

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022 年 3 月 24 日 (24.03.2022)



(10) 国际公布号
WO 2022/057433 A1

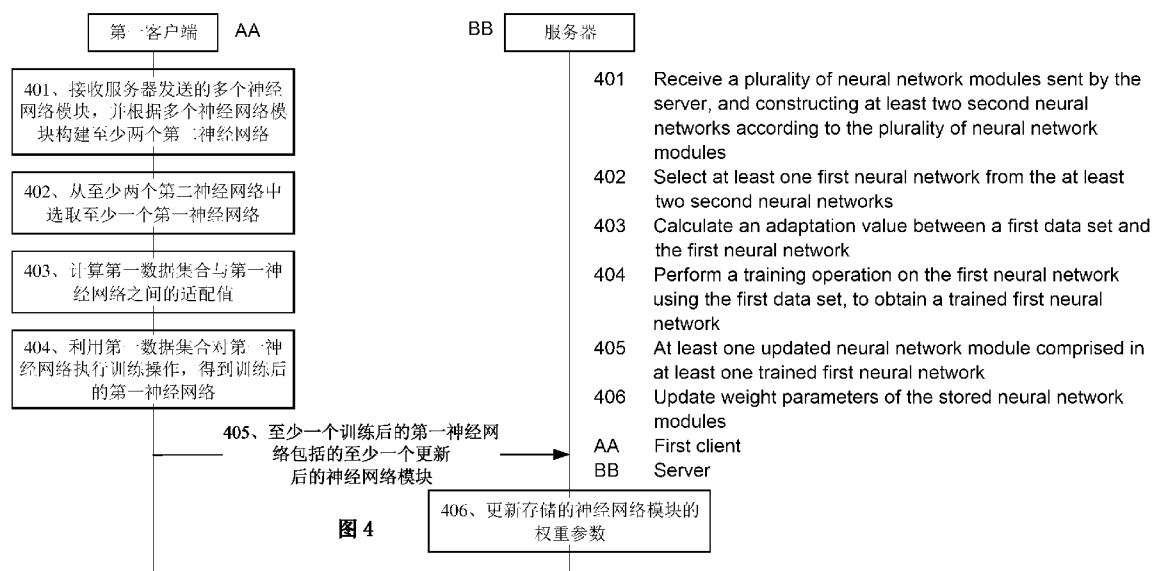
- (51) 国际专利分类号:
G06N 3/04 (2006.01)
- (21) 国际申请号: PCT/CN2021/107391
- (22) 国际申请日: 2021 年 7 月 20 日 (20.07.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202010989062.5 2020 年 9 月 18 日 (18.09.2020) CN
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 邵云峰 (SHAO, Yunfeng); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong

518129 (CN)。宋绍铭 (SONG, Shaoming); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。李温鹏 (LI, Wenpeng); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。郭凯洋 (GUO, Kaiyang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。钱莉 (QIAN, Li); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

- (74) 代理人: 深圳市深佳知识产权代理事务所 (普通合伙) (SHENPAT INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区南湖街道春风路庐山大厦 B 座 18C2、18D、18E、18E2, Guangdong 518001 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,

(54) **Title:** MACHINE LEARNING MODEL TRAINING METHOD AND RELATED DEVICE

(54) 发明名称: 一种机器学习模型的训练的方法以及相关设备



(57) **Abstract:** A machine learning model training method and a related device, relating to the field of artificial intelligence. The method is applied to a first client; a plurality of clients is in communication connection with a server; the server stores a plurality of modules; the plurality of modules is used for constructing at least two machine learning models. The method comprises: obtaining a first machine learning model, at least one first machine learning model being selected according to data characteristics of a first training data set stored in the first client; performing a training operation on the at least one first machine learning model using the first data set, to obtain at least one trained first machine learning model; and sending at least one updated module to the server, the updated module being used for the server to update weight parameters of the stored modules. Different neural networks are distributed to training data having different data characteristics, thereby achieving personalized matching between the neural networks and the data characteristics.



BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 一种机器学习模型的训练的方法以及相关设备, 涉及人工智能领域。该方法应用于第一客户端, 多个客户端与服务器通信连接, 服务器中存储有多个模块, 多个模块用于构建至少两个机器学习模型, 方法包括: 获取第一机器学习模型, 至少一个第一机器学习模型为根据第一客户端存储的第一训练数据集合的数据特性选取出来的; 利用第一数据集合对至少一个第一机器学习模型执行训练操作, 得到至少一个训练后的第一机器学习模型; 将至少一个更新后的模块发送给服务器, 更新后的模块用于供服务器更新存储的模块的权重参数, 对不同数据特性的训练数据分配不同的神经网络, 实现了神经网络与数据特性之间的个性化匹配。

一种机器学习模型的训练的方法以及相关设备

本申请要求于 2020 年 9 月 18 日提交中国专利局、申请号为 202010989062.5、发明名称为“一种机器学习模型的训练的方法以及相关设备”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

本申请涉及人工智能领域，尤其涉及一种机器学习模型的训练的方法以及相关设备。

背景技术

人工智能（Artificial Intelligence，AI）是利用计算机或者计算机控制的机器模拟、延伸和扩展人的智能。人工智能包括研究各种智能机器的设计原理与实现方法，使机器具有感知、推理与决策的功能。目前，随着用户对个人隐私数据的保护意愿日益提升，数据拥有者之间的用户数据无法互通，形成了大大小小的“数据孤岛”。“数据孤岛”对基于海量数据的人工智能提出了新的挑战。

针对“数据孤岛”的存在，联邦学习（federated learning）被提出，也即不同客户端利用本地存储的训练数据对同一神经网络进行训练，并将训练后的神经网络发送给服务器，由服务器对参数的更新情况进行汇聚。但由于不同客户端中存储的训练数据的数据特性不同，也即不同客户端的优化目标不一致，并且每一轮训练所选择的客户端也不完全相同，导致每一轮训练的优化目标也可能不一致，从而导致神经网络的训练过程容易产生震荡，导致训练效果不佳。

发明内容

本申请实施例提供了一种机器学习模型的训练的方法以及相关设备，对不同数据特性的训练数据分配不同的神经网络，实现了神经网络与数据特性之间的个性化匹配；每个客户端均根据客户端存储的训练数据集合的数据特性分配并训练神经网络，能够利用相同数据特性的训练数据训练相同的神经网络，有利于提高训练后神经网络的准确率。

为解决上述技术问题，本申请实施例提供以下技术方案：

第一方面，本申请实施例提供一种机器学习模型的训练方法，可用于人工智能领域中。方法应用于第一客户端，多个客户端与服务器通信连接，服务器中存储有多个模块，多个模块用于构建机器学习模型，第一客户端为多个客户端中的任一客户端；机器学习模型具体可以表现为神经网络、线性模型或其他类型的机器学习模型，对应的，组成机器学习模型的多个模块具体可以表现为神经网络模块、线性模型模块或组成其他类型的机器学习模型的模块。机器学习模型的训练包括多轮迭代，多轮迭代中的一轮迭代包括：第一客户端获取至少一个第一机器学习模型，至少一个第一机器学习模型为根据第一客户端存储的第一数据集合的数据特性选取出来的；具体的，第一客户端可以接收服务器发送的多个模块，并从至少两个第二机器学习模型中选取该至少一个第一机器学习模型，或者，第一客户端可以接收服务器发送的该至少一个第一机器学习模型。第一客户端利用第一数据集合对至

-2-

少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型；第一客户端将至少一个训练后的第一机器学习模型包括的至少一个更新后的模块发送给服务器，更新后的模块用于供服务器更新存储的模块的权重参数。

5 本实现方式中，服务器中存储有多个神经网络模块，多个神经网络模块能够组成至少两个不同的第二神经网络，在针对多个客户端中的一个第一客户端时，会选择与第一客户端存储的第一数据集合的数据特性匹配的至少一个第一神经网络，在利用第一客户端的训练数据对该至少一个第一神经网络进行训练后，再由服务器对参数的更新情况进行汇聚；通过前述方式，能够对不同数据特性的训练数据分配不同的神经网络，也即实现了神经网络与数据特性之间的个性化匹配；此外，由于第一客户端为多个客户端中的任一客户端，
10 对多个客户端中的每个客户端均根据客户端存储的训练数据集合的数据特性分配并训练神经网络，能够利用相同数据特性的训练数据训练相同的神经网络，不同数据特性的训练数据训练不同的神经网络，从而不仅实现了神经网络与数据特性之间的个性化匹配，而且有利于提高训练后神经网络的准确率。

在第一方面的一种可能实现方式中，多个模块用于构建至少两个第二机器学习模型，
15 至少一个第一机器学习模型为从至少两个第二机器学习模型中选取出来的；或者，用于构建至少一个第一机器学习模型的模块为从多个模块中选取出来的。本实现方式中，提供了第二机器学习模型的两种选取方式，提高了本方案的实现灵活性。

在第一方面的一种可能实现方式中，机器学习模型为神经网络，服务器中存储的多个模块为神经网络模块，第一客户端上存储有第一适配关系，第一适配关系包括多个适配值，
20 适配值用于表示第一数据集合与第二神经网络之间的适配程度。第一客户端获取至少一个第二机器学习模块之前，方法还包括：第一客户端接收服务器发送的多个神经网络模块。第一客户端获取至少一个机器学习模型包括：第一客户端根据第一适配关系，从至少两个第二神经网络中选取至少一个第一神经网络，该至少一个第一神经网络包括与第一数据集合的适配值高的第一神经网络。其中，适配值高的至少一个第一神经网络可以为适配值最高的 N 个第一神经网络，N 的取值为大于或等于 1 的整数，例如 N 的取值可以为 1、2、3、
25 4、5、6 或其他数值，此处不做限定。或者，适配值高的至少一个第一神经网络可以为大于第四阈值的至少一个第一神经网络，第四阈值的取值可以结合适配值的生成方式、取值范围等因素确定。可选地，该至少一个第一神经网络中还包括第一客户端从至少两个第二神经网络中随机选取的神经网络。

30 本实现方式中，在第一客户端上预先配置有第一适配关系，进而根据第一适配关系，从至少两个第二神经网络中选取与第一数据集合的适配值高的至少一个第一神经网络，以保证了选取的为与第一数据集合的数据特性适配的神经网络，保证了实现不同客户端的神经网络的个性化定制；此外，选取与第一数据集合的数据特性适配，有利于提高训练后的神经网络的准确率。

35 在第一方面的一种可能实现方式中，由于第一适配关系中可能存在空值，则第一客户端可以根据第一适配关系得到第一适配矩阵，第一适配矩阵中的每个元素代表一个适配值，在第一适配关系中不存在空值时，可以通过矩阵分解的方式对第一适配关系进行补全，补全

后的第一适配关系中不再包括空值，从而可以根据补全后的第一适配关系选取与第一数据集合的适配值高的至少一个第一神经网络。

在第一方面的一种可能实现方式中，第一数据集合与一个第二神经网络之间的适配值与第一损失函数的函数值对应，第一损失函数的函数值越小，第一数据集合与一个第二神经网络之间的适配值越大。其中，第一损失函数指示第一数据的预测结果与第一数据的正确结果之间的相似度，第一数据的预测结果通过一个第二神经网络得到，第一数据和第一数据的正确结果基于第一数据集合得到。第一数据可以为第一数据集合中的任一数据，也可以为对第一数据集合进行聚类操作之后得到至少两个数据子集合，第一数据是前述至少两个数据子集合中任一数据子集合的类中心。进一步地，在第一数据为第一数据集合中任一数据的情况下，由于第一数据集合中的数据会用来对第一神经网络执行训练操作（也即第一数据集合中会包括第一训练数据），还可能用来测试训练后的第一神经网络的正确率（也即第一数据集合中可能包括测试数据），还可能用来验证第一神经网络中超参数的正确性（也即第一数据集合中还可能包括验证数据），则第一数据可以是用来训练的数据，也可以是用来测试的数据，还可以是用来验证的数据。

本实现方式中，通过损失函数来计算第一数据集合与一个第一神经网络之间的适配值，方案简单，易于实现，且准确度高。

在第一方面的一种可能实现方式中，第一数据集合与一个第二神经网络之间的适配值与第一相似度对应，第一相似度越大，第一数据集合与一个第二神经网络之间的适配值越大。其中，第一相似度指的是一个第二神经网络和第三神经网络之间的相似度，第三神经网络为上一轮迭代中输出预测结果的准确率最高的神经网络。

本实现方式中，由于第三神经网络为上一轮迭代中输出预测结果的准确率最高的神经网络，且第三神经网络为已经利用第一数据集合训练过的神经网络，也即第三神经网络与第一数据集合的适配程度较高，若一个第一神经网络与该第三神经网络的相似度高，则证明该一个第一神经网络与第一数据集合之间的适配程度高，则适配值就会大；提供了适配值计算的另一种实现方案，提高了本方案的实现灵活性。

在第一方面的一种可能实现方式中，一个第二神经网络和第三神经网络之间的相似度通过以下任一种方式确定：第一客户端将相同数据分别输入至一个第二神经网络和第三神经网络，并对比一个第二神经网络的输出数据与第三神经网络的输出数据之间的相似度。或者，第一客户端计算一个第二神经网络的权重参数矩阵和第三神经网络的权重参数矩阵之间的相似度。其中，两者之间的相似度可以通过计算两者之间的欧式距离、马氏距离、余弦距离、交叉熵或其他方式获得。

