

一种联邦学习方法以及装置

本申请要求于 2021 年 04 月 25 日提交中国专利局、申请号为“202110450585.7”、申请名称为“一种联邦学习方法以及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及联邦学习领域，尤其涉及一种联邦学习方法以及装置。

背景技术

在用户隐私日益觉醒的今天，各方对于数据所有权的关注越来越多，各国对于数据安全也日趋严格，要解决数据孤岛的困境，仅仅靠传统的机器学习训练模式已经出现瓶颈。为了帮助公司充分利用用户的数据信息而不侵犯隐私，实现多方联合建模，以帮助企业应用提效降本，助力不同行业智慧升级，诞生了联邦学习（federated learning, FL）。联邦学习是一种人工智能（artificial intelligence, AI）联合建模的组织方式，在不直接获取设备私有数据的基础上，通过协调跨设备的模型训练和参数聚合完成算法模型任务，同时降低传统集中式机器学习的通信计算开销，具有重要理论创新和实践落地价值。

然而，在由于跨设备终端部署场景的特殊性，跨设备（cross-device）联邦学习还带来了新的挑战，如联邦学习的客户端设备异构，不同设备在存储、计算、通信和电池等硬件规格方面存在很大差异，对联邦学习性能产生影响。因此，如何提高跨设备的联邦学习的学习效率，成为亟待解决的问题。

发明内容

本申请提供一种联邦学习方法，用于在跨设备的联邦学习过程中，对各个服务端之间的聚合信息进行同步，从而使每轮迭代学习过程中服务端的数据保持同步，客户端在任何时刻接入服务端都可以获取到全量数据，提高整体的学习效率。

有鉴于此，第一方面，本申请提供一种联邦学习方法，应用于联邦学习系统，联邦学习系统中包括多个服务端，多个服务端之间互相连接，多个服务端用于进行迭代联邦学习，其中任意一轮迭代学习的过程包括：首先，第一服务端接收至少一个第一客户端发送的请求消息，请求消息用于请求第一服务端中存储的全局模型，前述的多个客户端包括该至少一个第一客户端；第一服务端向至少一个第一客户端发送全局模型的信息和训练配置参数，全局模型的信息和训练配置参数用于指示至少一个第一客户端使用训练配置参数对全局模型进行训练；第一服务端接收至少一个第一客户端分别反馈的第一模型更新参数，第一模型更新参数是至少一个第一客户端训练后得到的全局模型的参数；第一服务端对至少一个第一客户端反馈的第一模型更新参数进行聚合，以得到在本轮迭代中的第一聚合信息；第一服务端获取第二服务端发送的第二聚合信息，第二聚合信息为第二服务端在本轮迭代中，对接收到的第二模型更新参数进行聚合得到的信息；第一服务端基于第一聚合信息和第二聚合信息，更新第一服务端上存储的全局模型，以得到更新后的全局模型。

本申请实施方式中，在进行联邦学习的每轮迭代过程中，需要由客户端主动向服务端

发送请求消息以请求参与联邦学习，并将训练全局模型后得到的模型更新参数发送至服务端，因此客户端通过向服务端发送请求消息的方式接入服务端，客户端无需和服务端保持长期稳定的连接，以实现跨设备的联邦学习系统。多个服务端除了对自身接收到的模型更新参数进行聚合，还接收其他服务端发送的模型更新参数并进行聚合，从而使每个服务端都具有比自身的聚合信息更多的数据，客户端在接入服务端时，即可获取到更多的数据，从而可以基于更多的数据来进行每轮迭代训练，提高联邦学习的整体效率。可以理解为，本申请中，客户端通过请求与响应结合的方式接入服务器，客户端和服务端之间无需保持长期稳定连接，客户端在任何时刻访问任意服务端都可以获取到训练模型所需的更多的数据，提高客户端侧的联邦学习的学习效率。

在一种可能的实施方式中，在第一服务端接收至少一个第一客户端分别反馈的第一模型更新参数之前，第一服务端可以接收多个客户端发送的请求消息，以向服务端请求参与联邦学习，第一服务端在接收到多个客户端发送的请求消息之后，即可将本地存储的全局模型的信息和训练配置参数下发给该多个客户端。因此，本申请实施例可以通过请求和响应结合的方式使客户端接入服务端，服务端无需主动对客户端进行寻址，相当于客户端可以仅在有需要时接入服务端，从而实现跨设备的联邦学习。

在一种可能的实施方式中，若第一服务端接收到多个客户端发送的请求消息，第一服务端可以对该多个客户端进行筛选，筛选出至少一个客户端来参与联邦学习，如根据客户端的连接状态或者设备状态等确定是否允许客户端参与联邦学习，从而筛选出状态稳定的客户端来进行联邦学习，提高联邦学习的整体效率。

在一种可能的实施方式中，前述的多个服务端中包括的第二服务端的数量为多个，第一服务端获取第二服务端发送的第二聚合信息，可以包括：第一服务端接收多个第二服务端分别发送的第二聚合信息；前述的第一服务端基于第一聚合信息和第二聚合信息，更新第一服务端上存储的全局模型，可以包括：第一服务端基于第一聚合信息和多个第二服务端分别发送的第二聚合信息，更新第一服务端上存储的全局模型，以得到更新后的全局模型。

本申请实施方式中，联邦学习系统内的服务端之间可以互相传输数据，使每个服务端上都具有全量数据，从而使客户端在任意时刻接入服务端，都可以获取到全量数据，提高客户端获取到的数据的准确度，且在联邦学习的过程中客户端获取到全量数据，可以提高联邦学习的整体效率。

在一种可能的实施方式中，上述方法还可以包括：前述的多个服务端之间传输第一服务端向每个第二服务端发送第一聚合信息，以使每个第二服务端对第一聚合信息和第二聚合信息进行聚合得到第三聚合信息。

因此，本申请实施方式中，联邦学习系统内的服务端之间可以互相传输聚合信息，从而使每个服务端上都具有更多数据，从而保持联邦学习系统内的数据一致性。

在一种可能的实施方式中，多个服务端中的其中一个服务端作为主节点，以第一服务端作为主节点为例，主节点用于对多个服务端进行管理，第一服务端向第二服务器发送第一触发指示，第一触发指示用于指示第二服务端向第一服务端发送第二聚合信息，以使第

一服务端接收第二服务端发送的第二聚合信息。

因此,本申请实施方式中,联邦学习系统中的联邦学习过程可以由主节点来进行管理,在主节点的触发下各个服务端之间同步数据,从而实现联邦学习系统内的数据一致性。

在一种可能的实施方式中,主节点中包括计数器,计数器用于对发送请求消息的客户端进行计数,上述的第一服务端在主节点的触发下获取第二服务端中的第二聚合信息,可以包括:当计数器的数值满足第一阈值时,则第一服务端向第二服务器发送的第二触发指示,第二触发指示用于指示第二服务器进行下一轮迭代。

本申请实施方式中,可以通过计数器来对接收到的请求消息进行计数,当计数达到一定数值时,即可触发各个服务端之间同步数据,相当于限制每轮联邦学习迭代过程中参与的客户端的数量,避免过多客户端参与而导致的长尾效应,使各个服务端可以及时地进行数据同步。

在一种可能的实施方式中,第一阈值包括预设值,或者,该第一阈值与上一轮迭代过程中接入联邦学习系统内所有服务端的客户端的数量确定的值。因此,本申请实施方式中,计数器对应的第一阈值可以是预先设定的值,也可以根据接入联邦学习系统内所有服务端的客户端的数量确定的值,通常该第一阈值不大于接入联邦学习系统的客户端的数量,从而避免等待客户端的接入而导致联邦学习的等待时间过长,提高联邦学习的整体效率。

在一种可能的实施方式中,主节点中还设置有计时器,上述方法还可以包括:当计时器超过第二阈值时,即计时器超时时,第一服务端向每个服务端发送第二触发指示,该第二触发指示用于触发各个服务端进入下一轮迭代。因此,本申请实施方式中,可以通过计时器来设置每轮迭代的时间窗口,当计时器超时时即可进入下一轮迭代学习,从而避免因等待时长过长而导致的长尾效应,即使客户端在进行联邦学习时中途退出也不影响联邦学习的整体训练效率。

在一种可能的实施方式中,第二阈值为预设值、根据接入联邦学习系统的客户端的数量确定的值或者联邦学习系统内的所有服务端和客户端之间进行通信的数据量的大小确定的值。

因此,本申请实施方式中,可以根据客户端的数量或者通信量来确定时间窗口,使计时器对应的时间窗口与实际场景匹配,更合理地设置每轮迭代的时间窗口。

在一种可能的实施方式中,上述方法还可以包括:第一服务端接收第三客户端发送的查询消息,第三客户端是多个客户端中的任意一个;第一服务端针对查询消息向第三客户端发送更新后的全局模型的信息。

因此,本申请实施方式中,客户端可以向服务端发送查询消息以查询最新的全局模型。因每轮迭代学习过程中各个服务端之间进行了数据同步,因此每个服务端上都具有比自身的聚合信息更多的数据,客户端无论接入哪个服务端都可以获取到更多的数据,使客户端获取到更准确的全局模型。

第二方面,本申请提供一种联邦学习方法,应用于联邦学习系统,联邦学习系统包括多个服务端,多个服务端中的其中一个服务端作为主节点,多个服务端用于进行迭代学习以实现联邦学习;其中迭代学习中的任意一轮迭代的过程包括:

在多个服务端中任意一个服务端接收到第一个请求消息之后，主节点启动计数器和计时器，计数器用于对多个服务端在一轮迭代中接收到的请求消息进行计数，请求消息用于请求获取多个服务端中对应的服务端中存储的全局模型；若计数器的值达到第一阈值，主节点向多个服务端中的每个服务端发送第一触发指示，第一触发指示用于指示多个服务端之间互相传输本地存储的信息；若计数器的值未达到第一阈值，计时器的值达到第二阈值，则主节点向每个服务端发送第二触发指示，第二触发指示用于指示每个服务端进行下一轮迭代。

本申请实施方式中，将联邦学习系统中的其中一个服务端作为主节点，由主节点来维护计时器和计数器，在每轮迭代过程中，通过计数器对接收到的请求消息进行计数，并同时使用计时器来进行计时，从而限制每轮迭代参与的客户端的数量以及每轮迭代的训练时长，从而避免因客户端掉队而导致的无意义的等待，提高联邦学习的整体效率，避免出现长尾效应。

在一种可能的实施方式中，第一阈值为预设值，或者，第一阈值与上一轮迭代过程中接入联邦学习系统的客户端的数量相关。因此，本申请实施方式中，计数器对应的第一阈值可以是预先设定的值，也可以根据接入联邦学习系统内所有服务端的客户端的数量确定的值，通常该第一阈值不大于接入联邦学习系统的客户端的数量，从而避免等待客户端的接入而导致联邦学习的等待时间过长，提高联邦学习的整体效率。

在一种可能的实施方式中，第二阈值为预设值，或者，第二阈值与上一轮迭代中接入联邦学习系统的客户端的数量相关或者第二阈值与上一轮迭代中联邦学习系统内的每个服务端和对应的客户端之间进行通信的数据量的大小相关。因此，本申请实施方式中，可以通过计时器来设置每轮迭代的时间窗口，当计时器超时即可进入下一轮迭代学习，从而避免因等待时长过长而导致的长尾效应，即使客户端在进行联邦学习时中途退出也不影响联邦学习的整体训练效率。

第三方面，本申请提供一种联邦学习系统，包括多个服务端和多个客户端，多个服务端包括第一服务端和第二服务端，第一服务端和第二服务端上均存储了全局模型的信息，多个服务端用于进行迭代学习以实现联邦学习，其中迭代学习中的任意一轮迭代过程中：

第一服务端用于接收至少一个第一客户端分别发送的请求消息；

第一服务端用于针对至少一个第一客户端分别发送的请求消息向至少一个第一客户端发送全局模型的信息和训练配置参数，全局模型的信息和训练配置参数用于指示至少一个第一客户端使用训练配置参数对全局模型进行训练；

