地更新次数不同，进而影响异构网络下各边缘设备的本地更新梯度与全局更新的梯度存在很大的差异，使得训练出来的模型不准确，降低了联邦学习的模型训练效率，故采用预估梯度校准的方式。根据各个边缘设备的所述预估梯度校准值，得到目标本地梯度和目标本地模型参数，使得各个边缘设备的本地更新更贴近全局更新。相应的，为了得到目标本地梯度和目标本地模型参数，所述基于所述预估梯度校准值，得到目标本地梯度和目标本地模型参数包括如下步骤：获取数据样本；其中，所述数据样本是从边缘设备的样本中获取得到；根据所述数据样本和所述初始本地模型参数，得到损失函数偏导数；根据所述损失函数偏导数，得到目标本地梯度；根据所述损失函数偏导数、所述初始本地模型参数和所述预估梯度校准值，得到目标本地模型参数。

[0063] 具体地，从第i台设备的样本D_i中获取一个随机数据样本ε_k^i。在各个边缘设备进行模型训练的时候，各边缘设备会进行K_i次的本地更新，然后根据所述数据样本和所述初始本地模型参数，得到损失函数偏导数。相应的，为了得到损失函数偏导数，所述根据所述数据样本和所述初始本地模型参数，得到损失函数偏导数包括如下步骤：根据所述数据样本和所述初始本地模型参数，得到损失函数；对所述损失函数进行求偏导，得到损失函数偏导数。

[0064] 具体地，在k＝0时，是将全局模型参数赋值给初始本地模型参数$x_{t;0}^i$，在k从1到K_i-1次的循环更新过程中，各边缘设备更新过程中的本地模型参数是$x_{t;k}^i$，然后根据所述随机样本ε_k^i和各边缘设备更新过程中的本地模型参数$x_{t;k}^i$，得到损失函数$f_i(x_{t;k}^i, \varepsilon_k^i)$，然后对所述损失函数进行求偏导，得到损失函数偏导数$g_{t,k}^i = f_i(x_{t;k}^i, \varepsilon_k^i)$。当k＝0时，将$g_{t,0}^i$赋值给$v^i$，得到目标本地梯度$v^i$，这样，各边缘设备的起始端点一样，使得服务器更容易将各个边缘设备的目标本地梯度v^i聚合为跟全局梯度一样。此外，在k从0到K_i-1次的循环更新过程中，先将损失函数偏导数$g_{t,k}^i$加上所述预估梯度校准值c，得到补偿损失函数偏导数，然后将所述补偿损失函数偏导数乘以学习率步长η，得到目标损失函数偏导数$l = \eta * (g_{t,k}^i + c)$，再将各边缘设备更新过程中的本地模型参数$x_{t;k}^i$减去所述目标损失函数偏导数l，得到各边缘

设备更新过程中的后一次本地模型参数$x^{i}_{t;k+1}$，当各边缘设备更新过程结束时，也即当k＝K_i-1时，得到目标本地模型参数$x^{i}_{t;K_i}$。

[0065] 得到目标本地梯度和目标本地模型参数之后，执行如图1所示的如下步骤：S400、将所述目标本地梯度和所述目标本地模型参数发送至所述服务器，以使所述服务器生成更新后的全局梯度和全局模型参数。