本实现方式中，提供了一个第一神经网络和第三神经网络之间的相似度的两种计算方式，提高了本方案的实现灵活性。

在第一方面的一种可能实现方式中，一个第一神经网络的输出数据与第三神经网络的输出数据之间的相似度指的可以为整个第一神经网络的输出数据与整个第三神经网络的输出数据之间的第一相似度。或者，一个第一神经网络的输出数据与第三神经网络的输出数据之间的相似度指的还可以为第一神经网络中各个模块的输出数据与第三神经网络中各个

模块的输出数据之间的相似度，计算各个模块的输出数据之间相似度的乘积，得到整个第一神经网络的输出数据与整个第三神经网络的输出数据之间的相似度。

在第一方面的一种可能实现方式中，机器学习模型为神经网络，方法还包括：第一客户端接收服务器发送的选择器，选择器为用于从多个神经网络模块中选取与第一数据集合的数据特征匹配的至少一个神经网络模块的神经网络。第一客户端根据第一数据集合，将训练数据输入至选择器，得到选择器输出的指示信息。其中，指示信息包括多个神经网络模块中每个神经网络模块被选中的概率，用于指示构建至少一个第一神经网络的神经网络模块；进一步地，若该多个神经网络模块中包括 Z 个神经网络模块，则该指示信息具体可以表现为包括 Z 个元素的向量， Z 个元素中每个元素指示一个神经网络模块被选中的概率。

第一客户端从服务端接收用于构建至少一个第一神经网络的神经网络模块。

本实现方式中，根据第一数据集合，将训练数据输入至选择器，得到选择器输出的指示信息，并根据该指示信息选取用于构建第一神经网络的神经网络模块，选择器为用于从多个神经网络模块中选取与第一数据集合的数据特征匹配的神经网络模块的神经网络，提供了选取构建第一神经网络的神经网络模块的又一种实现方式，提高了本方案的实现灵活性；且通过神经网络来选取，有利于提高神经网络模块的选取过程的准确率。

在第一方面的一种可能实现方式中，针对将训练数据输入选择器的过程。第一客户端可以将第一数据集合中的每个第一训练数据均分别输入选择器一次，以得到与每个第一训练数据对应的指示信息。或者，第一客户端也可以对第一数据集合执行聚类操作，并分别将聚类后的几个类中心（训练数据的一种示例）输入选择器，以得到与每个类中心对应的指示信息。或者，第一客户端也可以对第一数据集合执行聚类操作，并分别从聚类后的几个数据子集合中抽样几个第一训练数据，将抽样得到的第一训练数据（训练数据的一种示例）分别输入选择器，以得到与每个抽样得到的第一训练数据对应的指示信息。

在第一方面的一种可能实现方式中，针对根据指示信息确定构建至少一个第一神经网络的神经网络模块的过程。第一客户端可以初始化一个用于指示每个神经网络模块被选中次数的数组，初始化值为 0，该数组也可以为表格、矩阵或其他形式。第一客户端在得到至少一个指示信息之后，针对每一个指示信息，对于选中概率大于第五阈值的神经网络模块，则数组中与该神经网络模块对应的次数加一，在遍历所有指示信息之后，第一客户端根据该数组统计被选中次数大于第六阈值的至少一个神经网络模块，并将前述至少一个神经网络模块确定为用于构建至少一个第一神经网络的神经网络模块。或者，第一客户端在得到多个指示信息之后，还可以对多个指示信息求平均值，得到一个包括 Z 个元素的向量，向量中每个元素指示一个神经网络模块被选中的概率，进而从 Z 个元素中获取平均值最大的 H 个元素，并将前述 H 个元素指向的 H 个神经网络模块确定为用于构建至少一个第一神经网络的神经网络模块， Z 为大于 1 的整数， H 为大于或等于 1 的整数。

在第一方面的一种可能实现方式中，机器学习模型为神经网络，服务器中存储的多个模块为神经网络模块，第一客户端获取至少一个第一机器学习模型之后，方法还包括：第一客户端计算第一数据集合与至少一个第一神经网络中每个第一神经网络之间的适配值。其中，第一数据集合包括多个第一训练数据，第一训练数据与第一神经网络之间的适配值

—5—

越高，在利用第一训练数据对第一神经网络进行一次训练的过程中，对第一神经网络的权重参数的修改程度越大。进一步地，调整对第一神经网络的权重参数的修改成的方式包括：调整学习率、调整惩罚项的系数或其他方式。学习率越大，在一次训练过程中对第一神经网络的权重参数的修改程度越大，学习率越小，再一次训练过程中对第一神经网络的修改程度越小；也即第一训练数据与第一神经网络之间的适配值越高，在利用该第一训练数据对第一神经网络进行一次训练的过程中的学习率越大。惩罚项的系数越小，在一次训练过程中对第一神经网络的修改程度越大；惩罚项的系数越大，在一次训练过程中对第一神经网络的修改程度越小；也即第一训练数据与第一神经网络之间的适配值越高，在利用该第一训练数据对第一神经网络进行一次训练的过程中的惩罚项系数越小。

本实现方式中，由于同一客户端中不同的训练数据与一个第一神经网络之间的适配程度有高有低，所有训练数据均以固定的能力修改一个第一神经网络的权重参数这是不合理的，一个第一训练数据与第一神经网络之间的适配值越高，则证明该第一神经网络越应该处理该一个第一训练数据，在利用该第一训练数据对第一神经网络进行一次训练的过程中，对第一神经网络的权重参数的修改程度越大，有利于提高一个第一神经网络的训练效率。

在第一方面的一种可能实现方式中，第一客户端计算第一数据集合与至少一个第一神经网络中每个第一神经网络之间的适配值，包括：第一客户端对第一数据集合进行聚类，得到至少两个数据子集合，第一数据子集合为第一数据集合的子集，第一数据子集合为至少两个数据子集合中的任一个；第一客户端根据第一数据子集合和第一损失函数，生成第一数据子集合与一个第一神经网络之间的适配值，第一损失函数的函数值越小，第一数据子集合与一个第一神经网络之间的适配值越大。其中，第一损失函数指示第一数据的预测结果与第一数据的正确结果之间的相似度。第一数据的预测结果通过一个第一神经网络得到，第一数据指的是第一数据子集合中的任一个数据，或者，第一数据指的是第一数据子集合的类中心。第一数据和第一数据的正确结果基于第一数据子集合得到，第一数据子集合与一个第一神经网络之间的适配值被确定为第一数据子集合中每个数据与一个第一神经网络之间的适配值。

本实现方式中，对第一数据集合进行聚类，得到至少两个数据子集合，同一数据子集合中不同训练数据与一个第一神经网络之间的适配值相同，也即同一类的训练数据对一个第一神经网络的修改能力相同，以满足同一客户端中存在至少两个不同数据特性的数据子集合的情况，以进一步提高神经网络的个性化定制能力，有利于提高训练后的神经网络的准确率。

在第一方面的一种可能实现方式中，机器学习模型为神经网络，服务器中存储的多个模块为神经网络模块。第一客户端利用第一数据集合对至少一个第一机器学习模型执行训练操作，包括：第一客户端根据第二损失函数，利用第一数据集合对第一神经网络执行训练操作。其中，第二损失函数包括第一项和第二项，第一项指示第一预测结果与第一训练数据的正确结果之间的相似度，第二项指示第一预测结果与第二预测结果之间的相似度，第二项可以称为惩罚项或约束项。进一步地，第一预测结果为将第一训练数据输入第一神经网络后，由第一神经网络输出的第一训练数据的预测结果；第二预测结果为将第一训练

—6—

数据输入第四神经网络后，由第四神经网络输出的第一训练数据的预测结果。第四神经网络为未执行过训练操作的第一神经网络，也即第四损失函数和第二损失函数的初始状态一致，但在对第二损失函数进行训练的过程中，第四损失函数的权重参数一直不会更新。

本实现方式中，由于第一客户端上的第一数据集合不一定与第一神经网络匹配，在利用第一数据集合对第一神经网络进行训练的过程中，第二损失函数还会指示第一预测结果与第二预测结果之间的相似度，也即避免第一神经网络在训练过程中被过多改动。

在第一方面的一种可能实现方式中，第一数据集合包括多个第一训练数据和每个第一训练数据的正确结果，方法还包括：第一客户端接收服务器发送的选择器，选择器为用于从多个神经网络模块中选取与第一数据集合的数据特征匹配的至少一个第一神经网络模块的神经网络。第一客户端利用第一数据集合对至少一个第一机器学习模型执行训练操作，包括：第一客户端将第一训练数据输入选择器，得到选择器输出的指示信息，指示信息包括多个神经网络模块中每个神经网络模块被选中的概率，用于指示构建第一神经网络的神经网络模块；根据多个神经网络模块，指示信息和第一训练数据，得到第一神经网络输出的第一训练数据的预测结果；根据第三损失函数，对第一神经网络和选择器执行训练操作，其中，第三损失函数指示第一训练数据的预测结果与正确结果之间的相似度，还指示该指示信息的离散程度。方法还包括：第一客户端向服务器发送训练后的选择器。

本实现方式中，在训练构建第一神经网络的神经网络模块的同时，训练选择器，节约了计算机资源；用选择器处理需要处理的数据来训练选择器，有利于提高选择器输出的指示信息的准确率。

第二方面，本申请实施例提供一种机器学习模型的训练方法，可用于人工智能领域中。方法应用于服务器，服务器与多个客户端通信连接，服务器中存储有多个模块，多个模块用于构建机器学习模型，第一客户端为多个客户端中的任一客户端，机器学习模型的训练包括多轮迭代，多轮迭代中的一轮迭代包括：服务器获取与第一客户端对应的至少一个第一机器学习模型，第一客户端为多个客户端中的一个客户端，至少一个第一机器学习模型与第一客户端存储的第一数据集合的数据特性对应。服务器将至少一个第一机器学习模型发送给第一客户端，至少一个第一机器学习模型指示第一客户端利用第一数据集合对至少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型。服务器从第一客户端接收至少一个训练后的第一机器学习模型包括的至少一个更新后的神经网络模块，并根据至少一个更新后的神经网络模块更新存储的神经网络模块的权重参数。

本实现方式中，能够对不同数据特性的训练数据分配不同的神经网络，也即实现了神经网络与数据特性之间的个性化匹配；由于第一客户端为多个客户端中的任一客户端，对多个客户端中的每个客户端均根据客户端存储的训练数据集合的数据特性分配并训练神经网络，能够利用相同数据特性的训练数据训练相同的神经网络，不同数据特性的训练数据训练不同的神经网络，从而不仅实现了神经网络与数据特性之间的个性化匹配，而且有利于提高训练后神经网络的准确率；由服务器选择与各个客户端适配的神经网络，既避免了将所有神经网络模块发送给客户端，以减少对客户端存储资源的浪费；且避免了对客户端计算机资源的占用，有利于提高用户体验。

在第二方面的一种可能实现方式中，多个模块用于构建至少两个第二机器学习模型，至少一个第一机器学习模型为从至少两个第二机器学习模型中选取出来的；或者，用于构建至少一个第一机器学习模型的模块为从多个模块中选取出来的。

在第二方面的一种可能实现方式中，服务器根据至少一个更新后的神经网络模块更新存储的神经网络模块的权重参数，可以包括：由于不同客户端发送的可以存在相同的神经网络模块，则服务器将不同客户端发送的相同的神经网络模块的权重参数进行加权平均，作为服务器中该神经网络模块的权重参数。对于不同客户端中没有重合的神经网络模块，则直接将客户端发送的神经网络模块的参数作为服务器中该神经网络模块的权重参数。其中，相同的神经网络模块指的是具体的神经网络相同，且位于相同的分组中。

在第二方面的一种可能实现方式中，服务器根据至少一个更新后的神经网络模块更新存储的神经网络模块的权重参数，可以包括：若服务器中存在训练数据，则还可以使用模型蒸馏的方法，利用多个客户端发送的多个更新后的神经网络模块，来更新服务器存储的神经网络模块的权重参数。也即使用服务器中存储的训练数据重新训练服务器中存储的多个神经网络模块，训练的目的为拉近服务器中存储的神经网络模块的输出数据与客户端发送的更新后的神经网络模块的输出数据之间的相似度。

在第二方面的一种可能实现方式中，机器学习模型为神经网络，服务器中存储的多个模块为神经网络模块，服务器上存储有第二适配关系，第二适配关系中包括多个适配值，适配值用于表示客户端中存储的训练数据与第二神经网络之间的适配程度。方法还包括：服务器接收第一客户端发送的第一数据集合与至少一个第二神经网络之间的适配值，并更新第二适配关系。服务器获取至少一个第一神经网络包括：服务器根据第二适配关系，从多个第二神经网络中选取至少一个第一神经网络，至少一个第一神经网络包括与第一数据集合的适配值高的神经网络。具体的，服务器可以得到与第二适配关系对应的第二适配矩阵，对第二适配矩阵进行矩阵分解，以得到分解后的神经网络的相似性矩阵和用户的相似性矩阵，神经网络的相似性矩阵和用户的相似性矩阵的乘积与第二适配关系中对对应位置的值需要相似。进而将神经网络的相似性矩阵和用户的相似性矩阵相乘，得到第二补全矩阵，进而根据第二补全矩阵选择与第一数据集合（也即第一客户端）的适配值高的至少一个第一神经网络。可选地，第一客户端选取的至少一个第一神经网络不仅可以包括适配值高的至少一个第一神经网络，还包括随机选取的至少一个第一神经网络。