第一服务端用于接收至少一个第一客户端分别反馈的第一模型更新参数，第一模型更新参数为至少一个第一客户端训练后得到的全局模型的参数；

第一服务端用于对至少一个第一客户端分别反馈的第一模型更新参数进行聚合，得到第一聚合信息；

第二服务端用于接收对应的至少一个第二客户端发送的第二模型更新参数，并对至少一个第二客户端发送的第二模型更新参数进行聚合，以得到第二聚合信息；

第一服务端用于接收每个第二服务端发送的第二聚合信息；

第一服务端用于基于第一聚合信息和每个第二服务端发送的第二聚合信息，更新第一服务端上存储的全局模型，以得到更新后的全局模型。

本申请实施方式提供的联邦学习系统中，在进行每轮迭代学习时，客户端可以主动向服务端发送请求来请求参与联邦学习，并将训练的全局模型的参数反馈给服务端，服务端对接收到的模型更新参数进行聚合，从而得到聚合信息，并且，服务端还接收其他服务端发送的聚合信息，从而使服务端具有比自身的聚合信息更多的信息。使客户端和服务端无需保持长连接，客户端在任意时刻接入服务端都可以获取到更多的数据，可以提高客户端获取到的数据的准确度，从而可以基于更多的数据来进行每轮迭代训练，提高联邦学习的整体效率。

在一种可能的实施方式中，多个服务端中包括的第二服务端的数量为多个；多个第二服务端中的每个第二服务端用于接收对应的至少一个第二客户端发送的第二模型更新参数，并对至少一个第二客户端发送的第二模型更新参数进行聚合，以得到第二聚合信息；第一服务端具体用于接收多个第二服务端分别发送的第二聚合信息；第一服务端具体用于基于第一聚合信息和多个第二服务端分别发送的第二聚合信息，更新第一服务端上存储的全局模型，以得到更新后的全局模型。

本申请实施方式中，联邦学习系统内的服务端之间可以互相传输数据，使每个服务端上都具有全量数据，从而使客户端在任意时刻接入服务端，都可以获取到全量数据，提高客户端获取到的数据的准确度，且在联邦学习的过程中客户端获取到全量数据，可以提高联邦学习的整体效率。

在一种可能的实施方式中，多个服务端中还包括作为主节点的第三服务端，主节点用于对多个服务端进行管理；主节点用于向多个服务端分别发送第一触发指示；第二服务器具体用于基于第一触发指示向第一服务器发送第二聚合信息。

因此，本申请实施方式中，可以由主节点来触发服务端之间互相传输聚合信息，从而使每个服务端都可以具有更全面的全量数据，从而使客户端在任意时刻接入服务端，都可以获取到全量数据，提高客户端获取到的数据的准确度，且在联邦学习的过程中客户端获取到全量数据，可以提高联邦学习的整体效率。

在一种可能的实施方式中，主节点中包括计数器，计数器用于对多个服务端接收到的请求消息的数量进行计数，请求消息用于请求获取多个服务端中对应的服务端中存储的全局模型；主节点具体用于当计数器的数值满足第一阈值时，向多个服务端中的每个服务端发送第一触发指示，第一触发指示用于触发每个第二服务器向第一服务器发送第二聚合信息。

本申请实施方式中，可以通过计数器来对接收到的请求消息进行计数，当计数达到一定数值时，即可触发各个服务端之间同步数据，相当于限制每轮联邦学习迭代过程中参与的客户端的数量，避免过多客户端参与而导致的长尾效应，使各个服务端可以及时地进行数据同步。

在一种可能的实施方式中，主节点还包括计时器，计时器从每一轮迭代过程中接收到的第一个请求消息开始计时；主节点还用于当计时器超过第二阈值时，向多个服务端中的

每个服务端分别发送第二触发指示，第二触发指示用于指示多个服务端进行下一轮迭代。

因此，本申请实施方式中，可以通过计时器来设置每轮迭代的时间窗口，当计时器超时时即可进入下一轮迭代学习，从而避免因等待时长过长而导致的长尾效应，即使客户端在进行联邦学习时中途退出也不影响联邦学习的整体训练效率。

5 在一种可能的实施方式中，第二阈值为预设值，或者，第二阈值与上一轮迭代过程中接入联邦学习系统中每个服务端的客户端的数量相关，或者，第二阈值与上一轮迭代过程中多个服务端和多个客户端之间进行通信的数据量的大小相关。因此，本申请实施方式中，可以根据客户端的数量或者通信量来确定时间窗口，使计时器对应的时间窗口与实际场景匹配，更合理地设置每轮迭代的时间窗口。

10 在一种可能的实施方式中，第一服务端接收第三客户端发送的查询消息，第三客户端包括接入联邦学习系统的任意一个客户端；第一服务端针对查询消息向第三客户端发送更新后的全局模型的信息。因此，本申请实施方式中，客户端可以向服务端发送查询消息以查询最新的全局模型。因每轮迭代学习过程中各个服务端之间进行了数据同步，因此每个服务端上都具有比自身的聚合信息更多的数据，客户端无论接入哪个服务端都可以获取到
15 更多的数据，使客户端获取到更准确的全局模型。

 在一种可能的实施方式中，第一服务端还用于向第二服务端发送第一聚合信息；第二服务器端具体用于结合第一聚合信息和第二聚合信息，更新本地存储的全局模型，以得到更新后的全局模型。因此，本申请实施方式中，联邦学习系统内的服务端之间可以互相传输聚合信息，从而使每个服务端上都具有更多数据，从而保持联邦学习系统内的数据一致性。
20 性。

 第四方面，本申请提供一种服务端，应用于联邦学习系统，联邦学习系统中包括多个服务端和多个客户端，多个服务端用于进行迭代联邦学习，多个服务端中的任意一个服务端包括：

 收发模块，用于接收至少一个第一客户端发送的请求消息，请求消息用于请求第一服务器中存储的全局模型，多个客户端包括至少一个第一客户端；
25 收发模块，还用于向至少一个第一客户端发送全局模型的信息和训练配置参数，全局模型的信息和训练配置参数用于指示至少一个第一客户端使用训练配置参数对全局模型进行训练；

 收发模块，还用于接收至少一个第一客户端分别反馈的第一模型更新参数，第一模型更新参数是至少一个第一客户端训练后得到的全局模型的参数；
30 聚合模块，用于对至少一个第一客户端反馈的第一模型更新参数进行聚合，以得到在本轮迭代中的第一聚合信息；

 收发模块，还用于获取第二服务端发送的第二聚合信息，第二聚合信息为第二服务端在本轮迭代中，对接收到的第二模型更新参数进行聚合得到的信息；
35 更新模块，用于基于第一聚合信息和第二聚合信息，更新第一服务端上存储的全局模型，以得到更新后的全局模型。

 在一种可能的实施方式中，收发模块，还用于接收多个第二服务端分别发送的第二聚

合信息；

更新模块，具体用于基于第一聚合信息和多个第二服务端分别发送的第二聚合信息，更新第一服务端上存储的全局模型，以得到更新后的全局模型。

在一种可能的实施方式中，第一服务器为联邦学习系统中的主节点，主节点用于对多个服务端进行管理，

收发模块，还用于向第二服务器发送第一触发指示，第一触发指示用于指示第二服务端向第一服务端发送第二聚合信息；

收发模块，还用于接收第二服务端中的第二聚合信息。

在一种可能的实施方式中，第一服务器中设置有计数器，计数器用于对多个服务端接收到的请求消息的数量进行计数，

收发模块，具体用于当计数器的数值满足第一阈值时，向第二服务端发送第一触发指示。

在一种可能的实施方式中，第一阈值为预设值，或者，第一阈值与上一轮迭代过程中接入联邦学习系统的客户端的数量相关。

在一种可能的实施方式中，主节点中包括计时器，计时器从每一轮迭代中接收到第一个请求消息之后开始计时，

收发模块，具体用于当计时器超过第二阈值时，接收主节点发送的第二触发指示，第二触发指示用于指示进行下一轮迭代学习。

在一种可能的实施方式中，第二阈值为预设值，或者，第二阈值与上一轮迭代中接入联邦学习系统的客户端的数量相关，或者，第二阈值与上一轮迭代中联邦学习系统内的每个服务端和对应的客户端之间进行通信的数据量的大小相关。

在一种可能的实施方式中，收发模块，还用于接收第三客户端发送的查询消息，第三客户端包括接入联邦学习系统中的客户端中的任意一个；

收发模块，还用于针对查询消息向第三客户端发送更新后的全局模型的信息。

在一种可能的实施方式中，收发模块，还用于向第二服务端发送第一聚合信息，以使第二服务端基于第一聚合信息和第二聚合信息，更新本地存储的全局模型以得到更新后的全局模型。

第五方面，本申请提供一种服务端，应用于联邦学习系统，联邦学习系统包括多个服务端，多个服务端中的其中一个服务端作为主节点，多个服务端用于进行迭代学习以实现联邦学习，主节点包括：

启动模块，用于在多个服务端中任意一个服务端接收到第一个请求消息之后，主节点启动计数器和计时器，计数器用于对多个服务端在一轮迭代中接收到的请求消息进行计数，请求消息用于请求获取多个服务端中对应的服务端中存储的全局模型；

收发模块，用于若计数器的值达到第一阈值，主节点向多个服务端中的每个服务端发送第一触发指示，第一触发指示用于指示多个服务端之间互相传输本地存储的信息；

收发模块，还用于若计数器的值未达到第一阈值，计时器的值达到第二阈值，则主节点向每个服务端发送第二触发指示，第二触发指示用于指示每个服务端进行下一轮迭代。

在一种可能的实施方式中，第一阈值为预设值，或者，第一阈值与上一轮迭代过程中接入联邦学习系统的客户端的数量相关。

在一种可能的实施方式中，第二阈值为预设值，或者，第二阈值与上一轮迭代中接入联邦学习系统的客户端的数量相关或者第二阈值与上一轮迭代中联邦学习系统内的每个服务端和对应的客户端之间进行通信的数据量的大小相关。第六方面，本申请实施例提供一种联邦学习装置，包括：处理器和存储器，其中，处理器和存储器通过线路互联，处理器调用存储器中的程序代码用于执行上述第一方面或者第二方面任一项所示的用于联邦学习方法中与处理相关的功能。可选地，该联邦学习装置可以是芯片。

第七方面，本申请实施例提供了一种联邦学习装置，该联邦学习装置也可以称为数字处理芯片或者芯片，芯片包括处理单元和通信接口，处理单元通过通信接口获取程序指令，程序指令被处理单元执行，处理单元用于执行如上述第一方面或第二方面中任一可选实施方式中与处理相关的功能。

第八方面，本申请实施例提供了一种计算机可读存储介质，包括指令，当其在计算机上运行时，使得计算机执行上述第一方面或第二方面中任一可选实施方式中的方法。

第九方面，本申请实施例提供了一种包含指令的计算机程序产品，当其在计算机上运行时，使得计算机执行上述第一方面或第二方面中任一可选实施方式中的方法。

附图说明

- 图 1 为本申请应用的一种人工智能主体框架示意图；
- 图 2 为本申请实施例提供的一种联邦学习系统的架构示意图；
- 图 3 为本申请实施例提供的一种服务器的结构示意图；
- 图 4 为本申请实施例提供的一种联邦学习方法的流程示意图；
- 图 5 为本申请实施例提供的另一种联邦学习系统的架构示意图；
- 图 6 为本申请实施例提供的另一种联邦学习方法的流程示意图；
- 图 7 为本申请实施例提供的另一种联邦学习方法的流程示意图；
- 图 8 为本申请实施例提供的一种聚合方式示意图；
- 图 9 为本申请实施例提供的一种服务器的结构示意图；
- 图 10 为本申请实施例提供的另一种服务器的结构示意图；
- 图 11 为本申请实施例提供的另一种服务器的结构示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

首先对人工智能系统总体工作流程进行描述，请参见图 1，图 1 示出的为人工智能主体框架的一种结构示意图，下面从“智能信息链”（水平轴）和“IT 价值链”（垂直轴）两个

维度对上述人工智能主题框架进行阐述。其中，“智能信息链”反映从数据的获取到处理的一系列过程。举例来说，可以是智能信息感知、智能信息表示与形成、智能推理、智能决策、智能执行与输出的一般过程。在这个过程中，数据经历了“数据—信息—知识—智慧”的凝练过程。“IT 价值链”从人智能的底层基础设施、信息（提供和处理技术实现）到系统的产业生态过程，反映人工智能为信息技术产业带来的价值。