[0066] 具体地，各个边缘设备将各自得到的所述目标本地梯度和所述目标本地模型参数都发送到服务器中，服务器接收边缘设备的所述目标本地梯度和所述目标本地模型参数，并对各边缘设备的所述目标本地梯度和所述目标本地模型参数进行整合，在本实施例中，各边缘设备与服务器的交互过程是循环进行的，其中，所述循环的全局次数为T，T可以设置为200次。可以理解的是，在联邦学习之初，是各边缘设备接收初始的全局梯度的值和初始的全局模型参数的值为0。而在各边缘设备进行了一次模型训练，并且所述服务器进行了聚合操作之后，所述全局梯度和所述全局模型参数的值不再是0，相应的，在第2次到第T次时，由于所述服务器与各边缘设备都有交互，在各边缘设备进行了一次模型训练后，在进行2-T次的全局交互过程中，各边缘设备会得到每一次全局交互后的目标本地梯度和目标本地模型参数；此时，所述服务器会接收到各边缘设备发送的所述目标本地梯度和所述目标本地模型参数，然后对所述目标本地梯度进行加权求和计算，得到更新后的全局梯度，例如：$\tilde{v}=\sum_{i=1}^{M} w_i v^i$，在上述公式中，$v^i$代表的是第i个边缘设备的目标本地梯度，$w_i$为第i个边缘设备相应的权重，将所有边缘设备的目标本地梯度乘以相应的权重后再进行累加，就可以得到更新后的全局梯度$\tilde{v}$。基于同样的道理，对所述目标本地模型参数进行加权求和计算，得到更新后的全局模型参数，例如：$\widetilde{x_{t+1}}=\sum_{i=1}^{M} w_i x^i_{t;K_i}$，上述公式中，$x^i_{t;K_i}$为在进行1-T次的全局交互过程中，得到的所述目标本地模型参数，w_i为第i个边缘设备相应的权重，也即将每次全局交互过程中得到的所述目标本地模型参数乘以第i个边缘设备相应的权重，得到每一个边缘设备的权重目标本地模型参数，然后将所有边缘设备的权重目标本地模型参数累加，就可以得到更新后的全局模型参数$\widetilde{x_{t+1}}$。

[0067] 此外，所述将所述目标本地梯度和所述目标本地模型参数发送至所述服务器，以使所述服务器生成更新后的全局梯度和全局模型参数之后包括：

[0068] 接收服务器统一发送的更新后的全局梯度和全局模型参数，并重复执行根据所述初始模型、所述全局梯度和所述全局模型参数，得到预估梯度校准值的步骤。

[0069] 具体地，在服务器和各边缘设备的第2次到第T次交互过程中，边缘设备都会接收服务器统一发送的更新后的全局梯度和全局模型参数，然后根据之前的初始模型以及更新后的全局梯度和全局模型参数继续进行运算，得到预估梯度校准值，再根据所述预估梯度校准值，得到更新后的目标本地梯度和目标本地模型参数；将更新后的目标本地梯度和目标本地模型参数发送至所述服务器的步骤。在服务器和各边缘设备的第2次到第T次交互过程中，对于边缘服务器而言，服务器会再次将更新后目标本地梯度和所述目标本地模型参数发送至所述服务器。如此循环往复T-1次。

[0070] 示例性设备

[0071] 如图3中所示,本发明实施例提供一种基于机器异构性的联邦学习装置,该装置包括服务器的数据接收单元501,预估梯度校准值获取单元502,目标参数获取单元503,目标参数发送单元504,其中:

[0072] 服务器的数据接收单元501,用于接收服务器统一发送的初始模型、全局梯度和全局模型参数;

[0073] 预估梯度校准值获取单元502,用于根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值;其中,所述预估梯度校准值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差以及各边缘设备因本地更新次数不同而产生的偏差;

[0074] 目标参数获取单元503,用于基于所述预估梯度校准值,得到目标本地梯度和目标本地模型参数;

[0075] 目标参数发送单元504,用于将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数。

[0076] 基于上述实施例,本发明还提供了一种智能终端,其原理框图可以如图4所示。该智能终端包括通过系统总线连接的处理器、存储器、网络接口、显示屏、温度传感器。其中,该智能终端的处理器用于提供计算和控制能力。该智能终端的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该智能终端的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种基于机器异构性的联邦学习方法。该智能终端的显示屏可以是液晶显示屏或者电子墨水显示屏,该智能终端的温度传感器是预先在智能终端内部设置,用于检测内部设备的运行温度。