本实现方式中，在服务器侧配置第二适配关系，由客户端生成适配值并发送给客户端，由服务器根据第二适配关系，选取与第一客户端适配的第一神经网络，既避免了对客户端计算机资源的占用，也避免了客户端的数据的泄露。

在第二方面的一种可能实现方式中，机器学习模型为神经网络，服务器中存储的多个模块为神经网络模块，方法还包括：服务器接收第一客户端发送的第一标识信息，第一标识信息为第一神经网络的标识信息，或者，第一标识信息为构建第一神经网络的神经网络模块的标识信息。服务器将至少一个第一机器学习模型发送给第一客户端，包括：服务器向第一客户端发送第一标识信息指向的第一神经网络，或者，向第一客户端发送第一标识信息指向的构建第一神经网络的神经网络模块。

在第二方面的一种可能实现方式中，机器学习模型为神经网络，服务器中存储的多个模块为神经网络模块，服务器还配置有选择器。方法还包括：服务器接收第一客户端发送的至少一个类中心，对第一数据集合执行聚类操作后，得到至少一个数据子集合，至少一个类中心中的一个类中心为至少一个数据子集合中一个数据子集合的类中心。服务器获取与第一客户端对应的至少一个第一机器学习模型，包括：服务器将类中心分别输入选择器，得到选择器输出的指示信息，并根据指示信息，确定构建至少一个第一神经网络的神经网络模块，指示信息包括多个神经网络模块中每个神经网络模块被选中的概率。服务器将至少一个第一机器学习模型发送给第一客户端，包括：服务器将构建至少一个第一神经网络的神经网络模块发送给第一客户端。

本实现方式中，通过选择器来执行神经网络模块的选择步骤，有利于提高选择过程的准确率，由服务器来执行选择步骤，有利于释放客户端的存储空间，和避免对客户端计算机资源的占用，且仅将类中心发送给服务器，也尽量避免客户端信息的泄露。

在第二方面的一种可能实现方式中，机器学习模型为神经网络，服务器中存储的多个模块为神经网络模块，一个神经网络被分为至少两个子模块，服务器存储的神经网络模块被分为与至少两个子模块对应的至少两个组，同一组中不同的神经网络模块的功能相同。服务器根据至少一个更新后的神经网络模块更新存储的神经网络模块的权重参数之后，方法还包括：服务器计算同一组包括的至少两个神经网络模块中不同的神经网络模块之间的相似度，并将相似度大于预设阈值的两个神经网络模块进行合并。具体的，服务器可以在两个不同的神经网络网络中随机选取一个神经网络模块；或者，若第二神经网络模块和第一神经网络模块具体表现为相同的神经网络，区别仅在于权重参数不同，则服务器还可以将第二神经网络模块和第一神经网络模块的权重参数求平均，以生成合并后的神经网络模块的权重参数。

本实现方式中，将相似度大于预设阈值的两个神经网络模块进行合并，也即将冗余的两个神经网络模块进行合并，不仅降低服务器对多个神经网络模块的管理难度；且避免客户端对相似度大于预设阈值的两个神经网络模块进行重复训练，以减少对客户端计算机资源的浪费。

在第二方面的一种可能实现方式中，不同的神经网络模块包括第二神经网络模块和第一神经网络模块，第二神经网络模块和第一神经网络模块之间的相似度通过以下任一种方式确定：服务器将相同数据分别输入至第二神经网络模块和第一神经网络模块，并对第二神经网络模块的输出数据与第一神经网络模块的输出数据之间的相似度；或者，计算第二神经网络模块的权重参数矩阵和第一神经网络模块的权重参数矩阵之间的相似度。两者之间相似度的计算方式包括但不限于：计算两者之间的欧氏距离、马氏距离、余弦距离或者交叉熵。

本实现方式中，提供了计算两个不同的神经网络模块之间相似度的两种具体实现方式，则用户可以结合实际情况灵活选择，提高了本方案的实现灵活性。

对于本申请实施例第二方面以及第二方面的各种可能实现方式中名词的具体含义，以及每种可能实现方式所带来的有益效果，均可以参考第一方面中各种可能的实现方式中的

描述，此处不再一一赘述。

第三方面，本申请实施例提供一种数据处理方法，可用于人工智能领域中。服务器获取与第二客户端存储的第二数据集合的数据特性对应的至少一个第三神经网络，向第二客户端发送至少一个第三神经网络，该至少一个第三神经网络用于供客户端生成待处理数据的预测结果。

在第三方面的一种实现方式中，服务器获取与第二客户端存储的第二数据集合的数据特性对应的至少一个第三神经网络，可以以下三项中的任一项或多项：服务器接收到至少一个第二类中心，将至少一个第二类中心分别输入选择器中，以得到用于构建至少一个第三神经网络的神经网络模块，每个第二类中心为一个第二数据子集合的类中心，至少一个第二数据子集合为对第二数据集合执行聚类操作得到的。或者，服务器根据第二客户端的标识信息和第二适配关系，从至少两个第二神经网络中选取至少一个第三神经网络，至少一个第三神经网络中包括与第二数据集合适配高的神经网络。或者，服务器从多个第二神经网络中随机选取至少一个第三神经网络。

对于本申请实施例第三方面以及第三方面的各种可能实现方式中步骤的具体实现方式、每种可能实现方式中名词的具体含义，以及每种可能实现方式所带来的有益效果，均可以参考第一方面中各种可能的实现方式中的描述，此处不再一一赘述。

第四方面，本申请实施例提供一种数据处理方法，可用于人工智能领域中。第二客户端得到与第二客户端存储的第二数据集合的数据特性对应的第二标识信息，并向服务器发送获取请求，获取请求中携带有第二标识信息，第二标识信息为第三神经网络的标识信息，或者，第二标识信息为构建第三神经网络的神经网络模块的标识信息。第二客户端接收第二标识信息指向的一个或多个第三神经网络，或者，接收第二标识信息指向的用于构建一个或多个第一神经网络的神经网络模块。

对于本申请实施例第四方面以及第四方面的各种可能实现方式中步骤的具体实现方式、每种可能实现方式中名词的具体含义，以及每种可能实现方式所带来的有益效果，均可以参考第一方面中各种可能的实现方式中的描述，此处不再一一赘述。

第五方面，本申请实施例提供一种机器学习模型的训练装置，可用于人工智能领域中。装置应用于第一客户端，多个客户端与服务器通信连接，服务器中存储有多个模块，多个模块用于构建机器学习模型，第一客户端为多个客户端中的任一客户端。机器学习模型的训练装置用于执行多轮迭代，机器学习模型的训练装置包括：获取单元、训练单元和发送单元，在多轮迭代中的一轮迭代中，获取单元，用于获取至少一个第一机器学习模型，至少一个第一机器学习模型为根据第一客户端存储的第一训练数据集合的数据特性选取出来的；训练单元，用于利用第一数据集合对至少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型；发送单元，用于将至少一个训练后的第一机器学习模型包括的至少一个更新后的模块发送给服务器，更新后的模块用于供服务器更新存储的模块的权重参数。

本申请实施例第五方面中，机器学习模型的训练装置还可以用于实现第一方面各种可能实现方式中第一客户端执行的步骤，对于本申请实施例第五方面以及第五方面的各种可

- 10 -

能实现方式中某些步骤的具体实现方式，以及每种可能实现方式所带来的有益效果，均可以参考第一方面中各种可能的实现方式中的描述，此处不再一一赘述。

第六方面，本申请实施例提供一种机器学习模型的训练装置，可用于人工智能领域中。装置应用于服务器，服务器与多个客户端通信连接，服务器中存储有多个模块，多个模块用于构建机器学习模型，第一客户端为多个客户端中的任一客户端。机器学习模型的训练装置用于执行多轮迭代，机器学习模型的训练装置包括：获取单元、发送单元和更新单元，在多轮迭代中的一轮迭代中，获取单元，用于获取与第一客户端对应的至少一个第一机器学习模型，第一客户端为多个客户端中的一个客户端，至少一个第一机器学习模型与第一客户端存储的第一数据集合的数据特性对应；发送单元，用于将至少一个第一机器学习模型发送给第一客户端，至少一个第一机器学习模型指示第一客户端利用第一数据集合对至少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型；更新单元，用于从第一客户端接收至少一个训练后的第一机器学习模型包括的至少一个更新后的神经网络模块，并根据至少一个更新后的神经网络模块更新存储的神经网络模块的权重参数。

本申请实施例第六方面中，机器学习模型的训练装置还可以用于实现第二方面各种可能实现方式中服务器执行的步骤，对于本申请实施例第六方面以及第六方面的各种可能实现方式中某些步骤的具体实现方式，以及每种可能实现方式所带来的有益效果，均可以参考第二方面中各种可能的实现方式中的描述，此处不再一一赘述。

第七方面，本申请实施例提供了一种服务器，可以包括处理器，处理器和存储器耦合，存储器存储有程序指令，当存储器存储的程序指令被处理器执行时实现上述第一方面的机器学习模型的训练的方法，或者，当存储器存储的程序指令被处理器执行时实现上述第一方面的机器学习模型的训练的方法。对于处理器执行第一方面的各个可能实现方式中第一客户端执行的步骤，或者，对于处理器执行第二方面的各个可能实现方式中服务器执行的步骤，具体均可以参阅第一方面或第二方面，此处不再赘述。

第八方面，本申请实施例提供了一种终端设备，可以包括处理器，处理器和存储器耦合，存储器存储有程序指令，当存储器存储的程序指令被处理器执行时实现上述第一方面的机器学习模型的训练的方法。对于处理器执行第一方面的各个可能实现方式中第一客户端执行的步骤，具体均可以参阅第一方面，此处不再赘述。

第九方面，本申请实施例提供了一种计算机可读存储介质，计算机可读存储介质中存储有计算机程序，当该计算机程序在计算机上运行时，使得计算机执行上述第一方面的机器学习模型的训练的方法，或者，使得计算机执行上述第二方面的机器学习模型的训练方法。

第十方面，本申请实施例提供了一种电路系统，电路系统包括处理电路，处理电路配置为执行上述第一方面的机器学习模型的训练的方法，或者，处理电路配置为执行上述第二方面的机器学习模型的训练方法。

第十一方面，本申请实施例提供了一种计算机程序，当该计算机程序在计算机上运行时，使得计算机执行上述第一方面的机器学习模型的训练的方法，或者，使得计算机执行

上述第二方面的机器学习模型的训练方法。

第十二方面，本申请实施例提供了一种芯片系统，该芯片系统包括处理器，用于支持训练设备或执行设备实现上述方面中所涉及的功能，例如，发送或处理上述方法中所涉及的数据和/或信息。在一种可能的设计中，芯片系统还包括存储器，存储器，用于保存服务器或通信设备必要的程序指令和数据。该芯片系统，可以由芯片构成，也可以包括芯片和其他分立器件。

附图说明

图 1 为本申请实施例提供的人工智能主体框架的一种结构示意图；

图 2 为本申请实施例提供的机器模型的训练系统的一种系统架构图；

图 3 为本申请实施例提供的机器学习模型的训练方法中多个不同的数据集合的一种示意图；

图 4 为本申请实施例提供的机器学习模型的训练方法的一种流程示意图；

图 5 为本申请实施例提供的机器学习模型的训练方法中多个神经网络模块的一种示意图；

图 6 为本申请实施例提供的机器学习模型的训练方法中多个神经网络模块的另一种示意图；

图 7 为本申请实施例提供的机器学习模型的训练方法中第二神经网络的三种结构示意图；

图 8 为本申请实施例提供的机器学习模型的训练方法中第二神经网络的另一种结构示意图；

图 9 为本申请实施例提供的机器学习模型的训练方法的另一种流程示意图；

图 10 为本申请实施例提供的机器学习模型的训练方法的又一种流程示意图；

图 11 为本申请实施例提供的机器学习模型的训练方法的再一种流程示意图；

图 12 为本申请实施例提供的机器学习模型的训练方法的又一种流程示意图；

图 13 为本申请实施例提供的数据处理方法的一种流程示意图；

图 14 为本申请实施例提供的机器学习模型的训练装置的一种结构示意图；

图 15 为本申请实施例提供的机器学习模型的训练装置的另一种结构示意图；

图 16 为本申请实施例提供的机器学习模型的训练装置的又一种结构示意图；

图 17 为本申请实施例提供的机器学习模型的训练装置的再一种结构示意图；

图 18 为本申请实施例提供的训练设备的一种结构示意图；

图 19 为本申请实施例提供的执行设备的一种结构示意图；

图 20 为本申请实施例提供的芯片的一种结构示意图。

具体实施方式

本申请实施例提供了一种机器学习模型的训练的方法以及相关设备，对不同数据特性的训练数据分配不同的神经网络，实现了神经网络与数据特性之间的个性化匹配；每个客户端均根据客户端存储的训练数据集合的数据特性分配并训练神经网络，能够利用相同数

据特性的训练数据训练相同的神经网络，有利于提高训练后神经网络的准确率。

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”等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术语在适当情况下可以互换，这仅仅是描述本申请的实施例中对相同属性的对象在描述时所采用的区分方式。

5 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，以便包含一系列单元的过程、方法、系统、产品或设备不必限于那些单元，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单元。