(1) 基础设施

基础设施为人工智能系统提供计算能力支持，实现与外部世界的沟通，并通过基础平台实现支撑。通过传感器与外部沟通；计算能力由智能芯片，如中央处理器（central processing unit, CPU）、网络处理器（neural-network processing unit, NPU）、图形处理器（英语：graphics processing unit, GPU）、专用集成电路（application specific integrated circuit, ASIC）或现场可编程逻辑门阵列（field programmable gate array, FPGA）等硬件加速芯片）提供；基础平台包括分布式计算框架及网络等相关的平台保障和支持，可以包括云存储和计算、互联互通网络等。举例来说，传感器和外部沟通获取数据，这些数据提供给基础平台提供的分布式计算系统中的智能芯片进行计算。

(2) 数据

基础设施的上一层的数据用于表示人工智能领域的数据来源。数据涉及到图形、图像、语音、文本，还涉及到传统设备的物联网数据，包括已有系统的业务数据以及力、位移、液位、温度、湿度等感知数据。

(3) 数据处理

数据处理通常包括数据训练，机器学习，深度学习，搜索，推理，决策等方式。

其中，机器学习和深度学习可以对数据进行符号化和形式化的智能信息建模、抽取、预处理、训练等。

推理是指在计算机或智能系统中，模拟人类的智能推理方式，依据推理控制策略，利用形式化的信息进行机器思维和求解问题的过程，典型的功能是搜索与匹配。

决策是指智能信息经过推理后进行决策的过程，通常提供分类、排序、预测等功能。

(4) 通用能力

对数据经过上面提到的数据处理后，进一步基于数据处理的结果可以形成一些通用的能力，比如可以是算法或者一个通用系统，例如，翻译，文本的分析，计算机视觉的处理，语音识别，图像的识别等等。

(5) 智能产品及行业应用

智能产品及行业应用指人工智能系统在各领域的产品和应用，是对人工智能整体解决方案的封装，将智能信息决策产品化、实现落地应用，其应用领域主要包括：智能终端、智能交通、智能医疗、自动驾驶、平安城市等。

本申请实施例可以应用于联邦学习领域，可以对通过客户端和服务端协同对神经网络进行训练，因此涉及了大量神经网络的相关应用，如在进行联邦学习时客户端训练的神经网络。为了更好地理解本申请实施例的方案，下面先对本申请实施例可能涉及的神经网络的相关术语和概念进行介绍。

(1) 神经网络

神经网络可以是由神经元组成的，神经元可以是指以 x_s 和截距 1 为输入的运算单元，该运算单元的输出可以如公式 (1-1) 所示：

$$h_{w,b}(x) = f(W^T x) = f(\sum_{s=1}^n W_s x_s + b)$$

5 其中， $s=1, 2, \dots, n$ ， n 为大于 1 的自然数， W_s 为 x_s 的权重， b 为神经单元的偏置。 f 为神经单元的激活函数 (activation functions)，用于将非线性特性引入神经网络中，来将神经单元中的输入信号转换为输出信号。该激活函数的输出信号可以作为下一层卷积层的输入，激活函数可以是 sigmoid 函数。神经网络是将多个上述单一的神经单元联结在一起形成的网络，即一个神经单元的输出可以是另一个神经单元的输入。每个神经单元的
10 输入可以与前一层的局部接受域相连，来提取局部接受域的特征，局部接受域可以是由若干个神经单元组成的区域。

(2) 深度神经网络

深度神经网络 (deep neural network, DNN)，也称多层神经网络，可以理解为具有多层中间层的神经网络。按照不同层的位置对 DNN 进行划分，DNN 内部的神经网络可以分为
15 三类：输入层，中间层，输出层。一般来说第一层是输入层，最后一层是输出层，中间的层数都是中间层，或者称为隐层。层与层之间是全连接的，也就是说，第 i 层的任意一个神经元一定与第 $i+1$ 层的任意一个神经元相连。

虽然 DNN 看起来很复杂，其每一层可以表示为线性关系表达式： $\vec{y} = \alpha(w \vec{x} + \vec{b})$ ，其中， $\vec{x}$ 是输入向量， $\vec{y}$ 是输出向量， $\vec{b}$ 是偏移向量或者称为偏置参数， w 是权重矩阵 (也称系数)， $\alpha()$ 是激活函数。每一层仅仅是对输入向量 $\vec{x}$ 经过如此简单的操作得到输出向量 $\vec{y}$ 。由于 DNN
20 层数多，系数 w 和偏移向量 $\vec{b}$ 的数量也比较多。这些参数在 DNN 中的定义如下所述：以系数 w 为例：假设在一个三层的 DNN 中，第二层的第 4 个神经元到第三层的第 2 个神经元的线性系数定义为 W_{24}^3 。上标 3 代表系数 w 所在的层数，而下标对应的是输出的第三层索引 2 和输入的第二层索引 4。

25 综上，第 $L-1$ 层的第 k 个神经元到第 L 层的第 j 个神经元的系数定义为 W_{jk}^L 。

需要注意的是，输入层是没有 W 参数的。在深度神经网络中，更多的中间层让网络更能够刻画现实世界中的复杂情形。理论上而言，参数越多的模型复杂度越高，“容量”也就越大，也就意味着它能完成更复杂的学习任务。训练深度神经网络的也就是学习权重矩阵的过程，其最终目的是得到训练好的深度神经网络的所有层的权重矩阵 (由很多层的向量 W 形成的权重矩阵)。
30

(3) 卷积神经网络

卷积神经网络 (convolutional neuron network, CNN) 是一种带有卷积结构的深度神经网络。卷积神经网络包含了一个由卷积层和子采样层构成的特征抽取器，该特征抽取器

可以看作是滤波器。卷积层是指卷积神经网络中对输入信号进行卷积处理的神经元层。在卷积神经网络的卷积层中，一个神经元可以只与部分邻层神经元连接。一个卷积层中，通常包含若干个特征平面，每个特征平面可以由一些矩形排列的神经元组成。同一特征平面的神经元共享权重，这里共享的权重就是卷积核。共享权重可以理解为提取图像信息的方式与位置无关。卷积核可以以随机大小的矩阵的形式初始化，在卷积神经网络的训练过程中卷积核可以通过学习得到合理的权重。另外，共享权重带来的直接好处是减少卷积神经网络各层之间的连接，同时又降低了过拟合的风险。

本申请以下提及的用于提取特征的网络，即可以包括一层或者多层卷积层，示例性地，该用于提取特征的网络即可以采用 CNN 来实现。

10 (4) 损失函数

在训练深度神经网络的过程中，因为希望深度神经网络的输出尽可能的接近真正想要预测的值，所以可以通过比较当前网络的预测值和真正想要的目标值，再根据两者之间的差异情况来更新每一层神经网络的权重向量（当然，在第一次更新之前通常会有初始化的过程，即为深度神经网络中的各层预先配置参数），比如，如果网络的预测值高了，就调整权重向量让它预测低一些，不断地调整，直到深度神经网络能够预测出真正想要的目标值或与真正想要的目标值非常接近的值。因此，就需要预先定义“如何比较预测值和目标值之间的差异”，这便是损失函数（loss function）或目标函数（objective function），它们是用于衡量预测值和目标值的差异的重要方程。其中，以损失函数举例，损失函数的输出值（loss）越高表示差异越大，那么深度神经网络的训练就变成了尽可能缩小这个 loss 的过程。该损失函数通常可以包括误差平方均方、交叉熵、对数、指数等损失函数。例如，可以使用误差均方作为损失函数，定义为 $mse = \frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2$ ，具体可以根据实际应用场景选择具体的损失函数。

(5) 反向传播算法

神经网络可以采用误差反向传播（back propagation, BP）算法在训练过程中修正初始的神经网络模型中参数的大小，使得神经网络模型的重建误差损失越来越小。具体地，前向传递输入信号直至输出会产生误差损失，通过反向传播误差损失信息来更新初始的神经网络模型中参数，从而使误差损失收敛。反向传播算法是以误差损失为主导的反向传播运动，旨在得到最优的神经网络模型的参数，例如权重矩阵。

在本申请中，客户端在进行模型训练时，即可通过损失函数或者通过 BP 算法来对全局模型进行训练，得到训练后的全局模型。

(6) 联邦学习（federated learning, FL）

一种分布式机器学习算法，通过多个客户端（client 或者 FL-Client），如移动设备或边缘服务器，和服务端在数据不出域的前提下，协作式完成模型训练和算法更新，得到训练后的全局模型。如客户端使用本地存储的训练样本对全局模型进行训练并将训练后得到的模型参数反馈给服务端，由服务端对接收到的模型参数进行聚合，并基于聚合得到的信息更新本地存储的全局模型，得到最新的全局模型。

(7) 横向联邦

在两个数据集的用户特征重叠较多而用户重叠较少的情况下，把数据集按照用户维度切分，并去除双方用户特征相同而用户不完全相同的部分数据进行训练。

通常，在进行联邦学习时，需要多个服务器和客户端参与，由客户端对模型进行训练，然后上传至服务器，由服务器聚合后得到最终的模型。

5 例如，在机器学习和深度学习领域，分布式的优化已经成了一种先决条件。常用的分布式训练是参数服务器（Parameter Server，PS）模式，包括了 Worker（参数服务器训练模式中
10 进行训练的计算节点）和 PS（参数服务器）角色，训练数据由计算中心分配到各 Worker，若干 Worker 与 PS 通过网络通信，交换模型或者梯度等信息，由 PS 进行聚合以及模型优化，从而高效训练得到最终模型。在 PS 模式中，每个服务器实际上都只负责分到的部分参数（多个服务器共同维持一个全局的共享参数），而每个计算节点也只分到部分数据
15 和处理任务。然而，PS 模式虽然解决了大规模分布式训练的问题，但是其每个服务器（server）只能处理部分参数的计算，并且要求参与方和中心服务器保持稳定长连接，在 Cross-Device 场景中，FL-Client 大多是无状态，不可靠，且不能寻址的，而且中心服务器需要聚合全量的参数并完成聚合（aggregation）计算，PS 模式训练模式显然也处理不了跨设备终端的挑战。

又例如，在已有的一些联邦学习框架中，如 TensorFlow Federated（TFF）框架或者 PaddleFL 框架，TFF 聚焦于提供联邦学习的仿真接口，支持在本地模拟任意数量的 FL-Client 和 FL-Server，但是其框架本身不存在跨网络通信，也就无法在真实生产环境中使用。PaddleFL 在提供了一种基于数据并行的分布式联邦学习的框架，包含有 FL-Server，
20 FL-Trainer，FL-Scheduler 三种角色，但是 FL-Client 和 FL-Server 通过远程过程调用系统（remote procedure calls，RPC）进行跨进程通信，两者存在的稳定的长连接，并且 FL-Scheduler 可以调度 Worker，在每个更新周期前，决定哪些 Worker 可以参与训练。这在跨设备的部署场景中是无法实现的。

因此，本申请提供一种联邦学习架构，通过更松散的接入方式，使客户端接入服务器，
25 各个服务器之间保持数据一致性，使客户端在任何时候接入服务器都可以获取到全量数据，提高客户端的模型训练效率。

本申请提供的联邦学习系统可以如图 2 所述，该系统（或者也可以简称为集群）中可以包括多个服务端，该多个服务端之间互相建立连接，每个服务端可以和一个或者多个客户端通信。

30 其中，服务端通常可以部署于服务器中，在一些场景中，服务端也可以部署于终端设备中，具体可以根据实际应用场景调整，本申请对此不作限定。客户端可以部署于终端中，如手机、手环、智能电视等终端中，也可以部署于服务器中，具体可以根据实际应用场景调整，本申请对此也不作限定。

该多个服务端中的其中一个服务端可以作为主节点，通过主节点来使集群内保持事务
35 一致性，从而使各个服务端之间可以协同工作。

该框架中还可以包括调度器（图中未示出）或者弹性负载均衡（elastic load balance，ELB）（图中未示出）等，此处不再一一赘述。如调度器可以部署于任意一个服务端上，可