[0077] 本领域技术人员可以理解,图4中的原理图,仅仅是与本发明方案相关的部分结构的框图,并不构成对本发明方案所应用于其上的智能终端的限定,具体的智能终端可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0078] 在一个实施例中,提供了一种智能终端,包括有存储器,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行以下操作的指令:

[0079] 接收服务器统一发送的初始模型、全局梯度和全局模型参数;

[0080] 根据所述初始模型、所述全局梯度和所述全局模型参数,得到预估梯度校准值;其中,所述预估梯度校准值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差以及各边缘设备因本地更新次数不同而产生的偏差;

[0081] 基于所述预估梯度校准值,得到目标本地梯度和目标本地模型参数;

[0082] 将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数。

[0083] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本发明所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括

随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM以多种形式可得,诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0084] 综上所述,本发明公开了基于机器异构性的联邦学习方法、智能终端、存储介质,所述方法包括:接收服务器统一发送的初始模型、全局梯度和全局模型参数;根据初始模型、全局梯度和全局模型参数,得到预估梯度校准值;其中,预估梯度校准值用于表征各边缘设备的本地梯度与服务器的全局梯度的偏差以及各边缘设备因本地更新次数不同而产生的偏差;基于预估梯度校准值,得到目标本地梯度和目标本地模型参数;将所述目标本地梯度和所述目标本地模型参数发送至所述服务器,以使所述服务器生成更新后的全局梯度和全局模型参数。本发明实施例通过对各边缘设备的预估梯度校准技术来实现移除各边缘设备与服务器的偏差,同时补偿本地更新次数不同导致的偏差,从而使得每次本地更新尽可能接近于全局更新,并且,模型更新的效果不受各边缘设备更新次数的限制。

[0085] 基于上述实施例,本发明公开了一种基于机器异构性的联邦学习方法,应当理解的是,本发明的应用不限于上述的举例,对本领域普通技术人员来说,可以根据上述说明加以改进或变换,所有这些改进和变换都应属于本发明所附权利要求的保护范围。