下面结合附图，对本申请的实施例进行描述。本领域普通技术人员可知，随着技术的发展和场景的出现，本申请实施例提供的技术方案对于类似的技术问题，同样适用。

10 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”

15 的凝练过程。“IT 价值链”从人工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提供，作为示例，该智能芯片

20 包括中央处理器（central processing unit, CPU）、神经网络处理器（neural-network processing unit, NPU）、图形处理器（graphics processing unit, GPU）、专用集成电路（application specific integrated circuit, ASIC）、现场可编程逻辑门阵列（field programmable gate array, FPGA）等硬件加速芯片；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

25 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指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。

35

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，图像的分类、图像的个性化管理、电池充电个性化管理、文本分析、计算机视觉的处理、语音识别等等。

(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、智能制造、智能交通、智能家居、智能医疗、智能安防、自动驾驶、智慧城市等。

本申请实施例主要用于对各种应用场景中采用到的机器学习模型进行训练，训练后的机器学习模型可以应用上述各种应用领域中以实现分类、回归或其他功能，训练后的机器学习模型的处理对象可以为图像样本、离散数据样本、文本样本或语音样本等，此处不做穷举。其中，机器学习模型具体可以表现为神经网络、线性模型或其他类型的机器学习模型等，对应的，组成机器学习模型的多个模块具体可以表现为神经网络模块、线性模型模块或组成其他类型的机器学习模型的模块等，此处不做穷举。在后续实施例中，仅以机器学习模型表现为神经网络为例进行说明，对于机器学习模型表现为除神经网络之外的其他类型时可类推理解，本申请实施例中不再赘述。

进一步地，本申请实施例可以应用于联邦学习和分布式训练两种训练方式中。为了便于理解，请先参阅图 2，图 2 为本申请实施例提供的机器模型的训练系统的一种系统架构图，先结合图 2 介绍联邦学习和分布式训练这两种训练方式。机器模型的训练系统中包括服务器 100 和多个客户端 200，服务器 100 和多个客户端 200 通信连接。其中，服务器 100 具体可以表现为一个服务器，也可以表现为由多个服务器组成的服务器集群，虽然图 2 中仅示出了一个服务器 100 和三个客户端 200，但实际情况中服务器 100 和客户端 200 的数量可以结合实际需求确定。客户端 200 可以配置于终端设备上，也可以配置于服务器上，此处不做限定。

具体的，在训练阶段，在采用联邦学习这一训练方式的情况下，服务器 100 上存储有多个神经网络模块，多个神经网络模块用于构建至少两个第二神经网络；每个客户端 200 上存储有数据集合，客户端 200 上存储的数据集合可以用来对神经网络执行训练操作。多个客户端 200 中的第一客户端中存储有第一数据集合，第一客户端为多个客户端 200 中的任一个客户端，在一轮迭代过程中，第一客户端获取与第一数据集合的数据特性适配的至少一个第一神经网络，并利用第一数据集合对该至少一个第一神经网络执行训练操作，得到至少一个训练后的第一神经网络，再将至少一个训练后的第一机器学习模型包括的至少一个更新后的神经网络模块发送给服务器 100。多个客户端 200 中每个客户端 200 均可以执行前述操作，则服务器 100 可以接收到多个客户端 200 发送的多个更新后的神经网络模块，服务器 100 根据接收到的多个更新后的神经网络模块更新存储的神经网络模块的权重参数，以完成一轮迭代的过程。通过多轮迭代过程以实现服务器 100 存储的多个神经网络模块的权重参数的更新。

分布式训练与联邦学习的区别在于，用于对神经网络进行训练的数据是由服务器 100 发送给各个客户端 200 的。在分布式训练这种训练方式中，服务器 100 存储有用于构建至

少两个第二神经网络的多个神经网络模块之外，还存储有数据集合；服务器 100 先对存储的数据集合进行聚类操作，得到聚类后的多个数据子集合，进而服务器 100 分别给每个客户端 200 发送与一个类或几个类对应的数据子集合，也即不同客户端 200 中可以存储有不同数据特性的数据子集合。进一步地，针对聚类的过程，服务器 100 可以对存储的整个数据集合进行聚类，也可以先根据数据集合中每个数据的正确标签将整个数据集合分成不同的数据子集合，进而依次对分类后的各个数据子集合执行聚类操作，以得到多个聚类后的数据子集合。针对数据分发的步骤，服务器 100 可以将聚类后的数据子集合直接发送给客户端 200，也可以从至少一个聚类后的数据子集合中抽样出一些数据发送给客户端 200 等，此处不做限定。在服务器 100 将用于执行训练操作的数据部署于各个客户端 200 之后，也需要通过多轮迭代以实现对服务器 100 存储的多个神经网络模块的权重参数的更新，每轮迭代过程的实现方法与联邦学习中每轮迭代过程的实现方法相同，此处不做赘述。

在推理阶段，客户端 200 获取与客户端 200 存储的数据集合的数据特性对应的第三神经网络，进而利用获取到的神经网络生成输入数据的预测结果。为了更直观地理解“数据集合的数据特性”这一概念，请参阅图 3，图 3 为本申请实施例提供的机器学习模型的训练方法中多个不同的数据集合的一种示意图。图 3 中以神经网络的任务为图像分类为例，示出了 4 个客户端中存储的数据集合，第一个客户端存储的数据集合为狗的图像集合，则第一客户端需要神经网络执行的任务为对狗进行分类；第二个客户端存储的数据集合为狼的图像集合，则第一客户端需要神经网络执行的任务为对狼进行分类；第三个客户端存储的数据集合中包括的狗的图像集合和沙发的图像集合，则第三客户端需要两个不同的神经网络，一个神经网络执行的任务为对狗进行分类，另一个神经网络执行的任务为对沙发进行分类。其中，狗的图像集合、狼的图像集合和沙发的图像集合的数据特性各不相同，第一客户端存储的狗的图像集合和第三客户端中均存储的狗的图像集合的数据特性相同。

以下对本申请实施例提供的机器学习模型的训练方法进行详细介绍，由于该方法对训练阶段和推理阶段均有影响，而训练阶段和推理阶段的实现流程有所不同，下面分别对前述两个阶段的具体实现流程进行描述。

一、训练阶段

本申请实施例中，服务器中存储有多个神经网络模块，而在一次迭代中，针对多个客户端中的一个第一客户端，需要获取与第一数据集合适配的至少一个第一神经网络，前述至少一个第一神经网络的选取操作可以由服务器执行，也可以由客户端执行，前述两种方式的实现流程不同。进一步地，服务器或客户端上可以根据神经网络与数据集合之间的适配关系来执行前述选取操作，也即服务器或客户端先利用多个神经网络模块构建多个第二神经网络，再从多个第二神经网络中选取与一个客户端中存储的数据集合的数据特性适配的至少一个第一神经网络。或者，服务器或客户端也可以利用选择器（一种神经网络）来执行前述选取操作，也即服务器或客户端先利用选择器从多个神经网络模块中选取与一个客户端中存储的数据集合的数据特性适配的至少一个神经网络模块，再利用选取出的神经网络模块构建至少一个第一神经网络。前述两种方式的实现流程也有所不同，以下分别进行描述。

(1) 由客户端根据第一适配关系选择与客户端存储的数据集合的数据特性适配的第一神经网络

具体的, 请参阅图 4, 图 4 为本申请实施例提供的机器学习模型的训练方法的一种流程示意图, 方法可以包括:

401、第一客户端接收服务器发送的多个神经网络模块, 并根据多个神经网络模块构建至少两个第二神经网络。

本申请的一些实施例中, 服务器会将存储的多个神经网络模块发送给第一客户端, 对应的, 第一客户端接收到多个神经网络模块, 并根据多个神经网络模块构建至少两个第二神经网络。前述多个神经网络模块可以为预先训练过的神经网络模块, 也可以为完全没训练过的神经网络模块。

其中, 至少两个第二神经网络中的每个第二神经网络均可以被分为至少两个子模块, 该多个神经网络模块包括与该至少两个子模块对应的至少两个组。不同组中可以包括相同数量的神经网络模块, 也可以包括不同数量的神经网络模块; 同一组中不同的神经网络模块的功能相同, 作为示例, 例如同一组神经网络模块的功能均为特征提取, 或者, 同一组的神经网络模块的功能均为特征变换, 或者, 同一组的神经网络模块的功能均为分类等等, 此处不做穷举。可选地, 在多轮迭代训练的过程中, 服务器可以将新的神经网络模块添加至该多个神经网络模块中, 也可以对该多个神经网络模块执行删减操作等。

不同的神经网络模块可以表现为不同的神经网络, 作为示例, 例如第一组有 3 个神经网络模块, 第一个神经网络模块采用 3 层的多层感知机 (multilayer perceptron, MLP), 第二个神经网络模块采用 2 层的 MLP, 第三个神经网络模块采用 2 层的卷积神经网络 (convolutional neural networks, CNN) 等。或者, 不同的神经网络模块也可以为相同的神经网络, 但权重参数不同。作为示例, 例如第一组有 3 个神经网络模块, 第一个神经网络模块采用 2 层的多层感知机 (multilayer perceptron, MLP), 第二个神经网络模块采用 2 层的 MLP, 第三个神经网络模块采用 2 层的卷积神经网络 (convolutional neural networks, CNN), 但第一个神经网络模块和第二个神经网络模块的权重参数不同, 应理解, 此处举例仅为方便理解本方案, 不用于限定本方案

为更直观的理解本方案, 请参阅图 5 和图 6, 图 5 和图 6 为本申请实施例提供的机器学习模型的训练方法中多个神经网络模块的两种示意图。图 5 和图 6 中以不同组中包括不同数量的神经网络模块, 且多个神经网络模块组成的层状结构共有 4 层 (也即对应将多个神经网络模块分为 4 个组), 分别为 SGL1、SGL2、SGL3、SGL4 为例。第一层 (也即第一组神经网络模块) 中包括 3 个神经网络模块, 第一个神经网络模块 SGL1M1 采用 3 层的 MLP, 第二个神经网络模块 SGL1M2 采用 2 层的 MLP, 第三个神经网络模块 SGL1M3 为 2 层的 CNN。第二层 (也即第二组神经网络模块) 中包括 4 个神经网络模块, 第一个神经网络模块 SGL2M1 采用 3 层的 MLP, 第二个神经网络模块 SGL2M2 采用 2 层的 MLP, 第三个神经网络模块 SGL2M3 为 3 层的 CNN, 第四个神经网络模块 SGL2M4 为 2 层的 CNN。第三层 (也即第三组神经网络模块) 中包括 2 个神经网络模块, 第一个神经网络模块 SGL3M1 采用 2 层的 MLP, 第二个神经网络模块 SGL3M2 采用 2 层的 CNN。第四层 (也即第四组

神经网络模块)中包括4个神经网络模块,第一个神经网络模块 SGL4M1 采用3层的MLP,第二个神经网络模块 SGL4M2 采用2层的MLP,第三个神经网络模块 SGL4M3 为1层的CNN+2层的MLP,第四个神经网络模块 SGL4M4 为1层的CNN+1层的MLP。

请继续参阅图6,多组神经网络模块层状结构主模型为二叉树树状结构。第一层(也即第一组神经网络模块)中包括的神经网络模块为 SGL1M1;第二层(也即第二组神经网络模块)包括的神经网络模块中,从左到右依次为 SGL2M1、SGL2M2;第三层(也即第三组神经网络模块)包括的神经网络模块中,从左到右依次为 SGL3M1、SGL3M2、SGL3M3、SGL3M4;第四层(也即第四组神经网络模块)包括的神经网络模块中,从左到右依次为 SGL4M1、SGL4M2、SGL4M3、SGL4M4、SGL4M5、SGL4M6、SGL4M7、SGL4M8。在利用图6中示出的多个神经网络模块构建多个第二神经网络时,可以从树状结构的多个神经网络模块中进行选择,作为示例,例如第二神经网络可以为 SGL1M1 + SGL2M1 + SGL3M1 + SGL4M1,作为另一示例,例如第二神经网络可以为 SGL1M1 + SGL2M2 + SGL3M3 + SGL4M5等,此处不做穷举。应理解,图5和图6中的示例均仅为方便理解本方案,不用于限定本方案。

针对多个神经网络模块构建至少两个第二神经网络的过程。第一客户端在接收到服务器发送的多个神经网络模块之后,可以从每组中只选择一个神经网络模块来构建第二神经网络,也即第二神经网络为单支路的。第一客户端也可以从每组中选择至少两个神经网络模块来构建第二神经网络,也即一个第二神经网络中包括多个支路。第一客户端还可以从某一组神经网络模块中不选择任何神经网络模块。