以用于对集群内的通信资源进行调度，如发送效率所使用的时间窗口、通道等。ELB 可以部署于任意一个服务端上，可以用于对服务端与客户端之间的通信进行路由，由 ELB 根据各个服务端的负载确定客户端接入的服务端等，如 ELB 可以将发送请求消息的客户端接入当前负载较少的服务端中，从而使各个服务端可以保持合适的负载，避免出现某一个或者多个服务端负载过大的情况。

具体地，在进行联邦学习时，每个服务端可以向与其建立了连接的客户端下发待训练的模型，客户端可以使用本地存储的训练样本对该模型进行训练，并将训练后的模型的数据反馈至服务端，服务端在接收到一个或者多个客户端反馈的训练后的一个或者多个模型之后，即可对接收到的一个或者多个模型的数据进行聚合，得到聚合后的数据，相当于聚合后的模型。在满足停止条件之后，即可输出最终的模型，完成联邦学习。

通常，随着客户端的增加，客户端和服务端之间的通信量也增加。例如，当千万级客户端同时进行联邦学习请求时，需要大量网络带宽的支持，并且单台服务端不能承接庞大的数据和参数，因此如何得到一种高效的联邦学习系统来应对跨设备的系统异构和远程通信挑战成为一个关键问题。

示例性地，本申请提供的联邦学习系统中的服务端的结构可以参阅图 3，该服务端中可以包括训练引擎、存储模块、限时通信模块、设备管理模块、加密模块、协议解析模块或通信模块等。

通信模块，用于 FL-Server 和 FL-Client 之间的数据通信，以及 FL-Server 集群动态组网实现中心服务端资源的弹性伸缩。其中在 FL-Server 与 FL-Client 进行通信时，可以采用超文本传输协议（hypertext transfer protocol, HTTP）协议进行传输，实现稳定的短时连接。而 FL-Server 与 FL-Server 进行通信时，则可以通过远程过程调用系统（remote procedure calls, RPC）协议进行跨进程通信，实现稳定的长连接。

设备管理模块：负责联邦学习设备选择等设备相关业务，如在接收到客户端的请求消息之后，筛选该客户端是否参与联邦学习。

加解密模块：对上传下载的模型权重加密，降低通过模型逆向攻击获得原始样本的可能性。

协议解析模块：对发送的通信消息进行序列化或者对接收到的消息反序列化等。

训练引擎：对 FL-Client 上传数据进行联邦学习计算的模块，包括聚合器和优化器。聚合器可以用于对数据进行聚合，如对 FL-Client 上传的数据进行求和、加权求和等。优化器可以用于计算模型梯度，并基于计算出来的梯度对模型进行优化等，得到输出更准确的模型。

存储模块：存储联邦学习需要的元数据以及训练所得的全局最优模型等数据。

限时通信装置：包括计时器和计数器，在计时器超时或者计数器值达到阈值时进行后续的处理。

下面结合前述图 2 的联邦学习系统以及图 3 所示出的系统中的服务端的结构，对本申请提供的联邦学习方法方法的流程进行介绍。

参与图 4，下面结合前述的架构对本申请提供的联邦学习方法的流程示意图。

401、第一服务端接收至少一个客户端发送的至少一个请求消息。

该第一服务端可以是前述图 2 中所示出的多个第一服务端中的任意一个，该至少一个客户端可以是与该第一服务端建立了连接的客户端，如该至少一个客户端发送的请求消息可以由 ELB 路由至该第一服务端中。

- 5 客户端可以通过该请求消息向第一服务端请求参与联邦学习，以请求使用本地的数据进行模型训练，通过本地数据影响全局模型，得到携带客户端的个性化特征的模型。

例如，当客户端上存储的数据超过预设数据量时，即可向联邦学习系统发送请求消息，由该系统中的 ELB 将该请求消息路由到适配的第一服务端中，以通过该第一服务端和客户端协同实现联邦学习。

- 10 又例如，客户端可以采集到与其他设备不同的个性化的训练数据，客户端在采集到训练数据后，可以主动向联邦学习系统发送请求消息，以请求使用本地存储的数据来进行联邦学习，从而得到与训练数据适配的模型，并将该模型同步至系统中，完成联邦学习，从而使集群内的全局模型可以适应性地学习到各个客户端采集到的各种数据，提高全局模型的输出精度。

- 15 402、第一服务端针对请求消息向至少一个客户端分别发送全局模型以及训练配置参数。

在第一服务端接收到至少一个客户端发送的请求消息之后，若确认允许该至少一个客户端参与联邦学习，则可以向该至少一个客户端分别发送本地存储的全局模型的信息以及训练配置参数，该全局模型即待训练的模型，该全局模型的信息可以包括全局模型的结构参数（如全局模型的深度或者宽度等）或者权重参数等，该训练配置参数即对该全局模型进行训练时使用的参数，如学习率、epoch 数量或者安全算法中类别等信息。

- 20 在一种可能的方式中，若第一服务端接收到多个客户端发送的多个请求消息，则第一服务端可以对该多个客户端进行筛选，筛选出其中的至少一个客户端来参与联邦学习。例如，请求消息中可以携带终端的信息，如网络连接状态、电量、负载等信息，第一服务端可以根据终端的信息筛选出连接稳定的客户端来进行联邦学习，并向筛选出来的至少一个客户端下发模型和训练配置参数，从而使客户端可以基于该训练配置参数对该模型进行训练。

403、客户端使用训练配置参数对全局模型进行训练。

- 30 客户端（或者称为第一客户端）接收到第一服务端下发的全局模型和训练配置参数之后，即可使用训练集对该全局模型进行训练，得到训练后的全局模型。该训练集中可以包括多个训练样本，该训练样本可以是客户端采集到的数据，也可以是客户端接收到的数据，此处不作限定。

- 例如，客户端可以是部署于用户的终端上的应用程序，在用户启用客户端的过程中，客户端可以采集用户启用客户端的期间产生的数据，并将该数据作为训练样本，向服务端请求参与联邦学习，从而将采集到的数据应用到模型更新中，从而使联邦学习系统中的全局模型可以与各个客户端适配。

404、客户端向第一服务端反馈模型更新参数。

在客户端对全局模型进行训练得到训练后的模型之后，即可向第一服务端反馈训练后

的模型的信息，即模型更新参数。

具体地，该模型更新参数中可以包括模型的权重参数。例如，在第 t 次迭代联邦学习的过程中，服务端 j 接收到的模型更新参数可以表示为： $\{w_t^{k,j}, n_{k,j}\}_{k \in K}^{s_j}$ ， n_k 表示第 k 个客户端的数据集大小， $w_t^{k,j}$ 即为客户端训练后得到的模型的权重参数。

5 405、第一服务端对至少一个客户端反馈的模型更新参数进行聚合，得到第一聚合信息。

在第一服务端接收到至少一个客户端分别反馈的模型更新参数之后，即可对该至少一个客户端分别反馈的模型更新参数进行聚合，得到第一聚合信息。相当于对该至少一个客户端反馈的模型进行聚合，得到聚合后的模型的信息。

具体地，具体方式可以包括多种，如平均、加权融合等，本申请对于聚合方式不作限定，具体可以根据实际应用场景调整。

10 406、第一服务端获取第二服务端中的第二聚合信息。

其中，第二服务端包括联邦学习系统中除第一服务端之外的服务端，第二服务端的数量可以是一个或者多个。第一服务端除了可以聚合与其连接的至少一个客户端的模型更新参数并进行聚合，还可以接收到其他服务端对接收到的模型更新参数聚合后得到的第二聚合信息，从而使联邦学习系统中的各个服务端中的聚合信息可以保持一致性。并且，每个服务端都可以获取到其他服务端的聚合信息，从而使每个服务端都可以保持更多的数据，每个客户端在接入服务端时，都可以获取到更全面的数据，而无需与服务端保持长连接状态。因此，本申请提供的联邦学习系统可以适应客户端部署在移动终端的场景。

15 在一种可能的场景中，第一服务端可以接收联邦学习系统内除自身外的所有服务器发送的聚合信息，即该联邦学习系统内的各个服务器之间可以互相传输聚合信息，从而使每个服务端上都可以具有全量数据。客户端在任意时刻接入服务端，都可以获取到全量数据，从而提高客户端获取到的数据的准确性。

需要说明的是，本申请以下实施方式中，以联邦学习系统内的所有服务端互相传输聚合信息为例进行示例性说明，在实际应用场景中，也可以是部分服务端之间传输数据，或者部分服务端仅接收聚合信息而不发送聚合信息等，具体可以根据实际应用场景调整，本申请对此并不作限定。

还需要说明的是，本申请实施方式中，第二服务端或者第二聚合信息的方式与第一服务端获取第一聚合信息的方式类似，如第二服务端接收到第二客户端发送的请求消息之后，向第二客户端下发全局模型和训练配置参数，该全局模型和训练配置参数通常与第一服务器向第一客户端下发的全局模型和训练配置参数相同。然后第二客户端使用本地存储的训练数据基于该训练配置参数对全局模型进行训练，并将训练的全局模型的参数（如权重参数、模型的宽度或深度等）反馈至第二服务端，第二服务端对接收到的模型更新参数进行聚合，即可得到第二聚合信息。本申请对第二服务端获取第二聚合信息的具体过程不再赘述。

35 在一种可能的实施方式中，可以由主节点来触发集群内的各个服务之间互相传输聚合

信息，来实现集群内的数据同步，从而使集群内的服务端都可以具有更全面的数据。

该主节点可以是联邦学习系统中的其中一个服务端，如前述的第一服务端或者第二服务端等。

5 在一种可能的实施方式中，主节点中可以设置计数器或者计时器中的至少一种，用于对每一轮迭代中参与的客户端数量或者时间窗口进行监控，从而有效控制参与联邦学习的客户端的数量，以及每一轮迭代过程的时长，避免无效的等待，减少“掉尾”效应。

具体地，计数器用于对联邦学习系统内接收到的由客户端发送的请求消息进行计数，在每一轮迭代完成之后，主节点即可对计数器进行初始化。在下一轮迭代中，联邦学习系统内的任意一个服务端接收到第一个请求消息时，主节点即可启动计时器，每接收到一个请求消息即进行计数。当计数器的值达到第一阈值时，即可触发第一服务端向客户端发送模型和训练配置参数，如主节点向各个服务端发送第一触发指示，以指示各个服务端将自身的聚合信息传输给其他服务端，如指示第二服务端将第二聚合信息发送给第一服务端。若主节点为第一服务端，则主节点可以触发自身将第一聚合信息发送给其他服务端（如所有的第二服务端），还向第二服务端发送第一触发指示，以指示各个第二服务端将自身的第二聚合信息发送给其他服务端。

15 因此，可以通过计数器来对参与联邦学习的客户端进行监测，避免因客户端太多而导致的学习效率低。例如，ELB 在将每个请求消息路由至服务端之后，可以向主节点发送通知消息，以通知主节点接收到的主节点的情况，如请求消息的数量、接收时间戳等，从而使主节点对请求消息的数量进行计数。当计数器的计数超过第一阈值时，主节点向各个服务器发送第一触发指示，触发联邦学习系统内各个服务端（包括主节点本身）之间互相传输数据，从而完成服务端之间的数据同步，使每个服务端上都具有更全面的数据，客户端在接入服务端之后，即可获取到更全面的数据，从而可以基于更全面的数据进行联邦学习或者应用等，提高客户端的用户的使用体验。

20 可选地，第一阈值包括预设值，或者，与本轮迭代或者上一轮迭代接入多个服务端的客户端的数量相关等。因此，本申请实施方式中，计数器对应的第一阈值可以是预先设定的值，也可以根据接入服务端的客户端的数量确定的值，通常该第一阈值不大于接入服务端的客户端的数量，从而避免等待客户端的接入而导致联邦学习的等待时间过长，提高联邦学习的整体效率。

30 计时器可以用于在联邦学习的过程中设置时间窗口，即仅在时间窗口内进行一轮联邦学习，当计时器的计时超过第二阈值时，主节点即可触发服务端进行下一轮学习，若当前学习过程中获取到了模型更新参数，则可以丢弃该模型更新参数，当然也可以保留并具体该模型更新参数，以提高数据的利用率。因此，本申请实施方式中，通过计时器来对每一轮迭代学习的时间窗口进行监控，每一轮迭代学习可以在计时器的时间窗口内进行，从而在某些客户端掉线或者训练效率过低等时，降低这些客户端对联邦学习过程的影响，提高联邦学习的效率。