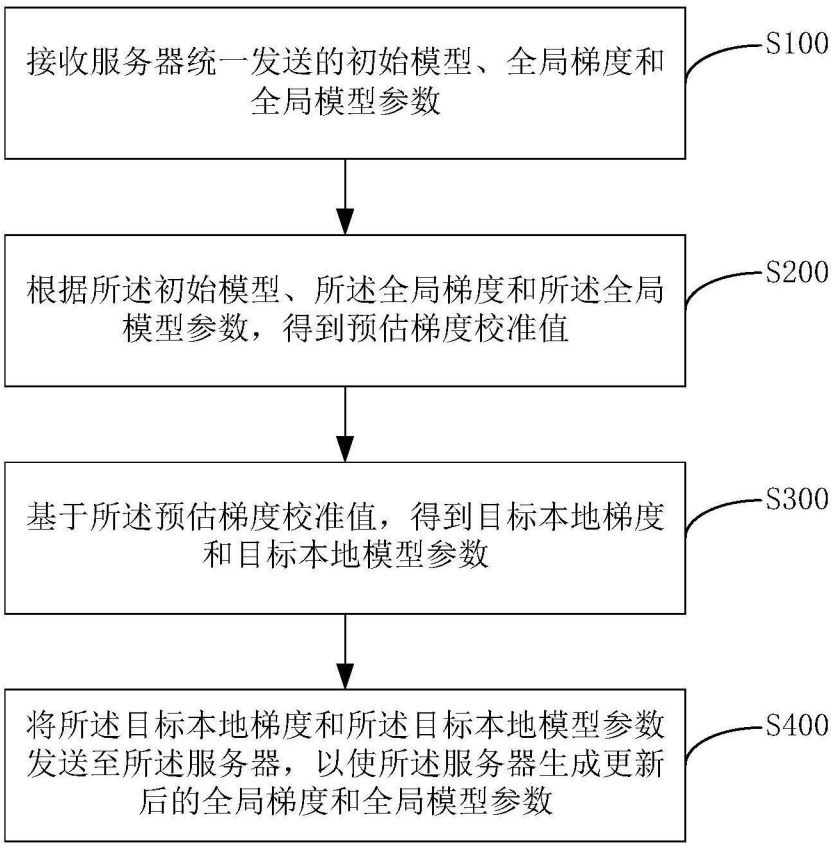

接收服务器统一发送的初始模型、全局梯度和全局模型参数
S100
根据所述初始模型、所述全局梯度和所述全局模型参数，得到预估梯度校准值
S200
基于所述预估梯度校准值，得到目标本地梯度和目标本地模型参数
S300
将所述目标本地梯度和所述目标本地模型参数发送至所述服务器，以使所述服务器生成更新后的全局梯度和全局模型参数
S400

图1

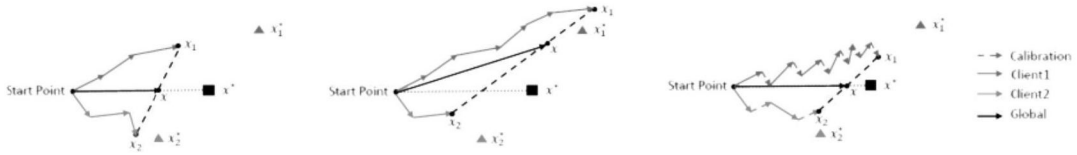

Start Point
Start Point
Start Point
Calibration
Client1
Client2
Global

图2

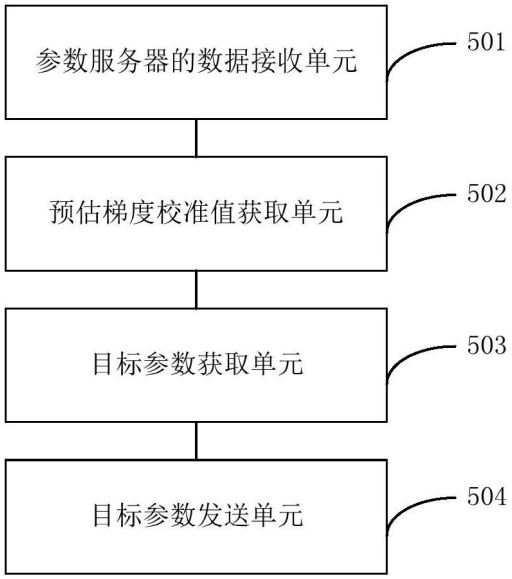

参数服务器的数据接收单元
501
预估梯度校准值获取单元
502
目标参数获取单元
503
目标参数发送单元
504

图3

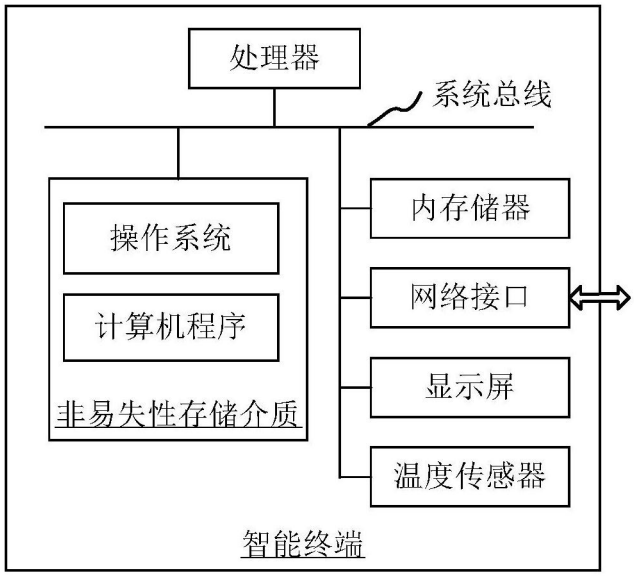

处理器
系统总线
操作系统
计算机程序
非易失性存储介质
内存储器
网络接口
显示屏
温度传感器
智能终端

图4