为更直观的理解本方案,结合图5进行举例,请参阅图7和图8,图7和图8为本申请实施例提供的机器学习模型的训练方法中第二神经网络的四种结构示意图。图7和图8中均以在图5示出的多组神经网络模块的基础上,构建第二神经网络。请先参阅图7,图7中以第一层的选取的神经网络模块为 SGL1M1,第二层选取的神经网络模块为 SGL2M1,第三层选取的神经网络模块为 SGL3M1 和 SGL3M2,第4层选取的神经网络模块为 SGL4M1。在一种实现方式中,如图7的(a)子示意图所示, SGL2M1 的输出分为作为 SGL3M1 和 SGL3M2 的输入, SGL3M1 和 SGL3M2 的输出加权平均作为 SGL4M1 的输入。在另一种实现方式中,如图7的(b)子示意图所示,在第二神经网络的第二层与第三层之间增设转化层 TL, SGL2M1 的输出分为作为 SGL3M1 和 SGL3M2 的输入, SGL3M1 和 SGL3M2 的输出作为转化层 TL 的输入,转化层 TL 的输出作为 SGL4M1 的输入。为更直观地理解图7的(b)子示意图中示出的第二神经网络,以下公开了第二神经网络的计算过程:

$$h_1 = \text{SGL1M1}(x);$$

$$h_2 = \text{SGL2M1}(h_1);$$

$$h_{3_1} = \text{SGL3M1}(h_2);$$

$$h_{3_2} = \text{SGL3M2}(h_2);$$

$$h_{\text{TL}} = \text{TL}(h_{3_1}, h_{3_2});$$

$$y = \text{SGL4M1}(h_{\text{TL}});$$

其中, x 代表输入数据, h1代表SGL1M1的输出, 将h1输入SGL2M1, 得到SGL2M1的输

-17-

出 h_2 ，将SGL2M1的输出分别输入SGL3M1和SGL3M2，分别得到 h_{3_1} 和 h_{3_2} ，将 h_{3_1} 和 h_{3_2} 输入至转化层 TL，得到转化层 TL 输出的 h_{TL} ，将 h_{TL} 输入 SGL4M1，得到整个第二神经网络输出的 x 的预测结果 y 。

在另一种实现方式中，SGL2M1 的输出分为作为 SGL3M1 和 SGL3M2 的输入，SGL3M1 和 SGL3M2 的输出作为转化层 TL 的输入通过 SGL2M 作为选择信号，TL 的输出作为 SGL4M1 的输入。为更直观地理解图 7 的 (c) 子示意图中示出的第二神经网络，以下公开了第二神经网络的计算过程：

$$\begin{aligned} h_1 &= \text{SGL1M1}(x); \\ h_2 &= \text{SGL2M1}(h_1); \\ h_{3_1} &= \text{SGL3M1}(h_2); \\ h_{3_2} &= \text{SGL3M2}(h_2); \\ h_{TL} &= h_{3_1} * \text{TL}(h_2) + h_{3_2} * (1 - \text{TL}(h_2)); \\ y &= \text{SGL4M1}(h_{TL}); \end{aligned}$$

其中， x 、 h_1 、 h_2 、 h_{3_1} 和 h_{3_2} 的含义与上一实现方式中的含义类似，可参阅理解；区别在于 h_{TL} 的生成方式不同，本实现方式中 h_{TL} 的生成方式可参阅上述公式， y 代表通过本实现方式中的第二神经网络输出的 x 的预测结果，应理解，图 7 中的示例均仅为方便理解本方案，不用于限定本方案。

请继续参阅图 8，图 8 示出的第二神经网络中，第一层选取的神经网络模块为 SGL1M1，第二层选取的神经网络模块为 SGL2M1，第三层空缺，第四层选取的神经网络模块为 SGL4M1，参阅图 8 的 (a) 子示意图，由于 SGL2M1 的输出维度和 SGL4M1 的输入维度不相等，因此不能直接相连。则可以在第二层和第三层之间增设中间转化层，将 SGL2M1 的输出作为中间转化层的输入，中间转化层的输出作为 SGL4M1 的输入，应理解，图 8 中的示例均仅为方便理解本方案，不用于限定本方案。

402、第一客户端从至少两个第二神经网络中选取至少一个第一神经网络。

本申请的一些实施例中，第一客户端在得到多个神经网络模块之后，在一种情况下，若第一客户端为初次选取第一神经网络，则第一客户端可以为从至少两个第二神经网络中随机选取至少两个第一神经网络。随机选取的第一神经网络的数量可以预先设定好，作为示例，例如 4 个、5 个或 6 个等，此处不做限定。

可选地，还可以为在第一适配关系中的适配值数量未超过第一阈值的情况下，第一客户端从至少两个第二神经网络中随机选取至少两个第一神经网络。作为示例，利于第一阈值的取值可以为百分之十、百分之十二、百分之十五等等，此处不做限定。

在另一种情况下，在第一客户端不是初次选取第一神经网络的情况下，第一客户端可以根据第一适配关系，从至少两个第二神经网络中选取与第一数据集合的适配值高的至少一个第一神经网络。本申请实施例中，在第一客户端上预先配置有第一适配关系，进而根据第一适配关系，从至少两个第二神经网络中选取与第一数据集合的适配值高的至少一个第一神经网络，以保证了选取的为与第一数据集合的数据特性适配的神经网络，保证了实现不同客户端的神经网络的个性化定制；此外，选取与第一数据集合的数据特性适配，有

利于提高训练后的神经网络的准确率。

其中，适配值高的至少一个第一神经网络可以为适配值最高的 N 个第一神经网络， N 的取值为大于或等于 1 的整数，作为示例，例如 N 的取值可以为 1、2、3、4、5、6 或其他数值等等，此处不做限定。或者，适配值高的至少一个第一神经网络可以为大于第四阈值的至少一个第一神经网络，第四阈值的取值可以结合适配值的生成方式、取值范围等因素灵活确定，此处不做限定。

具体的，第一客户端可以根据第一适配关系得到第一适配矩阵，第一适配矩阵中的每个元素代表一个适配值，在第一适配关系中存在空值时，可以通过矩阵分解的方式对第一适配关系进行补全，补全后的第一适配关系中不再包括空值，从而可以根据补全后的第一适配关系选取与第一数据集合的适配值高的至少一个第一神经网络。

可选地，在第一客户端不是初次选取第一神经网络，且第一适配关系中适配值的数量大于第一阈值的情况下，第一客户端根据第一适配关系，从至少两个第二神经网络中选取与至少一个第一神经网络，该至少一个第一神经网络中包括与第一数据集合的适配值高的神经网络。

可选地，第一客户端选取的至少一个第一神经网络不仅可以包括适配值高的至少一个第一神经网络，还包括随机选取的至少一个第一神经网络。

403、第一客户端计算第一数据集合与第一神经网络之间的适配值。

本申请的一些实施例中，第一客户端上可以通过表格、矩阵、索引、数组等形式存储有第一适配关系，该第一适配关系包括多个适配值，该适配值用于表示第一数据集合与第二神经网络之间的适配程度；第一适配关系中还可以包括每个第二神经网络的标识信息，用于唯一标识每个第一网络。为更直观的理解本方案，以下以表格为例来展示第一适配关系。

表 1

	神经网络 ID1	神经网络 ID2	神经网络 ID3	神经网络 ID4	...	神经网络 ID95	神经网络 ID96
适配值							

其中，ID 为身份标识 (Identity) 的缩写，表 1 中以多个神经网络模块被分为 4 组，第一组神经网络模块中包括 4 个神经网络模块、第二组神经网络模块中包括 3 个神经网络模块，第三组神经网络模块中包括 2 个神经网络模块，第四组神经网络模块中包括 4 个神经网络模块，且第一客户端从每组中只选择一个神经网络模块来构建第二神经网络为例，则共可以构建 96 个第二神经网络，对应的，第一适配关系中有与 96 个第二神经网络一一对应的 96 个标识信息，需要说明的是，第一适配关系中不一定会包括第一数据集合与每个第二神经网络之间的适配值，第一客户端可以根据已有的适配值，通过矩阵分解等方法计算得到第一数据集合与每个第二神经网络之间的适配值，具体计算过程在后续步骤中进行说明。

第一客户端上存储由第一数据集合，第一数据集合中包括多个第一训练数据和每个第一训练数据的准确结果。第一客户端在获取到至少一个第一神经网络之后，需要计算第一

数据集合与第一神经网络之间的适配值，并将通过步骤 403 计算得到的适配值写入第一适配关系，也即根据通过步骤 403 计算得到的适配值更新第一适配关系。对于适配值的生成方式也可参阅后续步骤中的描述，此处先不做介绍。

其中，第一数据集合与一个第一神经网络之间的适配值可通过如下两种方式计算得到。

（一）、通过计算损失函数的函数值得到适配值

本实施例中，第一数据集合与一个第一神经网络之间的适配值与第一损失函数的函数值对应。其中，第一损失函数指示第一数据的预测结果与第一数据的正确结果之间的相似度。第一数据的预测结果通过一个第一神经网络得到，第一数据和第一数据的正确结果基于第一数据集合得到。第一损失函数的函数值越大，第一数据集合与一个第一神经网络之间的适配值越小；第一损失函数的函数值越小，第一数据集合与一个第一神经网络之间的适配值越大。本申请实施例中，通过损失函数来计算第一数据集合与一个第一神经网络之间的适配值，方案简单，易于实现，且准确度高。

具体的，在一种实现方式中，第一客户端对第一数据集合进行聚类，得到至少一个数据子集合，第一数据子集合为第一数据集合的子集，第一数据子集合为至少一个数据子集合中的任一个。进而第一客户端根据第一数据子集合和第一损失函数，生成第一数据子集合与一个第一神经网络之间的适配值。其中，第一损失函数指示第一数据的预测结果与第一数据的正确结果之间的相似度；第一数据的预测结果通过一个第一神经网络得到，第一数据和第一数据的正确结果基于第一数据子集合得到。第一损失函数的函数值越小，第一数据子集合与一个第一神经网络之间的适配值越大。第一客户端对至少两个数据子集合中每个数据子集合均执行前述操作，以得到每个数据子集合与一个第一神经网络之间的适配值。第一客户端可以将多个数据子集合与该一个第一神经网络之间的适配值求平均值，以得到整个第一数据集合与该一个第一神经网络之间的适配值，并更新第一适配关系。

针对第一数据子集合与一个第一神经网络之间的适配值的生成过程。更具体的，在一种情况下，第一数据指的是第一数据子集合中的任一个数据，由于第一数据子集合中的数据会用来对第一神经网络执行训练操作（也即第一数据子集合中会包括第一训练数据），还可能用来测试训练后的第一神经网络的正确率（也即第一数据子集合中可能包括测试数据），还可能用来验证第一神经网络中超参数的正确性（也即第一数据子集合中还可能包括验证数据），则第一数据可以是用来训练的数据，也可以是用来测试的数据，还可以是用来验证的数据。第一客户端可以将第一数据子集合中的每个第一数据输入至该一个第一神经网络，得到该一个第一神经网络输出的第一数据的预测结果，并根据第一数据的预测结果和第一数据的正确结果计算得到第一损失函数的函数值，第一客户端对第一数据子集合中的每个第一数据均执行前述操作，得到多个损失函数的函数值，对前述多个损失函数的函数值求平均值，以得到整个第一数据子集合与该一个第一神经网络之间的适配值。进一步地，第一客户端可以将多个损失函数的函数值的平均值的倒数，确定为整个第一数据子集合与该一个第一神经网络之间的适配值。

在另一种情况下，第一数据指的是第一数据子集合的类中心，第一客户端还可以根据第一数据子集合，计算第一数据子集合中所有数据的类中心，将该类中心输入该一个第一

神经网络，得到该一个第一神经网络输出的第一数据的预测结果；第一客户端对第一数据子集合中所有的数据的结果进行求平均值，得到一个第一数据的结果，进而计算第一损失函数的函数值，以得到整个第一数据子集合与该一个第一神经网络之间的适配值。进一步地，第一客户端可以前述一个损失函数的函数值取倒数，并将该倒数确定为整个第一数据子集合与该一个第一神经网络之间的适配值。

可选地，结合步骤 404 中的描述，在利用第一数据集合对第一神经网络执行训练操作的过程中，可以将第一数据子集合与一个第一神经网络之间的适配值被确定为第一数据子集合中每个训练数据与一个第一神经网络之间的适配值，不同数据子集合中的训练数据与该一个第一神经网络之间的适配值不同。本申请实施例中，对第一数据集合进行聚类，得到至少两个数据子集合，同一数据子集合中不同训练数据与一个第一神经网络之间的适配值相同，也即同一类的训练数据对一个第一神经网络的修改能力相同，以满足同一客户端中存在至少两个不同数据特性的数据子集合的情况，以进一步提高神经网络的个性化定制能力，有利于提高训练后的神经网络的准确率。

在另一种实现方式中，第一数据指的是第一数据集合中的任一个数据，由于第一数据集合中的数据会用来对第一神经网络执行训练操作（也即第一数据集合中会包括第一训练数据），还可能用来测试训练后的第一神经网络的正确率（也即第一数据集合中可能包括测试数据），还可能用来验证第一神经网络中超参数的正确性（也即第一数据集合中还可能包括验证数据），则第一数据可以是用来训练的数据，也可以是用来测试的数据，还可以是用来验证的数据。第一客户端可以将第一数据集合中的每个第一数据逐次输入到该一个第一神经网络，并得到与每个第一数据对应的损失函数的函数值，将多个损失函数的函数值进行求平均值，以得到与整个第一数据集合对应的一个损失函数的函数值，进而根据前述与整个第一数据集合对应的一个损失函数的函数值，生成整个第一数据集合与该一个第一神经网络之间的适配值。

可选地，结合步骤 404 中的描述，在利用第一数据集合对第一神经网络执行训练操作的过程中，将第一数据集合与一个第一神经网络之间的适配值被确定为第一数据集合中每个第一数据与该一个第一神经网络之间的适配值，也即第一数据集合中所有第一训练数据与该一个第一神经网络之间的适配值均相同。