35 具体地，当联邦学习系统内的任意一个服务端接收到第一个请求消息时，主节点即可启动计时器，在计时器的时间窗口内完成本轮迭代。当计时器超时，主节点即可向联邦

学习系统中的各个服务端（包括主节点本身）发送第二触发指示，触发各个服务端进入下一轮迭代。针对主节点本身，主节点在确定计时器超时，即可直接确定进入下一轮迭代，可以无需针对自身发送第二触发指示，当然也可以针对自身生成第二触发指示，本申请对此不作限定。

5 可选地，该第二阈值可以是预先设定的值，也可以是与本轮或者上一轮接入联邦学习系统内的客户端的数量相关，或者，与本轮或者上一轮客户端和服务端之间的通信的数据量的大小相关。例如，该第二阈值可以和接入联邦学习系统内的客户端的数量确定的值或者根据客户端和服务端之间的通信的数据量的大小等呈正相关关系，如上一轮迭代中接入联邦学习系统的客户端的数量越多，该第二阈值也就越大，上一轮迭代中客户端和服务端之间的通信的数据量越大，该第二阈值也就越大等，具体可以根据实际应用场景进行调整。

10 因此，本申请实施方式中，可以为计时器设置和实际应用场景匹配的时间窗口，可以避免时间窗口过长或者过短，从而提高联邦学习的整体效率。

在一种可能的场景中，当计时器和计数器协同工作时，在每轮迭代过程中，联邦学习系统内的服务端接收到第一个请求消息时，主节点启动计时器和计数器。若计数器的数值达到第一阈值，而计时器未超时，则可以关闭计时器，在本轮迭代过程中所有的客户端反馈的模型更新参数进行聚合完成之后，即可进入下一轮迭代。若计数器的数值未达到第一阈值，而计时器超时，则可以确定结束本轮迭代，进入下一轮迭代，本轮迭代中的数据可以丢弃也可以保留。若计数器的数值未达到第一阈值，计时器也未超时，则可以继续本轮迭代。

20 因此，本申请实施方式中，可以通过计时器和计数器协同工作，对本轮迭代中参与联邦学习的客户端的数量进行限制，以及对本轮迭代的时长进行限制，从而避免因等待客户端接入或者客户端中途退出等情况而导致的长尾效应。

407、第一服务端对第一聚合信息和第二聚合信息进行聚合，得到第三聚合信息。

在第一服务端得到第一聚合信息和其他服务端发送的第二聚合信息之后，即可基于接收到的第一聚合信息和第二聚合信息更新本地存储的全局模型，即步骤 407 和步骤 408。

其中，第一服务端除了可以对自身接收到的一个或者多个客户端反馈的模型更新参数进行聚合，还可以接收其他服务端得到的聚合信息，并对自身计算得到的第一聚合信息和接收到的第二聚合信息进行再次聚合，得到数据更完整的第三聚合信息。

因此，本申请实施方式中，各个服务端不仅对自身接收到的模型更新参数进行聚合，还对联联邦学习系统内的其他服务端聚合得到的信息进行再次聚合，从而每个服务端上都可以具有联邦系统内的每个服务端中包括的数据，从而使各个服务端的数据保持一致，客户端在任意时刻接入服务端都可以获取到全量数据。

例如，在 $t+1$ 次 Iteration 时，联邦学习系统内总的客户端的数量记为 K ，服务端 j 收集到部分客户端（记为集合 S_j ）上传的权重更新信息表示为 $\{w_{t+1}^{k,j}, n_{k,j}\}_{k \in S_j}^{S_j}$ ， n_k 表示第 k 个客户端的数据集大小。每个服务端先计算出各自的 $w_{t+1}^j \leftarrow \sum_{k \in S_j} n_{k,j} w_{t+1}^{k,j}$ 和 $n_j = \sum_{k \in S_j} n_{k,j}$ ，

35 然后由主节点触发服务端之间的数据聚合，得到聚合计算后的 w_{t+1}^j 和 n_j ，各服务端都可以

得到集群维度的计算结果，即全量数据。在下一轮迭代中，客户端请求任意服务端都能得到正确的全量数据。

此外，第一服务端也可以将第一聚合信息发送给第二服务端，以使第二服务端对自身聚合得到的第二聚合信息和接收到的第一聚合信息进行再次聚合，得到第三聚合信息。通常，为保持各个服务端中的数据的数据的异质性，第二服务端在进行聚合时可以采用与第一服务端相同的聚合方式，从而最终得到相同的第三聚合信息，使每个服务端中可以具有全量数据。

408、第一服务端更新存储的全局模型。

在得到第三聚合信息之后，该第三聚合信息中可以包括聚合后的模型的权重参数，第一服务端可以使用最终聚合后得到权重参数来更新本地存储的全局模型，从而得到更新后的全局模型。

因此，本申请实施方式中，客户端可以通过向服务端发送请求的方式来接入服务端，在服务端允许之后即可参与联邦学习，相当于客户端以松散的方式接入服务端，而无需与服务端保持长连接，在可以适应一些客户端部署于移动终端的场景。并且，各个服务端除了对与其连接的客户端反馈的模型更新参数进行聚合，还可以聚合其他服务端聚合后得到的信息，从而使各个服务端都可以具有联邦学习系统内的全量数据，客户端在任何时刻接入服务端，都可以获取到全量数据，从而使最终训练得到的模型的输出精度更高。

并且，本申请还可以通过设置计时器和/或计数器的方式，来对接入服务端的客户端进行计数，并在进行联邦学习的过程中进行计时，在计时器的时间窗口内进行联邦学习，若超出时间窗口未完成一个迭代学习过程，则可以直接进行下一轮学习，从而避免存在部分客户端因连接不稳定或者训练效率低而影响整体的训练过程，提高联邦学习的整体学习效率。并且，客户端在进行联邦训练的过程中可以随时退出，而不影响联邦学习的效率，避免了对客户端的绑定。

需要说明的是，前述的步骤 401-408 可以是联邦学习过程中的其中一个迭代过程，在联邦学习的过程中，可以包括一次或者多次迭代，本申请仅以一次迭代为例进行示例性说明，并不作为限定，具体可以根据实际应用场景确定迭代次数。

409、客户端向第一服务端发送查询消息。

其中，客户端（或者称为第二客户端）在需要获取最新的模型时，可以向服务端发送查询消息，以请求查询最新的全局模型。

具体地，客户端可以通过联邦学习系统的地址来访问联邦学习系统，可以由 ELB 来对客户端发送的查询消息进行路由，从而将请求消息路由至适配的服务端中，此处的客户端向第一服务端发送查询消息仅仅是示例性说明，此处的第一服务端可以替换为其他服务端，如第二服务端或者第三服务端等等，此处并不作限定。

410、第一服务端向客户端下发更新的全局模型。

第一服务端在接收到客户端的查询消息之后，即可将本地最新的模型下发给客户端。

在前述步骤 408 中，第一服务端使用最新的聚合信息更新了存储的全局模型，在接收到查询消息之后，即可将最新的全局模型下发给客户端。具体可以是向该客户端下发模型

的结构和参数，如神经网络的层数、卷积核大小、输入输出的分辨率、各个网络层内的权重参数等等，从而使客户端获取到最新的模型结构。

需要说明的是，本申请中的步骤 409 和步骤 410 为可选步骤。

本申请实施方式中，各个服务端之间可以互相传输最新的聚合信息，从而使联邦学习系统内的各个服务端都具有全量数据，客户端在任意时刻接入服务端都可以获取到全量数据，而无需与服务端保持长连接，也可以获取到全量数据，提高客户端能获取到的模型的准确度。

前述对本申请提供的联邦学习方法的流程进行了介绍，为便于理解，下面结合具体的应用场景对本申请提供的联邦学习方法进行更详细的说明。

首先，对本申请提供的联邦学习方法应用的联邦学习系统进行示例性介绍。

示例性地，本申请提供的联邦学习系统可以如图 5 所示，此处以三个服务端（即 FL-Server0、FL-Server1 和 FL-Server2）进行示例性说明，也可以替换为更多的服务端。调度器（即 FL-Scheduler）可以用于对服务端集群内的通信资源或者存储资源进行调度，使各个服务端之间可以进行数据交换。

服务端之间或者服务端与调度器可以建立传输控制协议（transmission control protocol, TCP）连接，并保持稳定的长期连接，通过 TCP 私有协议进行通信。当然，此处的 TCP 也可以替换为其他协议，如因特网分组交换协议（internet packet exchange, IPX）、顺序分组交换（sequenced Packet Exchange, SPX）等协议，本申请对此并不作限定。

客户端可以通过 IP 地址接入联邦学习系统，服务端对外部暴露超文本传输协议（hypertext transfer protocol, HTTP）端口，对外部提供分布式服务。客户端可以与服务端建立 HTTP 连接，即短连接。

服务端 0 可以作为主节点，其他服务端作为从节点，主节点负责集群范围内的数据同步，如联邦学习过程中的触发所有节点进行聚合、全局计时或者计数等，从而保持集群内的数据一致性。

通常，若客户端部署于移动终端中，则服务端可能无法随时获取客户端的状态，如移动终端可能与服务端之间的连接不稳定，检索不到移动终端或者获取不到移动终端的连接状态等，可能导致服务端和移动终端之间通信中断，可能影响联邦学习的效率。

参阅图 6，本申请提供的另一种联邦学习方法的流程示意图。

首先，在每轮迭代之前，即可初始化计时器和计数器，如设置为 0。在每轮迭代过程中，联邦学习系统接收到由客户端发送的第一个请求消息之后，主节点上的计数器即可进行计数。在计数器开始计数时，可以同步开启计时器，在计时器的时间窗口内进行一轮联邦学习。

若计数器超过第一阈值，则可以关闭计时器，将当前轮迭代中发送请求的客户端作为当前轮次联邦学习中进行训练的客户端。该第一阈值可以是根据上一轮参与联邦学习的客户端的数量确定的，例如，上一轮参与联邦学习的客户端相对更上一轮迭代中的客户端减少，则可以在已设置的第一阈值的基础上进行降低，若上一轮参与联邦学习的客户端相对更上一轮迭代中的客户端增多，则可以在原设置的第一阈值的基础上进行增加，从而使计

数器对应的阈值和每轮迭代中参与训练的客户端适配，最大化训练效率。

以其中一个 FL-server 1 和 FL-client 为例，FL-server 1 可以向 FL-client 下发待训练的全局模型和训练配置参数。该全局模型可以包括各种神经网络，如 DNN、CNN、残差网络（ResNet）或者构建得到的网络等，具体可以根据实际应用场景进行调整。训练配置参数可以包括学习率、epoch 数量或者安全算法中类别等训练模型时涉及到的参数，用于在客户端进行模型训练时，确定训练中的学习率、迭代次数或者传输数据时的加密算法类型等。

在 FL-client 使用训练配置参数以及本地存储的训练样本进行模型训练之后，即可将得到的模型更新参数反馈至 FL-server 1，该模型更新参数可以包括训练后的模型相关的参数，如更新的全局模型的结构（如网络层的深度或者宽度等）、权重参数、或者降低稀疏性后的权重参数等。

本实施例仅以一个 FL-client 进行示例性说明，通常可以具有多个 FL-client，每个 FL-client 都可以使用本地存储的训练样本对接收到的全局模型进行训练。例如，每个训练样本中包括样本和对应的真值标签，在一次训练过程中，将样本作为模型的输入，然后通过损失函数来衡量模型的输出和真值标签之间的差异，并通过差异计算全局模型的更新梯度，使用梯度对全局模型的权重参数进行反向更新，从而得到更新后的模型。

若 FL-server 1 接收到多个 FL-client 反馈的模型更新参数，则 FL-server 1 可以对该多个 FL-client 反馈的模型更新参数进行聚合，相当于数据归约。具体的聚合方式可以包括求和、求均值、加权求和等，具体可以根据实际应用场景进行调整。例如，FL-server 1 对模型跟新数据进行聚合得到的数据可以表示为： $w_{t+1} \leftarrow \sum_{k=1}^{k \in S} n_k w_{t+1}^k$ ，其中 S 为客户端集合， w_{t+1}^k 表示各个客户端反馈的模型更新参数， n_k 表示每个客户端反馈的模型更新参数的大小。

在计数器的值超过第一阈值时，由主节点，即 FL-server 0 触发自身所在的服务端和从节点（如 FL-server 1、FL-server 2 等等）对各个服务端之间互相传输的聚合信息进行再次聚合，需要说明的是，此处为便于理解，仅以 FL-server 0 触发从节点进行聚合为例进行示例性说明。例如，在 FL-server 0 的触发下，FL-server 2 将对接收到的模型更新参数聚合得到的聚合信息发送至 FL-server 1，FL-server 1 对接收到的聚合信息和本地存储的聚合信息进行聚合，得到最终的聚合信息。

在进行联邦学习的一轮迭代过程中，若计数器不超过第一阈值，而计时器超时，则可以终止当前次迭代，进入下一轮迭代中，即初始化计数器。如 FL-server 0 向 FL-server 1、FL-server 2 发送第二触发指示，以指示进入下一轮迭代。此时，当前次迭代过程中的数据可以丢弃或者启用。例如，若数据完整也可以继续采用，仅丢弃部分未响应的客户端的数据，具体可以根据实际应用场景调整。

因此，在本申请实施方式中，可以由主节点维护计数器和计时器，对发送了请求的客户端进行计数，在计时器达到阈值时即可触发集群维度的聚合操作，即各个服务端对接收到的其他服务端的聚合信息进行再次聚合，从而使各个服务端都可以具有其他服务端的聚

合信息，每个服务端都可以具有全量数据，当客户端在任意时刻接入任意服务端都可以获取到全量数据，提高联邦学习的效率。并且，主节点还维护计时器，当计时器超时，即可进行下一轮迭代，相当于仅在计时器的时间窗口内进行一轮迭代学习，防止因等待客户端响应的时长太长而影响联邦学习的效率，解决因客户端掉队而导致的长尾效应。

5 此外，本申请提供的联邦学习系统中，主节点作为对每轮迭代时长和参与迭代的客户端的数量进行监控的重要节点，主要可以用于执行以下步骤：

在多个服务端中任意一个服务端接收到第一个请求消息之后，主节点启动计数器和计时器，计数器用于对多个服务端在一轮迭代中接收到的请求消息进行计数，请求消息用于请求获取多个服务端中对应的服务端中存储的全局模型；若计数器的值达到第一阈值，主节点向多个服务端中的每个服务端发送第一触发指示，第一触发指示用于指示多个服务端之间互相传输本地存储的信息；若计数器的值未达到第一阈值，计时器的值达到第二阈值，则主节点向每个服务端发送第二触发指示，第二触发指示用于指示每个服务端进行下一轮迭代。

本申请实施方式中，将联邦学习系统中的其中一个服务端作为主节点，由主节点来维护计时器和计数器，在每轮迭代过程中，通过计数器对接收到的请求消息进行计数，并同时使用计时器来进行计时，从而限制每轮迭代参与的客户端的数量以及每轮迭代的训练时长，从而避免因客户端掉队而导致的无意义的等待，提高联邦学习的整体效率，避免出现长尾效应。

在一种可能的实施方式中，第一阈值为预设值，或者，第一阈值与上一轮迭代过程中接入联邦学习系统的客户端的数量相关。因此，本申请实施方式中，计数器对应的第一阈值可以是预先设定的值，也可以根据接入联邦学习系统内所有服务端的客户端的数量确定的值，通常该第一阈值不大于接入联邦学习系统的客户端的数量，从而避免等待客户端的接入而导致联邦学习的等待时间过长，提高联邦学习的整体效率。

在一种可能的实施方式中，第二阈值为预设值，或者，第二阈值与上一轮迭代中接入联邦学习系统的客户端的数量相关或者第二阈值与上一轮迭代中联邦学习系统内的每个服务端和对应的客户端之间进行通信的数据量的大小相关。因此，本申请实施方式中，可以通过计时器来设置每轮迭代的时间窗口，当计时器超时即可进入下一轮迭代学习，从而避免因等待时长过长而导致的长尾效应，即使客户端在进行联邦学习时中途退出也不影响联邦学习的整体训练效率。

30 为便于理解，参阅图 7，下面结合联邦学习系统内由主节点和各个服务端执行的流程进行示例性介绍。

701、初始化计数器和计时器。

其中，在进行每轮迭代之前，对计数器和计时器进行初始化，如将计数器和计时器都设置为 0。计数器和计时器可以由主节点维护，如前述图 5 中所述示出的 FL-Server 0。

35 例如，在每轮迭代开始时，初始化计数器，设置本轮迭代参与的客户端的请求消息的数量，如设置为 C，通常这个值可以设置为与服务端有通信连接的客户端的数量的子集，即通常不超过与服务端建立了连接的客户端的数量，以避免每轮迭代中对客户端的等待时

长过长。如若在联邦学习系统中，有 100 个客户端和服务端建立了连接，而仅有 10 个客户端需要参与联邦学习，则此时若 C 设置过大，可能导致等待客户端参与的时长过长，降低联邦学习的效率。

702、FL-Server 接收 FL-Client 的请求消息。

5 FL-Client 发送的请求消息可以由 ELB 路由至 FL-Server，此处的 FL-Server 可以是联邦学习系统内的任意一个服务端，如前述图 5 中的 FL-Server0、FL-Server1 或者 FL-Server2 等服务端中的任意一个。

此外，在 FL-Server 接收到 FL-Client 请求消息之后，FL-Server 可以确认是否允许该 FL-Client 参与联邦学习。如可以采用随机筛选、先到先得或者其他筛选方式来确定是否
10 允许 FL-Client 参与联邦学习。例如，还可以通过请求消息确定 FL-Client 的连接状态，如时延、通信时长等，或者 FL-Client 的状态，如电量过低或者负载过高等，则可以不允许该 FL-Client 参与联邦学习，以避免 FL-Client 影响联邦学习的效率。

703、计数器累加。

在 FL-Server 接收到 FL-Client 请求消息时，主节点也可以对发送了请求消息的
15 FL-Client 数量进行记录，计数器进行累加。

704、判断计数器是否为 1，若是，则执行步骤 708，若否，则执行步骤 705。

其中，在对计数器进行累加之后，可以对计数器的值进行监控，若计数器的值为 1，即在本轮迭代过程中，接收到了第一个客户端发送的请求消息，此时即可开启计时器，即执行步骤 709，而此时仅因仅接收到一个客户端发送的请求消息，因此无需进行数据同步，
20 在接收到多个客户端发送的请求消息之后，即可后在集群维度进行数据同步，使集群内的服务端保持数据一致性，即执行步骤 705。

705、FL-Server 处理请求并在集群维度同步数据。

在接收到多个 FL-Client 发送的请求消息之后，即可向该多个 FL-Client 中的全部或者部分 FL-Client 下发模型以及训练配置参数。FL-Client 基于训练配置参数，使用本地
25 存储的训练样本对该模型进行训练，并将训练后的模型的相关数据反馈至 FL-Server。

若多个 FL-Client 参与了联邦学习，则 FL-Server 可以接收多个 FL-Client 反馈的模型更新参数，FL-Client 可以对接收到的多个 FL-Client 反馈的模型更新参数进行聚合，得到聚合信息（即第一聚合信息）。各个 FL-Server 之间可以同步各自计算得到的聚合信息，从而实现集群维度的数据同步。

30 706、FL-Server 之间进行聚合。

在各个 FL-Server 之间同步了各自的聚合信息之后，即可对自身聚合得到的聚合信息（即第一聚合信息）其他 FL-Server 发送的聚合信息（即第二聚合信息）进行聚合，得到最终的全局聚合信息（即第三聚合信息），从而使各个 FL-Server 都可以具有全量数据。

示例性地，FL-Server 之间进行聚合的方式具体可以如图 8 所示，其中，联邦学习系统
35 内包括了多个服务端，该多个服务端呈环形互相连接，即如图 8 中所示出的 Server 0、Server 1、Server 2 和 Server 3 等，此处仅以 4 个服务端作为示例性说明，实际应用场景中也可以替换为更多或者更少的服务端数量，此处不作限定。客户端包括与服务端连接的 K 个客

户端,如图 8 中所示出的 client 1 至 client K,每个客户端接收接入的服务端发送的初始权重 w_t ,然后使用本地存储的训练样本对初始权重对应的模型进行训练,得到训练后的模型的权重 $w_{t+1}^1-w_{t+1}^K$,以及对应的数据量 n_1-n_K , n_k 表示第 k 个客户端的数据集大小,然后将训练后的模型的权重以及数据量反馈给接入的服务端,如 client 1 将训练后的模型

5 权重反馈给 Server 0,client K 将训练后得到的模型的权重反馈给 Server 2 等。服务端 j 对各自接收到的客户端发送的权重进行聚合,得到 $w_{t+1}^j \leftarrow \sum_{k=1}^{k \in S_j} n_{k,j} w_{t+1}^{k,j}$, $n_j = \sum_{k=1}^{k \in S_j} n_{k,j}$,然后主节点触发各个服务端对服务端之间传输的数据进行聚合,得到聚合信息,如每个服务端聚合后的数据表示为 w_{t+1}^j 和 n_j 。

707、计数是否达到第一阈值,若是,则执行步骤 710,若否,则执行步骤 702。

10 在本轮的联邦学习迭代过程层中,可以实时对计数器的数值进行监测,若计数器的计数达到第一阈值,则表示本轮联邦学习迭代过程中参与的客户端的数量已足够,无需在增加更多的客户端来参与本轮联邦学习,即停止计时,并进入下一轮联邦学习的迭代过程,即执行步骤 710。

708、触发计时器计时。

15 在联邦学习的每轮迭代过程中,在接收到第一个由客户端发送的请求消息之后,即可触发计时器开始计时。

例如,在 ELB 将第一个请求消息路由至对应的服务器的同时或者之后,ELB 还通知主节点已接收到第一个请求消息,主节点在接收到 ELB 的通知消息之后即可启动计时器和计数器,以对联邦学习系统接收到的请求消息进行计数,并对本轮迭代的过程进行计时。

20 709、是否超时,若是,则执行步骤 710。

在每轮迭代学习的过程中,可以对计时器的状态进行实时监测,判断计时器是否超时,若超时,则停止计时,并进行下一轮迭代学习。

可以理解为,相当于通过计时器来设置时间窗口,在计时器的时间窗口内进行本轮迭代学习过程,若超出该时间窗口,则可以进入下一轮迭代中,从而避免每轮迭代消耗的时长过长。

25 在计时器的时间窗口 T 内所接收到的客户端发送的请求消息或者数据可以被服务端正常处理,如执行加解密或者对接收到的数据进行聚合等,若在时间窗口 T 之外若计数器还未达到 C,则计时器和计数器都重置为初始状态,进入下一轮迭代;而若在时间窗口 T 之内计数器达到 C,则停止计时器的计时,并在处理完本轮迭代中客户端发送的数据之后,

30 进入下一轮迭代。因此,在本申请实施方式中,在计数器和计时器相互配合的基础上,每轮迭代耗时最长即为时间窗口 T 的时长,并且在迭代过程中,可以记录每轮迭代的通信时长,从而动态调整时间窗口的时长,从而消除客户端的长尾效应,提升训练效率。

此外,可选地,该计时器的作用也可以是触发集群内的服务端之间进行数据同步,即前述步骤 705-706,在完成步骤 705-706 之后,即可进入下一轮迭代,具体可以根据实际

35 应用场景进行调整。

710、停止计时。

在计时器停止计时之后，即表示结束本轮迭代学习，在初始化计时器和计数器之后，即可进入下一轮迭代，即执行步骤 701。

其中，步骤 702、步骤 705 和步骤 706 可以由集群内的任意一个 FL-Server 执行，步骤 5 701、步骤 703、步骤 704、步骤 707-步骤 710 则由主节点来执行。