在另一种实现方式中，第一客户端可以将第一数据集合中的每个第一数据逐次输入到该一个第一神经网络，并得到与每个第一数据对应的损失函数的函数值，并生成每个第一数据与该一个第一神经网络之间的适配值，进而对所有第一数据与该一个第一神经网络之间的适配值求平均值，以得到整个第一数据集合与该一个第一神经网络之间的适配值。可选地，结合步骤 404 中的描述，在利用第一数据集合对第一神经网络执行训练操作的过程中，每个第一数据均有一个与该一个第一神经网络之间的适配值。

（二）、通过计算第一神经网络与第三神经网络之间的相似度得到适配值

本实施例中，第一数据集合与一个第一神经网络之间的适配值与第一相似度对应。其中，第一相似度越大，第一数据集合与一个第一神经网络之间的适配值越大；第一相似度越小，第一数据集合与一个第一神经网络之间的适配值越小。第一相似度指的是一个第一

神经网络和第三神经网络之间的相似度。第三神经网络为上一轮迭代中输出预测结果的准确率最高的神经网络；或者，若此轮迭代不是第一轮迭代过程，第三神经网络还可以为与第一神经网络网络结构相同的神经网络，也即第一神经网络与第三神经网络对应于相同的标识信息。第三神经网络与第一神经网络的区别在于，第三神经网络是第一客户端上一次利用第一数据集合对第三神经网络执行训练操作得到的训练后的神经网络。

本申请实施例中，由于第三神经网络为上一轮迭代中输出预测结果的准确率最高的神经网络，且第三神经网络为已经利用第一数据集合训练过的神经网络，也即第三神经网络与第一数据集合的适配程度较高，若一个第一神经网络与该第三神经网络的相似度高，则证明该一个第一神经网络与第一数据集合之间的适配程度高，则适配值就会大；提供了适配值计算的另一种实现方案，提高了本方案的实现灵活性。

具体的，一个第一神经网络和第三神经网络之间的相似度通过以下任一种方式确定：

在一种实现方式中，第一客户端将相同数据分别输入至一个第一神经网络和第三神经网络，并对比一个第一神经网络的输出数据与第三神经网络的输出数据之间的相似度。其中，该相似度可以通过计算两者之间的欧式距离、马氏距离、余弦距离、交叉熵或其他方式获得。

进一步地，一个第一神经网络的输出数据与第三神经网络的输出数据之间的相似度指的可以为整个第一神经网络的输出数据与整个第三神经网络的输出数据之间的第一相似度，则可以将该第一相似度，直接确定为第一神经网络与第三神经网络之间的相似度；或者，将该第一相似度进行转换后，得到第一神经网络与第三神经网络之间的相似度。

一个第一神经网络的输出数据与第三神经网络的输出数据之间的相似度指的还可以为第一神经网络中各个模块的输出数据与第三神经网络中各个模块的输出数据之间的相似度，计算各个模块的输出数据之间相似度的乘积，得到整个第一神经网络的输出数据与整个第三神经网络的输出数据之间的相似度，进而可以得到第一神经网络与第三神经网络之间的相似度。

在另一种实现方式中，若构建该一个第一神经网络与第三神经网络的神经网络模块为相同的神经网络，则第一客户端还可以通过计算一个第一神经网络的权重参数矩阵和第三神经网络的权重参数矩阵之间的第二相似度，进而可以将第二相似度确定为该一个第一神经网络与第三神经网络之间的相似度；或者，将该第二相似度进行转换后，得到该一个第一神经网络与第三神经网络之间的相似度。其中，该第二相似度可以通过计算两者之间的欧式距离、马氏距离、余弦距离、交叉熵或其他方式获得。

本申请实施例，提供了一个第一神经网络和第三神经网络之间的相似度的两种计算方式，提高了本方案的实现灵活性。

需要说明的是，若第三神经网络与第一神经网络对应相同的标识信息，则还可以给第三神经网络增加不置信度，第三神经网络与第一神经网络的间隔时间越久，不置信度越大。可以将不置信度与计算得到的适配值共同决定最终的适配值，可以采用相加或者相乘的方式。

404、第一客户端利用第一数据集合对第一神经网络执行训练操作，得到训练后的第一

神经网络。

本申请实施例中，第一客户端在得到至少一个第一神经网络之后，会利用第一数据集合对第一神经网络执行训练操作，以得到训练后的第一神经网络。具体的，第一数据集合中包括多个第一训练数据以及每个第一训练数据的正确结果，第一客户端将一个第一训练数据输入一个第一神经网络中，得到该一个第一神经网络输出的第一训练数据的预测结果。进而根据第一训练数据的预测结果和第一训练数据的正确结果，生成第四损失函数的函数值，根据第四损失函数的函数值进行梯度求导，以反向更新该一个第一神经网络的权重参数，以完成对该一个第一神经网络的一次训练操作，第一客户端对该一个第一神经网络进行迭代训练，直至满足预设条件，得到训练后的一个第一神经网络。

其中，第四损失函数指示第一训练数据的预测结果和第一训练数据的正确结果，第四损失函数的类型与第一神经网络的任务类型相关，作为示例，例如第一神经网络的任务为分类，则第四损失函数可以为交叉熵损失函数、0-1 损失函数或其他损失函数等，此处不做限定。第一客户端对该一个第一神经网络进行迭代训练的目标为拉近第一训练数据的预测结果和第一训练数据的正确结果之间的相似度；预设条件可以为满足第四损失函数的收敛条件，也可以为迭代次数达到预设次数。

为更直观地理解本方案，如下公开了第四损失函数的一个示例：

$$\text{LossM}_1 = \frac{\sum_{j=1}^{J_1} (M_k(x_{ij}) - y_{ij})^2}{J_1}; \quad (1)$$

其中， LossM_1 代表第四损失函数，第一客户端中的第一数据集合 $d_{ij} = \{x_{ij}, y_{ij}\}$ ， j 的取值为 1 到 J_1 ， M_k 代表一个第二损失函数，应理解，式 (1) 中的示例仅为方便理解本方案，不用于限定本方案。

进一步地，第一客户端在得到至少一个第一神经网络之后，利用第一数据集合对第一神经网络执行训练操作之前，还需要初始化第一神经网络的参数。在一种方式中，第一客户端可以直接使用服务器发送给第一客户端时第一神经网络的参数；在另一种方式中，还可以为利用第一客户端上一次训练第一神经网络时得到的权重参数初始化本次第一神经网络的权重参数；在另一种实现方式中，还可以根据服务器发送给第一客户端时第一神经网络的参数，和，第一客户端上一次训练第一神经网络时得到的权重参数进行加权平均，以初始化本次第一神经网络的权重参数；在另一种实现方式中，还可以随机初始化第二神经网络外网络的参数等，此处不做限定。

可选地，步骤 404 可以包括：第一客户端根据第二损失函数，利用第一数据集合对第一神经网络执行训练操作。其中，第二损失函数包括第一项和第二项，第一项指示第一预测结果与第一训练数据的正确结果之间的相似度，第二项指示第一预测结果与第二预测结果之间的相似度，第二项可以称为惩罚项或约束项。进一步地，第一预测结果为将第一训练数据输入第一神经网络后，由第一神经网络输出的第一训练数据的预测结果；第二预测结果为将第一训练数据输入第四神经网络后，由第四神经网络输出的第一训练数据的预测结果。第四神经网络为未执行过训练操作的第一神经网络，也即第四损失函数和第二损失函数的初始状态一致，但在对第二损失函数进行训练的过程中，第四损失函数的权重参数

一直不会更新。本申请实施例中，由于第一客户端上的第一数据集合不一定与第一神经网络匹配，在利用第一数据集合对第一神经网络进行训练的过程中，第二损失函数还会指示第一预测结果与第二预测结果之间的相似度，也即避免第一神经网络在训练过程中被过多改动。

- 5 也即第二损失函数在第四损失函数的基础上加入了惩罚项，加入惩罚项后的目的为拉近第一神经网络输出的第一训练数据的预测结果与第四神经网络输出的第一训练数据的预测结果之间的相似度。为更直观地理解本方案，如下公开了第二损失函数的一个示例：

$$\text{LossM}_2 = \frac{\sum_{j=1}^{J_i} (M_k(x_{ij}) - y_{ij})^2}{J_i} + \gamma_1 \frac{\sum_{j=1}^{J_i} (M_k(x_{ij}) - y'_{ij})^2}{J_i}; \quad (2)$$

其中， LossM_2 代表第二损失函数， γ_1 是一个超参数， y'_{ij} 代表将第一训练数据输入第四

- 10 损失函数之后，由第四损失函数输出的第一训练数据的预测结果， $\frac{\sum_{j=1}^{J_i} (M_k(x_{ij}) - y_{ij})^2}{J_i}$ 代表的含义以及式(2)中其他字母的含义均可参阅上述对式(1)的描述，此处不再赘述。应理解，式(2)中的举例仅为方便理解本方案，不用于限定本方案。

- 可选地，步骤404还可以包括：第一客户端根据第五损失函数，利用第一数据集合对第一神经网络执行训练操作。其中，第五损失函数指示第一预测结果与第一训练数据的正
15 确结果之间的相似度，还指示第一神经网络与第四神经网络之间的相似度，也即第二损失函数在第四损失函数的基础上加入了惩罚项，加入惩罚项后的目的为拉近第一神经网络与第四神经网络之间的相似度。为更直观地理解本方案，如下公开了第五损失函数的一个示例：

$$\text{LossM}_3 = \frac{\sum_{j=1}^{J_i} (M_k(x_{ij}) - y_{ij})^2}{J_i} + \gamma_2 ||M_k - M_0||; \quad (3)$$

- 20 其中， LossM_3 代表第五损失函数， γ_2 是一个超参数， M_0 代表第四损失函数， $\frac{\sum_{j=1}^{J_i} (M_k(x_{ij}) - y_{ij})^2}{J_i}$ 代表的含义以及式(3)中其他字母的含义均可参阅上述对式(1)的描述，此处不再赘述。应理解，式(3)中的举例仅为方便理解本方案，不用于限定本方案。

- 可选地，第一训练数据与第一神经网络之间的适配值越高，在利用第一训练数据对第一神经网络进行一次训练的过程中，对第一神经网络的权重参数的修改程度越大。进一步
25 地，在一次训练过程中，调整对第一神经网络的权重参数的修改成的方式包括：调整学习率、调整惩罚项的系数或其他方式等。本申请实施例中，由于同一客户端中不同的训练数据与一个第一神经网络之间的适配程度有高有低，所有训练数据均以固定的能力修改一个第一神经网络的权重参数这是不合理的，一个第一训练数据与第一神经网络之间的适配值越高，则证明该第一神经网络越应该处理该一个第一训练数据，在利用该第一训练数据对
30 第一神经网络进行一次训练的过程中，对第一神经网络的权重参数的修改程度越大，有利于提高一个第一神经网络的训练效率。

其中，学习率越大，在一次训练过程中对第一神经网络的权重参数的修改程度越大，学习率越小，再一次训练过程中对第一神经网络的修改程度越小；也即第一训练数据与第一神经网络之间的适配值越高，在利用该第一训练数据对第一神经网络进行一次训练的过

-24-

程中的学习率越大。为更直观地理解本方案，结合上述式（1）至式（3）进行举例。

$$M_{k+1} = M_k - \eta_i * \nabla \text{Loss}M;$$

$$\eta_i = \eta * E;$$

其中， M_{k+1} 代表对 M_k 执行一次训练操作后的第一神经网络， η_i 代表学习率， η 为一个超参数， E 代表第一训练数据与正在训练的该一个第一神经网络之间的适应值， $\text{Loss}M$ 代表 $\text{Loss}M_1$ 、 $\text{Loss}M_2$ 和 $\text{Loss}M_3$ 中的任意一个，应理解，上述举例仅为方便理解本方案，不用于限定本方案。

惩罚项的系数越小，在一次训练过程中对第一神经网络的修改程度越大；惩罚项的系数越大，在一次训练过程中对第一神经网络的修改程度越小；也即第一训练数据与第一神经网络之间的适配值越高，在利用该第一训练数据对第一神经网络进行一次训练的过程中的惩罚项系数越小。结合上述式（2）和上述式（3）进行举例， γ_1 和 γ_2 的取值均可以为 $1/E$ ，也即 γ_1 和 γ_2 的取值均可以为第一训练数据与正在训练的一个第一神经网络之间适配值的倒数。

进一步地，在一种实现方式中，第一数据集合中不同的第一训练数据与该一个第一神经网络之间的适配值可以不同。可以为对第一数据集合进行聚类后得到至少两个数据子集合，同一数据子集合中的训练数据与该一个第一神经网络之间的适配值相同，不同数据子集合中的训练数据与该一个第一神经网络之间的适配值不同。也可以为第一数据集合中每个第一训练数据与该一个第一神经网络之间的适配值均不同。在另一种实现方式中，也可以为将整个第一数据集合视为一个整体，第一数据集合中所有第一训练数据与该一个第一神经网络之间的适配值均相同。

需要说明的是，由于第一客户端会选取出一个或多个第一神经网络，第一客户端需要计算第一训练数据与一个或多个第一神经网络中每个第一神经网络之间的适配值，并对一个或多个第一神经网络中每个第一神经网络执行训练操作。则可以为在每次执行步骤 403 和 404 时的执行对象仅为一个或多个第一神经网络中的一个第一神经网络，则第一客户端需要重复执行步骤 403 和 404 多次。或者，第一客户端可以先通过步骤 403 分别计算第一训练数据与一个或多个第一神经网络中所有第一神经网络之间的适配值，之后再通过步骤 404 分别对每个第一神经网络执行迭代操作。