因此，本申请实施方式中，客户端可以以请求和响应结合的方式接入服务端，通过设置计数器和计时器的方式来对每轮迭代参与的客户端和迭代时长进行监控，即使客户端在联邦学习的中途退出，也不影响集群的联邦学习的整体学习过程，避免因寻址不到客户端而导致的掉尾问题，提高联邦学习的整体效率。并且，各个服务端之间可以进行数据同步，10 即每个服务端都具有集群内的全量数据，客户端在任意时刻接入任意服务端都可以获取到全量数据，从而提高客户端得到的模型的准确度。

示例性地，以一个训练简单的分类网络为例。联邦学习系统可以由多台服务端组成，客户端部署于用户的移动终端，如手机、相机或平板等移动终端。移动终端可以通过摄像头采集到图像，用户可以在相册手动标记出拍摄到的图像中的对象的类型，如图像中的“猫”、15 “狗”等，移动终端可以将用户手动标记后的图像作为训练样本，当训练样本的数量达到一定量时，移动终端可以向联邦学习系统请求参与联邦学习，从而将自身采集到的数据应用到分类网络中。

移动终端可以获取到联邦学习系统对外暴露的 HTTP 端口，并通过该 HTTP 端口向集群发送请求消息，主节点上设置的计数器对该请求消息进行计数，且主节点还维护计时器。20 计数器的上限值可以是预先设定的，也可以是根据上一轮迭代中参与的客户端的数量确定的，或者根据与集群建立了连接的客户端的数量确定的。计时器的上限值可以是预先设定的，也可以是根据上一轮迭代所花费的时长确定的等。

且若该移动终端发送的请求消息为本轮迭代过程中发送的第一个请求消息，则启用计时器。若移动终端发送的请求消息为本轮迭代中的最后一个请求消息，即计数器的值达到25 预先设定的第一阈值，则该移动终端为本轮迭代中确定的最后一个参与联邦学习的客户端，在移动终端完成模型训练之后，即可进入下一轮迭代学习。

集群内部的 ELB 将移动终端的请求消息路由至适配的服务端中，由该服务端确定是否允许移动终端参与联邦学习。若服务端允许该移动终端参与联邦学习，服务端将本地存储的分类网络下发给移动终端，同时还下发训练配置参数，如训练时所使用的学习率、数据的加解密方式等。移动终端使用本地采集到的训练样本，根据训练配置参数对分类网络进行30 训练，得到训练后的分类网络。移动终端将训练后的分类网络的权重参数反馈给服务端，若服务端接收到该权重参数的时刻在计时器的时间窗口内，则服务端对接收到的权重参数进行聚合，得到第一聚合信息。

服务端之间可以互相传输聚合得到的信息，从而使每个服务端上都具有全量数据。移动35 终端可以在任何时刻向服务端发送查询消息，以查询最新的分类网络，服务端可以将存储的最新的分类网络发送给移动终端，从而时移动终端在任意时刻接入任意服务端都可以获取到最新的全量数据。

因此，在本申请实施方式中，当客户端部署于移动终端时，可以通过请求和响应结合的方式，使客户端以松散的方式接入联邦学习系统并参与联邦学习，无需保持稳定的长连接，适用于客户端部署于移动终端的场景。并且，联邦学习系统内部还部署了计时器和计数器，通过计时器和计数器配合的方式，避免因等待客户端而导致的联邦学习效率低，减少长尾效应，提高了联邦学习的整体效率。此外，计时器或者计数器的上限值可以根据上一轮迭代或者前几轮迭代训练过程中的信息来调整的，可以根据实际场景动态调整计时器或者计数器的上限值，进一步提高联邦学习的效率。

前述对本申请提供的联邦学习系统和联邦学习方法的流程进行了详细介绍，下面对本申请提供的服务端的结构进行介绍。

参阅图 9，本申请提供一种服务端的结构示意图，该服务端应用于联邦学习系统，联邦学习系统中包括多个服务端和多个客户端，多个服务端用于进行迭代联邦学习，多个服务端中的任意一个服务端包括：

收发模块 901，用于接收至少一个第一客户端发送的请求消息，请求消息用于请求第一服务器中存储的全局模型，多个客户端包括至少一个第一客户端；

收发模块 901，还用于向至少一个第一客户端发送全局模型的信息和训练配置参数，全局模型的信息和训练配置参数用于指示至少一个第一客户端使用训练配置参数对全局模型进行训练；

收发模块 901，还用于接收至少一个第一客户端分别反馈的第一模型更新参数，第一模型更新参数是至少一个第一客户端训练后得到的全局模型的参数；

聚合模块 902，用于对至少一个第一客户端反馈的第一模型更新参数进行聚合，以得到在本轮迭代中的第一聚合信息；

收发模块 901，还用于获取第二服务端发送的第二聚合信息，第二聚合信息为第二服务端在本轮迭代中，对接收到的第二模型更新参数进行聚合得到的信息；

更新模块 903，用于基于第一聚合信息和第二聚合信息，更新第一服务端上存储的全局模型，以得到更新后的全局模型。

在一种可能的实施方式中，收发模块，还用于接收多个第二服务端分别发送的第二聚合信息；

更新模块 903，具体用于基于第一聚合信息和多个第二服务端分别发送的第二聚合信息，更新第一服务端上存储的全局模型，以得到更新后的全局模型。

在一种可能的实施方式中，第一服务器为联邦学习系统中的主节点，主节点用于对多个服务端进行管理，

收发模块，还用于向第二服务器发送第一触发指示，第一触发指示用于指示第二服务端向第一服务端发送第二聚合信息；

收发模块 901，还用于接收第二服务端中的第二聚合信息。

在一种可能的实施方式中，第一服务器中设置有计数器，计数器用于对多个服务端接收到的请求消息的数量进行计数，

收发模块 901，具体用于当计数器的数值满足第一阈值时，向第二服务端发送第一触

发指示。

在一种可能的实施方式中，第一阈值为预设值，或者，第一阈值与上一轮迭代过程中接入联邦学习系统的客户端的数量相关。

在一种可能的实施方式中，主节点中包括计时器，计时器从每一轮迭代中接收到第一个请求消息之后开始计时，

收发模块 901，具体用于当计时器超过第二阈值时，接收主节点发送的第二触发指示，第二触发指示用于指示进行下一轮迭代学习。

在一种可能的实施方式中，第二阈值为预设值，或者，第二阈值与上一轮迭代中接入联邦学习系统的客户端的数量相关，或者，第二阈值与上一轮迭代中联邦学习系统内的每个服务端和对应的客户端之间进行通信的数据量的大小相关。

在一种可能的实施方式中，收发模块，还用于接收第三客户端发送的查询消息，第三客户端包括接入联邦学习系统中的客户端中的任意一个；

收发模块 901，还用于针对查询消息向第三客户端发送更新后的全局模型的信息。

在一种可能的实施方式中，收发模块，还用于向第二服务端发送第一聚合信息，以使第二服务端基于第一聚合信息和第二聚合信息，更新本地存储的全局模型以得到更新后的全局模型。

参阅图 10，本申请提供另一种服务端的结构示意图，即前述的主节点。该服务端应用于联邦学习系统，联邦学习系统包括多个服务端，多个服务端中的其中一个服务端作为主节点，多个服务端用于进行迭代学习以实现联邦学习，主节点包括：

启动模块 1001，用于在多个服务端中任意一个服务端接收到第一个请求消息之后，主节点启动计数器和计时器，计数器用于对多个服务端在一轮迭代中接收到的请求消息进行计数，请求消息用于请求获取多个服务端中对应的服务端中存储的全局模型；

收发模块 1002，用于若计数器的值达到第一阈值，主节点向多个服务端中的每个服务端发送第一触发指示，第一触发指示用于指示多个服务端之间互相传输本地存储的信息；

收发模块 1002，还用于若计数器的值未达到第一阈值，计时器的值达到第二阈值，则主节点向每个服务端发送第二触发指示，第二触发指示用于指示每个服务端进行下一轮迭代。

在一种可能的实施方式中，第一阈值为预设值，或者，第一阈值与上一轮迭代过程中接入联邦学习系统的客户端的数量相关。

在一种可能的实施方式中，第二阈值为预设值，或者，第二阈值与上一轮迭代中接入联邦学习系统的客户端的数量相关或者第二阈值与上一轮迭代中联邦学习系统内的每个服务端和对应的客户端之间进行通信的数据量的大小相关。

本申请实施例还提供一种服务端，请参阅图 11，图 11 为本申请实施例提供的服务端的一种结构示意图。服务端 1100 上可以部署有图 9 对应实施例中所描述的服务端，用于实现图 4 至图 8 对应实施例中服务端的功能。具体的，服务端 1100 可因配置或性能不同而产生比较大的差异，可以包括一个或一个以上中央处理器 CPU1122（例如，一个或一个以上处理器）和存储器 1132，一个或一个以上存储应用程序 1142 或数据 1144 的存储介质 1130

(例如一个或一个以上海量存储设备)。

其中，存储器 1132 和存储介质 1130 可以是短暂存储或持久存储。在一个实施例中，存储器 1132 为随机存储存储器 RAM，可以与中央处理器 1122 直接交换数据，用于加载数据 1144 和应用程序 1142 和/或操作系统 1141 以供中央处理器 1122 直接运行与运用，通常
5 作为操作系统或其他正在运行中的程序的临时数据存储媒介。存储在存储介质 1130 的程序可以包括一个或一个以上模块(图 11 中未示出)，每个模块可以包括对服务端中的一系列指令操作。

更进一步地，中央处理器 1122 可以设置为与存储介质 1130 通信，在服务端 1100 上执行存储介质 1130 中的一系列指令操作。在一种可能的实施方式中，存储介质 1130 中存储了前述图 4-图 8 中任一实施例所示的方法步骤对应的程序指令以及数据。
10

服务端 1100 还可以包括一个或一个以上电源 1126，一个或一个以上有线或无线网络接口 1150，一个或一个以上输入输出接口 1158，和/或，一个或一个以上操作系统 1141，例如 Windows Server™，Mac OS X™，Unix™，Linux™，FreeBSD™ 等等。

在一种可能的实施方式中，中央处理器 1122 用于执行前述图 4-图 8 中任一实施例所示的服务端执行的方法步骤。
15

本申请实施例还提供了一种联邦学习装置，该联邦学习装置也可以称为数字处理芯片或者芯片，芯片包括处理单元和通信接口，处理单元可以通过通信接口获取程序指令，程序指令被处理单元执行，处理单元用于执行前述图 4-图 8 中任一实施例所示的服务端执行的方法步骤。

本申请实施例还提供一种数字处理芯片。该数字处理芯片中集成了用于实现上述中央处理器 1122，或者中央处理器 1122 的功能的电路和一个或者多个接口。当该数字处理芯片中集成了存储器时，该数字处理芯片可以完成前述实施例中的任一个或多个实施例的方法步骤。当该数字处理芯片中未集成存储器时，可以通过通信接口与外置的存储器连接。该数字处理芯片根据外置的存储器中存储的程序代码来实现上述实施例中服务端执行的动
20 作。

更详细地，当本申请实施例提供的联邦学习装置为芯片时，该芯片具体包括：处理单元和通信单元，所述处理单元例如可以是处理器，所述通信单元例如可以是输入/输出接口、管脚或电路等。该处理单元可执行存储单元存储的计算机执行指令，以使服务端内的芯片执行上述图 4-图 8 所示实施例描述的联邦学习方法。

可选地，前述的存储单元可以为芯片内的存储单元，如寄存器、缓存等，所述存储单元还可以是所述无线接入设备端内的位于所述芯片外部的存储单元，如只读存储器(read-only memory, ROM)或可存储静态信息和指令的其他类型的静态存储设备，随机存取存储器 RAM 等。
30

具体地，前述的处理单元或者处理器可以是中央处理器、网络处理器(neural-network processing unit, NPU)、图形处理器(graphics processing unit, GPU)、数字信号处理器(digital signal processor, DSP)、专用集成电路(application specific integrated circuit, ASIC)或现场可编程逻辑门阵列(field programmable gate array, FPGA)或
35

者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者也可以的任何常规的处理器等。

其中，上述任一处提到的处理器，可以是一个通用中央处理器，微处理器，ASIC，或一个或多个用于控制上述图 4-图 8 的方法的程序执行的集成电路。

5 本申请实施例中还提供一种计算机可读存储介质，该计算机可读存储介质中存储有程序，当其在计算机上运行时，使得计算机执行如前述图 4-图 8 所示实施例描述的方法中的步骤。

本申请实施例中还提供一种计算机程序产品，当其在计算机上运行时，使得计算机执行如前述图 4-图 8 所示实施例描述的方法中服务端所执行的步骤。

10 另外需说明的是，以上所描述的装置实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。另外，本申请提供的装置实施例附图中，模块之间的连接关系表示它们之间具有通信连接，具体可以实现为一条或多条通信总线或信号线。

通过以上的实施方式的描述，所属领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件的方式来实现，当然也可以通过专用硬件包括专用集成电路、专用 CPU、专用存储器、专用元器件等来实现。一般情况下，凡由计算机程序完成的功能都可以很容易地用相应的硬件来实现，而且，用来实现同一功能的具体硬件结构也可以是多种多样的，20 例如模拟电路、数字电路或专用电路等。但是，对本申请而言更多情况下软件程序实现是更佳的实施方式。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在可读取的存储介质中，如计算机的软盘、U 盘、移动硬盘、只读存储器 ROM、随机存取存储器 RAM、磁碟或者光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务端，或者网络设备25 等）执行本申请各个实施例所述的方法。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。

所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是30 通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机、服务端或数据中心通过有线（例如同轴电缆、光纤、数字用户线（DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务端或数据中心进行传输。所述计算机可读存储介质可以是计算机35 能够存储的任何可用介质或者是包含一个或多个可用介质集成的服务端、数据中心等数据存储设备。所述可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，DVD）、或者半导体介质（例如固态硬盘（solid state disk，SSD））等。

本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等（如果存在）是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的实施例能够以除了在这里图示或描述的内容以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

最后应说明的是：以上，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。

权 利 要 求

1. 一种联邦学习方法，其特征在于，

应用于联邦学习系统，所述联邦学习系统中包括多个服务端和多个客户端，所述多个服务端包括第一服务端和第二服务端，所述多个服务端用于进行迭代学习以实现联邦学习，其中所述迭代学习中的任意一轮迭代的过程包括：

所述第一服务端接收至少一个第一客户端发送的请求消息，所述请求消息用于请求所述第一服务器中存储的全局模型；

所述第一服务端向所述至少一个第一客户端发送所述全局模型的信息和训练配置参数；

所述第一服务端接收至少一个第一客户端分别反馈的第一模型更新参数，所述第一模型更新参数是所述至少一个第一客户端训练的全局模型的参数；

所述第一服务端对所述至少一个第一客户端反馈的第一模型更新参数进行聚合，以得到在本轮迭代中的第一聚合信息；

所述第一服务端获取所述第二服务端发送的第二聚合信息，所述第二聚合信息为所述第二服务端在本轮迭代中，对接收到的第二模型更新参数进行聚合得到的信息；

所述第一服务端基于所述第一聚合信息和所述第二聚合信息，更新所述第一服务端上存储的所述全局模型，以得到更新后的全局模型。

2. 根据权利要求1所述的方法，其特征在于，所述多个服务端中包括的所述第二服务端的数量为多个，所述第一服务端获取所述第二服务端发送的第二聚合信息，包括：

所述第一服务端接收所述多个第二服务端分别发送的所述第二聚合信息；

所述第一服务端基于所述第一聚合信息和所述第二聚合信息，更新所述第一服务端上存储的所述全局模型，包括：

所述第一服务端基于所述第一聚合信息和所述多个第二服务端分别发送的所述第二聚合信息，更新所述第一服务端上存储的所述全局模型，以得到更新后的全局模型。

3. 根据权利要求1或2所述的方法，其特征在于，所述第一服务端为所述联邦学习系统中的主节点，所述主节点用于对所述多个服务端进行管理，所述第一服务端获取第二服务端发送的第二聚合信息，还包括：

所述第一服务端向所述第二服务器发送第一触发指示，所述第一触发指示用于指示所述第二服务端向所述第一服务端发送所述第二聚合信息；

所述第一服务端接收所述第二服务端中的第二聚合信息。

4. 根据权利要求2所述的方法，其特征在于，所述第一服务器中设置有计数器，所述计数器用于对所述多个服务端接收到的所述请求消息的数量进行计数，所述第一服务端向所述第二服务器发送第一触发指示，包括：

当所述计数器的数值满足第一阈值时，所述第一服务端向所述第二服务端发送所述第一触发指示。

5. 根据权利要求4所述的方法，其特征在于，所述第一阈值为预设值，或者，所述第一阈值与上一轮迭代过程中接入所述联邦学习系统的客户端的数量相关。

6. 根据权利要求3-5中任一项所述的方法，其特征在于，所述主节点中包括计时器，

所述计时器从每一轮迭代中接收到第一个请求消息之后开始计时，所述方法还包括：

当所述计时器超过第二阈值时，所述第一服务端向所述第二服务器发送的第二触发指示，所述第二触发指示用于指示所述第二服务器进行下一轮迭代。

7. 根据权利要求6所述的方法，其特征在于，所述第二阈值为预设值，或者，所述第二阈值与上一轮迭代中接入所述联邦学习系统的客户端的数量相关，或者，所述第二阈值与上一轮迭代中联邦学习系统内的每个服务端和对应的客户端之间进行通信的数据量的大小相关。

8. 根据权利要求1-7中任一项所述的方法，其特征在于，所述方法还包括：

所述第一服务端接收第三客户端发送的查询消息，所述第三客户端包括接入所述联邦学习系统中的客户端中的任意一个；

所述第一服务端针对所述查询消息向所述第三客户端发送更新后的全局模型的信息。

9. 根据权利要求1-8中任一项所述的方法，其特征在于，所述方法还包括：

所述第一服务端向所述第二服务端发送所述第一聚合信息，以使所述第二服务端基于所述第一聚合信息和所述第二聚合信息，更新本地存储的所述全局模型以得到所述更新后的全局模型。

10. 一种联邦学习方法，其特征在于，应用于联邦学习系统和多个客户端，所述联邦学习系统包括多个服务端，所述多个服务端中的其中一个服务端作为主节点，所述多个服务端用于进行迭代学习以实现联邦学习；

其中所述迭代学习中的任意一轮迭代的过程包括：

在所述多个服务端中任意一个服务端接收到第一个请求消息之后，所述主节点启动计数器和计时器，所述计数器用于对所述多个服务端在一轮迭代中接收到的请求消息进行计数，所述请求消息用于客户端请求获取所述多个服务端中对应的服务端中存储的全局模型；

若所述计数器的值达到第一阈值，所述主节点向所述多个服务端中的每个服务端发送第一触发指示，所述第一触发指示用于指示所述多个服务端之间互相传输本地存储的信息；

若所述计数器的值未达到所述第一阈值，所述计时器的值达到第二阈值，则所述主节点向所述每个服务端发送第二触发指示，所述第二触发指示用于指示所述每个服务端进行下一轮迭代。

11. 根据权利要求10所述的方法，其特征在于，所述第一阈值为预设值，或者，所述第一阈值与上一轮迭代过程中接入所述联邦学习系统的客户端的数量相关。

12. 根据权利要求10或11所述的方法，其特征在于，所述第二阈值为预设值，或者，所述第二阈值与上一轮迭代中接入所述联邦学习系统的客户端的数量相关或者所述第二阈值与上一轮迭代中联邦学习系统内的每个服务端和对应的客户端之间进行通信的数据量的大小相关。

13. 一种联邦学习系统，其特征在于，包括多个服务端和多个客户端，所述多个服务端包括第一服务端和第二服务端，所述第一服务端和所述第二服务端上均存储有全局模型的信息，所述多个服务端用于进行迭代学习以实现联邦学习，其中所述迭代学习中的任意一轮迭代过程中：

所述第一服务端用于接收至少一个第一客户端分别发送的请求消息；

所述第一服务端用于针对所述至少一个第一客户端分别发送的请求消息向所述至少一个第一客户端发送所述全局模型的信息和训练配置参数；

5 所述第一服务端用于接收所述至少一个第一客户端分别反馈的第一模型更新参数，所述第一模型更新参数为所述至少一个第一客户端训练后得到的全局模型的参数；

所述第一服务端用于对所述至少一个第一客户端分别反馈的第一模型更新参数进行聚合，得到第一聚合信息；

所述第二服务端用于接收对应的至少一个第二客户端发送的第二模型更新参数，并对所述至少一个第二客户端发送的第二模型更新参数进行聚合，以得到第二聚合信息；

10 所述第一服务端用于接收所述每个第二服务端发送的第二聚合信息；

所述第一服务端用于基于所述第一聚合信息和所述每个第二服务端发送的第二聚合信息，更新所述第一服务端上存储的所述全局模型，以得到更新后的全局模型。

14. 根据权利要求 13 所述的系统，其特征在于，所述多个服务端中包括的所述第二服务端的数量为多个；

15 所述多个第二服务端中的每个第二服务端用于接收对应的至少一个第二客户端发送的第二模型更新参数，并对所述至少一个第二客户端发送的第二模型更新参数进行聚合，以得到所述第二聚合信息；

所述第一服务端具体用于接收所述多个第二服务端分别发送的所述第二聚合信息；

20 所述第一服务端具体用于基于所述第一聚合信息和所述多个第二服务端分别发送的所述第二聚合信息，更新所述第一服务端上存储的所述全局模型，以得到更新后的全局模型。

15. 根据权利要求 13 或 14 所述的系统，其特征在于，所述多个服务端中还包括作为主节点的第三服务端，所述主节点用于对所述多个服务端进行管理；

所述主节点用于向所述多个服务端分别发送第一触发指示；

25 所述第二服务器具体用于基于所述第一触发指示向所述第一服务器发送所述第二聚合信息。

16. 根据权利要求 15 所述的系统，其特征在于，所述主节点中包括计数器，所述计数器用于对所述多个服务端接收到的请求消息的数量进行计数，所述请求消息用于请求获取所述多个服务端中对应的服务端中存储的全局模型；

30 所述主节点具体用于当所述计数器的数值满足第一阈值时，向所述多个服务端中的每个服务端发送所述第一触发指示，所述第一触发指示用于触发所述每个第二服务器向所述第一服务器发送所述第二聚合信息。

17. 根据权利要求 15 或 16 所述的系统，其特征在于，所述主节点还包括计时器，所述计时器从每一轮迭代过程中接收到的第一个所述请求消息开始计时；

35 所述主节点还用于当所述计时器超过第二阈值时，向所述多个服务端中的每个服务端分别发送第二触发指示，所述第二触发指示用于指示所述多个服务端进行下一轮迭代。

18. 根据权利要求 17 所述的系统，其特征在于，所述第二阈值为预设值，或者，所述第二阈值与上一轮迭代过程中接入所述联邦学习系统中每个服务端的客户端的数量相关，

或者，所述第二阈值与上一轮迭代过程中所述多个服务端和所述多个客户端之间进行通信的数据量的大小相关。

19. 根据权利要求 13-18 中任一项所述的系统，其特征在于，

所述第一服务端接收第三客户端发送的查询消息，所述第三客户端包括接入所述联邦
5 学习系统的任意一个客户端；

所述第一服务端针对所述查询消息向所述第三客户端发送更新后的全局模型的信息。

20. 根据权利要求 13-19 中任一项所述的系统，其特征在于，

所述第一服务端还用于向所述第二服务端发送所述第一聚合信息；

所述第二服务器端具体用于结合所述第一聚合信息和所述第二聚合信息，更新本地存
10 储的所述全局模型，以得到所述更新后的全局模型。

21. 一种联邦学习装置，其特征在于，包括处理器，所述处理器和存储器耦合，所述
存储器存储有程序，当所述存储器存储的程序指令被所述处理器执行时实现权利要求 1-9
或者 10-12 中任一项所述的方法的步骤。

22. 一种计算机可读存储介质，包括程序，当其被处理单元所执行时，执行如权利要
15 求 1 至 9 或者 10-12 中任一项所述的方法的步骤。

23. 一种计算机程序产品，其特征在于，所述计算机程序产品包括软件代码，所述软
件代码用于执行如权利要求 1 至 9 或者 10-12 中任一项所述的方法的步骤。