此外，在步骤 404 所描述的整个训练过程中，若训练后的至少一个第一神经网络中所有第一神经网络的准确率均没有达到第二阈值，第一客户端可以直接生成新的神经网络模块，并结合接收到的多个神经网络模块构建新的第一神经网络。可选地，在通过步骤 404 训练结束后，可以对比包括新增的神经网络模块的第一神经网络与不包括新增的神经网络模块的第一神经网络的准确率，如果准确率增益没有超过第三阈值则不保留新增的神经网络模块。

405、第一客户端将至少一个训练后的第一神经网络包括的至少一个更新后的神经网络模块发送给服务器。

本申请实施例，第一客户端在得到至少一个训练后的第一神经网络之后，会将至少一个训练后的第一神经网络包括的至少一个更新后的神经网络模块发送给服务器，对应的，服务器会接收到第一客户端发送的至少一个更新后的神经网络模块发送给服务器，由于第

一客户端为多个客户端中的任一个客户端，则服务器会接收到多个客户端中每个客户端发送的至少一个更新后的神经网络模块。

可选地，若第一客户端还将新增的神经网络模块发送给服务器，则服务器还可以接收到新增的神经网络模块。

5 406、服务器更新存储的神经网络模块的权重参数。

本申请实施例中，服务器在接收到多个客户端中每个客户端发送的至少一个更新后的神经网络模块之后，服务器需要根据接收到的多个更新后的神经网络模块，更新存储的神经网络的权重参数，以完成多轮迭代中的一轮迭代。

具体的，在一种实现方式中，由于不同客户端发送的可以存在相同的神经网络模块，
10 则将不同客户端发送的相同的神经网络模块的权重参数进行加权平均，作为服务器中该神经网络模块的权重参数。对于不同客户端中没有重合的神经网络模块，则直接将客户端发送的神经网络模块的参数作为服务器中该神经网络模块的权重参数。其中，相同的神经网络模块指的是具体的神经网络相同，且位于相同的分组中。

可选地，若服务器接收到新增的神经网络模块，则可以将新增的神经网络模块放入到
15 对应的分组中。进一步可选地，为增加各个客户端的隐私性，若多个客户端在同一分组中均加入了新增的神经网络模块，则可以将同一分组中所有新增的神经网络模块加权平均为一个神经网络模块之后，再放入该分组中。

在另一种实现方式中，若服务器中存在训练数据，则还可以使用模型蒸馏的方法，利用多个客户端发送的多个更新后的神经网络模块，来更新服务器存储的神经网络模块的权重参数。也即使用服务器中存储的训练数据重新训练服务器中存储的多个神经网络模块，
20 训练的目的为拉近服务器中存储的神经网络模块的输出数据与客户端发送的更新后的神经网络模块的输出数据之间的相似度。

一个神经网络被分为至少两个子模块，服务器存储的多个神经网络模块被分为与至少两个子模块对应的至少两个组，同一组中不同的神经网络模块的功能相同。则可选地，服务器在更新存储的神经网络模块的权重参数之后，还会计算同一组包括的至少两个神经网络模块中不同的神经网络模块之间的相似度，并将相似度大于预设阈值的两个神经网络模块进行合并。本申请实施例中，将相似度大于预设阈值的两个神经网络模块进行合并，也即将冗余的两个神经网络模块进行合并，不仅降低服务器对多个神经网络模块的管理难度；且避免客户端对相似度大于预设阈值的两个神经网络模块进行重复训练，以减少对客户端
30 计算机资源的浪费。

具体的，针对相似度判断的过程。同一组中不同的神经网络模块包括第二神经网络模块和第一神经网络模块，第二神经网络模块和第一神经网络模块之间的相似度通过以下任一种方式确定：

在一种实现方式中，服务器将相同数据分别输入至第二神经网络模块和第一神经网络模块，并对比第二神经网络模块的输出数据与第一神经网络模块的输出数据之间的相似度。该相似度的计算方式包括但不限于：计算两者之间的欧氏距离、马氏距离、余弦距离或者交叉熵等等，此处不做穷举。
35

在另一种实现方式中,若第二神经网络模块和第一神经网络模块具体表现为相同的神经网络,区别仅在于权重参数不同,则可以计算第二神经网络模块的权重参数矩阵和第一神经网络模块的权重参数矩阵之间的相似度。计算相似度的方式与上一实现方式类似,可参照理解。结合上述图5进行举例,先针对第一层(也即第一组),服务器需要计算 SGL1M1, SGL1M2 和 SGL1M3 任意两两之间的相似度 $D_{mn} = \langle SGL1Mm, SGL1Mn \rangle$,如果 D_{mn} 大于预设阈值,则在SGL1Mm 和 SGL1Mn中随机选取一个;比如 D_{12} 大于预设阈值,则随机选择 SGL1M1,并删除 SGL1M2。依次类推处理第二层(也即第二组)、第三层(也即第三组)和第四层(也即第四组)中的神经网络模块。

针对两个神经网络模块的合并过程,服务器可以在两个不同的神经网络网络中随机选取一个神经网络模块;若第二神经网络模块和第一神经网络模块具体表现为相同的神经网络,区别仅在于权重参数不同,则还可以将第二神经网络模块和第一神经网络模块的权重参数求平均,以生成合并后的神经网络模块的权重参数。

本申请实施例中,提供了计算两个不同的神经网络模块之间相似度的两种具体实现方式,则用户可以结合实际情况灵活选择,提高了本方案的实现灵活性。

需要说明的是,服务器在更新存储的神经网络模块的权重参数之后,会重新进入步骤401,以重新执行步骤401至407,也即重新执行下一轮次的迭代。

为了更直观地理解本方案,请参阅图9,图9为本申请实施例提供的机器学习模型的训练方法的一种流程示意图。如图9所示,服务器中存储有多个神经网络模块,服务器会将存储的多个神经网络模块分别发给每个客户端,客户端在得到多个神经网络模块之后,针对多个客户端中的任一客户端(例如第一客户端),先根据多个神经网络模块构建多个第二神经网络,并从多个第二神经网络中选取与第一客户端存储的第一数据集合的数据特性适配的至少一个第二客户端,计算第一数据集合与每个第一神经网络之间的适配值,并利用第一数据集合对每个第一神经网络执行训练操作,得到用于构建第一神经网络的多个更新后的神经网络模块,再将多个更新后的神经网络模块发送给服务器。服务器接收到多个客户端中每个客户端发送的更新后的神经网络模块之后,根据所有客户端发送的更新后的神经网络模块更新存储的多个神经网络模块的权重参数,从而完成多轮迭代中的一轮迭代。虽然图9中仅示出了两个客户端,但实际情况中,服务器可以与更多客户端建立通信连接,图9中的示例仅为方便理解本方案,不用于限定本方案。

(2)由客户端利用选择器选择与客户端存储的数据集合的数据特性适配的第一神经网络

具体的,请参阅图10,图10为本申请实施例提供的机器学习模型的训练方法的一种流程示意图,方法可以包括:

1001、第一客户端接收服务器发送的选择器。

本申请的一些实施例中,服务器向第一客户端发送选择器,对应的,第一客户端接收服务器发送的选择器,选择器为用于从多个神经网络模块中选取与第一数据集合的数据特征匹配的至少一个神经网络模块的神经网络。服务器还可以向第一客户端发送服务器存储的多个神经网络模块中每个神经网络模块的标识信息。

1002、第一客户端根据第一数据集合，将训练数据输入选择器，得到选择器输出的指示信息，指示信息包括多个神经网络模块中每个神经网络模块被选中的概率，用于指示构建至少一个第一神经网络的神经网络模块。

本申请的一些实施例中，第一客户端根据第一数据集合，将训练数据输入选择器，得到选择器输出的指示信息；其中，指示信息包括多个神经网络模块中每个神经网络模块被选中的概率，若多个神经网络模块中共包括 Z 个神经网络模块，则指示信息具体可以表现为包括 Z 个元素的向量，每个元素指代一个神经网络模块被选中的概率。结合图 5 进行举例，图 5 中共包括 18 个神经网络模块，则 Z 的取值为 18，应理解，此处举例仅为方便理解本方案，不用于限定本方案。

针对将训练数据输入选择器的过程。在一种实现方式中，第一客户端可以将第一数据集合中的每个第一训练数据（训练数据的一种示例）均分别输入选择器一次，以得到与每个第一训练数据对应的指示信息。在另一种实现方式中，第一客户端也可以对第一数据集合执行聚类操作，并分别将聚类后的几个类中心（训练数据的一种示例）输入选择器，以得到与每个类中心对应的指示信息。在另一种实现方式中，第一客户端也可以对第一数据集合执行聚类操作，并分别从聚类后的几个数据子集合中抽样几个第一训练数据，将抽样得到的第一训练数据（训练数据的一种示例）分别输入选择器，以得到与每个抽样得到的第一训练数据对应的指示信息等，第一客户端还可以通过其他方式生成指示信息，此处不做穷举。

针对根据指示信息确定构建至少一个第一神经网络的神经网络模块的过程。在一种实现方式中，第一客户端可以初始化一个用于指示每个神经网络模块被选中次数的数组，初始化值为 0，该数组也可以为表格、矩阵等其他形式，此次不做穷举。第一客户端在得到至少一个指示信息之后，针对每一个指示信息，对于选中概率大于第五阈值的神经网络模块，则数组中与该神经网络模块对应的次数加一，在遍历所有指示信息之后，第一客户端根据该数组统计被选中次数大于第六阈值的至少一个神经网络模块，并将前述至少一个神经网络模块确定为用于构建至少一个第一神经网络的神经网络模块，第五阈值和第六阈值的取值均可结合实际情况设定，此处不做限定。

在另一种实现方式中，第一客户端在得到多个指示信息之后，还可以对多个指示信息求平均值，得到一个包括 Z 个元素的向量，向量中每个元素指示一个神经网络模块被选中的概率，进而从 Z 个元素中获取平均值最大的 H 个元素，并将前述 H 个元素指向的 H 个神经网络模块确定为用于构建至少一个第一神经网络的神经网络模块， Z 为大于 1 的整数， H 为大于或等于 1 的整数，对于 Z 和 H 的取值均可以结合实际情况灵活设定，此处不做限定。

1003、第一客户端向服务器发送第一标识信息，第一标识信息为构建第一神经网络的神经网络模块的标识信息。

本申请的一些实施例中，第一客户端上还可以存储有每个神经网络模块的标识信息，在第一客户端确定用于构建第一神经网络的多个神经网络模块之后，还会获取前述多个神经网络模块的标识信息，以组成第一标识信息，第一标识信息中包括构建第一神经网络的

所有神经网络模块的标识信息。

1004、服务器向第一客户端发送第一标识信息指向的构建第一神经网络的神经网络模块。

本申请的一些实施例中，服务器在接收到第一标识信息之后，从存储的所有神经网络模块（也即 L 个神经网络模块）中获取第一标识信息指向的所有神经网络模块，并向第一客户端发送第一标识信息指向的构建第一神经网络的神经网络模块。

1005、第一客户端将第一训练数据输入选择器，得到选择器输出的指示信息。

本申请的一些实施例中，第一客户端将一个第一训练数据输入训练器，得到选择器输出的一个指示信息，该一个指示信息具体可以表现为包括 Z 个元素的向量，指示 Z 个神经网络模块中每个神经网络模块被选中的概率，用于指示构建第一神经网络的神经网络模块。结合图 5 举例，该指示信息可以 $[M_{SGL1M1}, M_{SGL1M2}, M_{SGL1M3}, M_{SGL2M1}, \dots, M_{SGL4M3}, M_{SGL4M4}]$ ，其中， M_{SGL1M1} 代表第一组中第一个神经网络模块被选中的概率，剩下的元素依次类推，此处不再赘述。

本申请实施例中，根据第一数据集合，将训练数据输入至选择器，得到选择器输出的指示信息，并根据该指示信息选取用于构建第一神经网络的神经网络模块，选择器为用于从多个神经网络模块中选取与第一数据集合的数据特征匹配的神经网络模块的神经网络，提供了选取构建第一神经网络的神经网络模块的又一种实现方式，提高了本方案的实现灵活性；且通过神经网络来选取，有利于提高神经网络模块的选取过程的准确率。

1006、第一客户端根据接收到的多个神经网络模块、指示信息和第一训练数据，得到第一神经网络输出的第一训练数据的预测结果。

本申请的一些实施例中，第一客户端在通过步骤 1005 得到一个指示信息之后，可以根据接收到的多个神经网络模块、指示信息和第一训练数据，得到第一神经网络输出的第一训练数据的预测结果。为更直观地理解本方案，结合图 5 进行举例，如下公开了计算第一训练数据的预测结果的公式的一个示例：

$$h1 = \sum_{q=1}^3 M_{SGL1Mq} * SGL1Mq(x);$$

$$h2 = \sum_{q=1}^4 M_{SGL2Mq} * SGL2Mq(h1);$$

$$h3 = \sum_{q=1}^2 M_{SGL3Mq} * SGL3Mq(h2);$$

$$y = \sum_{q=1}^4 M_{SGL4Mq} * SGL4Mq(h3);$$

其中， M_{SGL1Mq} 、 M_{SGL2Mq} 、 M_{SGL3Mq} 和 M_{SGL4Mq} 均来自选择器输出的指示信息， $SGL1Mq$ 代表第一组神经网络模块中的一个神经网络模块， $SGL1Mq(x)$ 代表将第一训练数据输入第一组神经网络模块中的一个神经网络模块之后该神经网络模块的输出，若第一客户端未从服务器获取到某个神经网络模块，则视为该神经网络模块的输出为 0， $h1$ 代表整个第一组的输出数据，可对上述公式中的其他公式进行类推理解， y 代表整个第一神经网络的输出，也即第一训练数据的预测结果，应理解，此处举例仅为方便理解本方案，不用于限定本方案。

1007、第一客户端根据第三损失函数，对第一神经网络和选择器执行训练操作，其中，第三损失函数指示第一训练数据的预测结果与正确结果之间的相似度，还指示该指示信息的离散程度。

本申请的一些实施例中，第一客户端在生成第一训练数据的预测结果之后，会根据第一训练数据的预测结果与第一训练数据的正确结果以及选择器生成的指示信息，生成第三损失函数的函数值，并根据第三损失函数的函数值进行梯度求导，以反向更新第一神经网络（也即更新接收到的多个神经网络模块）和选择器的权重参数，以完成对接收到的多个神经网络模块和选择器的一次训练。训练的目的为拉近第一训练数据的预测结果与正确结果之间的相似度，且增大选择器输出的指示信息的离散程度。

其中，第三损失函数包括第三项和第四项，第三项指示第一训练数据的预测结果与正确结果之间的相似度，第四项指示该指示信息的离散程度。第三项可以为基于第一训练数据的预测结果与正确结果之间的交叉熵距离、一阶距离、二阶距离等得到，第四项可以为对该指示信息进行正则化处理，例如对该指示信息进行 L1 正则化、进行 LP 正则化等等，此处均不做限定。为更直观地理解本方案，以下公开第三损失函数的一个示例：

$$\text{LossM}_4 = \frac{\sum_{j=1}^{J_i} (M_k(x_{ij}) - y_{ij})^2}{J_i} + \gamma_3 |\text{MS}(x)|; \quad (4)$$

其中， LossM_4 代表第三损失函数， $\frac{\sum_{j=1}^{J_i} (M_k(x_{ij}) - y_{ij})^2}{J_i}$ 的含义可参阅图 4 对应实施例中对应式 (1) 的描述，此处不做赘述， $\text{MS}(x)$ 代表选择器输出的指示信息， γ_3 为一个超参数，应理解，式 (4) 中的举例仅为方便理解本方案，不用于限定本方案。

第一客户端重复执行步骤 1005 至 1007，直至达到预设条件，得到第一标识信息指向的多个更新后的神经网络模块和训练后的训练器。预设条件可以为迭代训练的迭代次数达到预设次数，也可以为第三损失函数满足收敛条件。

1008、第一客户端将至少一个更新后的神经网络模块和训练后的训练器发送给服务器。

1009、服务器更新存储的神经网络模块的权重参数。

本申请实施例中，服务器在可以接收到多个客户端（包括第一客户端）发送的多个更新后的神经网络模块，需要更新存储的 Z 个神经网络模块的权重参数。具体实现方式可参阅图 3 对应实施例中步骤 406 中的描述，此处不做赘述。

1010、服务器更新选择器的权重参数。

本申请的一些实施例中，服务器可以接收到多个客户端发送的训练后的训练器，对多个训练后的训练器中对应位置的权重参数进行求平均值，以更新服务器存储的选择器的权重参数，从而完成了多轮迭代中的一轮迭代。需要说明的是，在执行完步骤 1010 之后，可以重新进入步骤 1001，以进入下一轮的迭代。

本申请实施例中，在训练构建第一神经网络的神经网络模块的同时，训练选择器，节约了计算机资源；用选择器处理需要处理的数据来训练选择器，有利于提高选择器输出的指示信息的准确率。

本申请实施例中，通过上述方式，能够对不同数据特性的训练数据分配不同的神经网络，也即实现了神经网络与数据特性之间的个性化匹配；此外，由于第一客户端为多个客

户端中的任一客户端，对多个客户端中的每个客户端均根据客户端存储的训练数据集合的数据特性分配并训练神经网络，能够利用相同数据特性的训练数据训练相同的神经网络，不同数据特性的训练数据训练不同的神经网络，从而不仅实现了神经网络与数据特性之间的个性化匹配，而且有利于提高训练后神经网络的准确率。

(3) 由服务器根据第二适配关系选择与第一客户端存储的数据集合的数据特性适配的第一神经网络

具体的，请参阅图 11，图 11 为本申请实施例提供的机器学习模型的训练方法的一种流程示意图，方法可以包括：

1101、服务器获取与第一客户端对应的至少一个第一神经网络。

本申请的一些实施例中，服务器中可以配置有多个神经网络模块，服务器会根据存储的多个神经网络模块构建多个第二神经网络，对于多个神经网络模块以及多个第二神经网络的描述可以参阅图 4 对应实施例中步骤 401 中的描述，此处不做赘述。

当服务器选择要为第一客户端分配至少一个第一神经网络时，需要获取与第一客户端对应的至少一个第一神经网络。具体的，与图 4 对应实施例中步骤 402 中描述的类似，服务器可以从多个第二神经网络中随机选取至少一个第一神经网络模块，也可以根据第二适配关系，从至少两个第二神经网络中选取与第一数据集合的适配值高的至少一个第一神经网络。前述两种方式的应用情况可参阅图 4 对应实施例中步骤 402 中的描述，此处不做赘述。

与图 4 对应的实施例中步骤 402 不同的地方在于，服务器中存储的为第二适配关系，第二适配关系中包括多个适配值，适配值用于表示多个客户端中存储的训练数据与第二神经网络之间的适配程度。结合图 5 进行举例。

表 2

	神经网络 ID1	神经网络 ID2		神经网络 ID96
客户端 ID1	E1_1	E1_2		E1_96
客户端 ID2	E2_1	Null		Null
.....				
客户端 ID100	Null	E100_2		E100_96

其中，表 2 中以共可以构建 96 个第二神经网络，共有 100 个客户端为例，E1_1、E1_2 等代表适配值，如表 2 所示，第二适配关系中包括空值，应理解，表 2 中的示例仅为方便理解本方案，不用于限定本方案。

在服务器为非初次执行第一神经网络的分配操作，或者，在第二适配关系包括的适配值的数量所占比例大于第一阈值的情况下，服务器会根据第二适配关系，从至少两个第二神经网络中选取与第一数据集合（也即第一客户端）的适配值高的至少一个第一神经网络。

具体的，服务器可以得到与第二适配关系对应的第二适配矩阵，对第二适配矩阵进行矩阵分解，以得到分解后的神经网络的相似性矩阵和用户的相似性矩阵，神经网络的相似性矩阵和用户的相似性矩阵的乘积与第二适配关系中对对应位置的值需要相似。进而将神经

- 31 -

网络的相似性矩阵和用户的相似性矩阵相乘，得到第二补全矩阵，进而根据第二补全矩阵选择与第一数据集合（也即第一客户端）的适配值高的至少一个第一神经网络。

可选地，第一客户端选取的至少一个第一神经网络不仅可以包括适配值高的至少一个第一神经网络，还包括随机选取的至少一个第一神经网络。

1102、服务器将选取的至少一个第一神经网络发送给第一客户端。

1103、第一客户端计算第一数据集合与第一神经网络之间的适配值。

1104、第一客户端利用第一数据集合对第一神经网络执行训练操作，得到训练后的第一神经网络。

1105、第一客户端将至少一个训练后的第一神经网络包括的至少一个更新后的神经网络模块发送给服务器。

本申请实施例中，步骤 1103 至 1105 的具体实现方式可参阅图 4 对应实施例中步骤 403 至 405 中的描述，此处不做赘述。

1106、第一客户端将第一数据集合与每个第一神经网络之间的适配值发送给服务器。

本申请的一些实施例中，第一客户端还会将通过步骤 1103 计算得到的每个第一神经网络与第一数据集合（也即第一客户端）之间的适配值发送给服务器。其中包括神经网络的标识信息和第一客户端的标识信息，用于指示服务器第一客户端与哪几个神经网络之间的适配值。应理解，步骤 1106 可以与步骤 1105 一起执行，也可以在步骤 1104 和 1105 任一步骤的之前或之后执行，此处不限定步骤 1106 的执行顺序。

1107、服务器更新第二适配关系。

本申请的一些实施例中，由于第一客户端为多个客户端中的任一个客户端，则服务器可以获取到每个客户端发送的适配值，也即服务器客户得到多组适配关系，每组适配关系为一个客户端标识和一个神经网络标识之间的适配值。则服务器可以根据接收到的多个适配值更新第二适配关系。服务器还可以从第二适配关系中删除长时间不更新适配值，长时间不更新指的是超过 20 轮未更新过。

1108、服务器更新存储的神经网络模块的权重参数。

本申请实施例中，步骤 1108 的具体实现方式可参阅图 4 对应实施例中步骤 405 中的描述，此处不做赘述。可选地，服务器还可以删除跟所有客户端适配度都很低的神经网络模块，或者，还可以删除长时间不被选中的神经网络。

在执行完步骤 1108 之后，还可以更新第二适配关系，以将与删除掉的神经网络模块所对应的信息从第二适配关系中删除，需要说明的是，可以先执行步骤 1107，再执行步骤 1108，也可以先执行步骤 1108，再执行步骤 1107，此处不做限定。

本申请实施例中，在服务器侧配置第二适配关系，由客户端生成适配值并发送给客户端，由服务器根据第二适配关系，选取与第一客户端适配的第一神经网络，既避免了对客户端计算机资源的占用，也避免了客户端的数据的泄露。

(4) 由服务器利用选择器选择与第一客户端存储的数据集合的数据特性适配的第一神经网络

具体的，请参阅图 12，图 12 为本申请实施例提供的机器学习模型的训练方法的一种

流程示意图，方法可以包括：

1201、第一客户端对第一数据集合执行聚类操作后，得到至少一个数据子集合，并生成与至少一个数据子集合一一对应的至少一个第一类中心。

本申请的一些实施例中，第一客户端对第一数据集合执行聚类操作后，得到至少一个数据子集合之后，会生成每个数据子集合的第一类中心，从而生成与该至少一个数据子集合一一对应的至少一个第一类中心。

1202、服务器接收第一客户端发送的至少一个第一类中心。

本申请的一些实施例中，第一客户端在生成至少一个第一类中心之后，会将该至少一个第一类中心发送给服务器，对应的，服务器接收第一客户端发送的至少一个第一类中心。

1203、服务器将至少一个第一类中心分别输入选择器，得到选择器输出的指示信息，并根据指示信息，确定构建至少一个第一神经网络的神经网络模块。

本申请的一些实施例中，服务器将至少一个第一类中心分别输入选择器，得到选择器输出的与至少一个第一类中心对应的至少一个指示信息。进而根据指示一个指示信息选取用于构建至少一个第一神经网络的神经网络模块。前述选取过程可以参阅图 10 对应实施例中步骤 1002 中的描述，此处不做赘述。

1204、服务器将选择器和构建至少一个第一神经网络的神经网络模块发送给第一客户端。

1205、第一客户端将第一训练数据输入选择器，得到选择器输出的指示信息，指示信息包括多个神经网络模块中每个神经网络模块被选中的概率，用于指示构建第一神经网络的神经网络模块。

1206、第一客户端根据接收到的多个神经网络模块、指示信息和第一训练数据，得到第一神经网络输出的第一训练数据的预测结果。

1207、第一客户端根据第三损失函数，对第一神经网络和选择器执行训练操作，其中，第三损失函数指示第一训练数据的预测结果与正确结果之间的相似度，还指示该指示信息的离散程度。

1208、第一客户端将至少一个更新后的神经网络模块和训练后的训练器发送给服务器。

1209、服务器更新存储的神经网络模块的权重参数。

1210、服务器更新选择器的权重参数。

本申请实施例中，步骤 1205 至 1210 的具体实现方式可参阅图 10 对应实施例中步骤 1005 至 1010 的描述，此处不做赘述。需要说明的是，服务器可以重复执行步骤 1201 至 1208，以实现与多个客户端中每个客户端的交互，再执行步骤 1209 和 1210，以完成多轮迭代中的一轮迭代。服务器在执行完步骤 1210 之后，再重新进入步骤 1201，以进入下一轮迭代。

本申请实施例中，通过选择器来执行神经网络模块的选择步骤，有利于提高选择过程的准确率，由服务器来执行选择步骤，有利于释放客户端的存储空间，和避免对客户端计算机资源的占用，且仅将类中心发送给服务器，也尽量避免客户端信息的泄露。

本申请实施例中，通过上述方式，能够对不同数据特性的训练数据分配不同的神经网络，也即实现了神经网络与数据特性之间的个性化匹配；由于第一客户端为多个客户端中

的任一客户端，对多个客户端中的每个客户端均根据客户端存储的训练数据集合的数据特性分配并训练神经网络，能够利用相同数据特性的训练数据训练相同的神经网络，不同数据特性的训练数据训练不同的神经网络，从而不仅实现了神经网络与数据特性之间的个性化匹配，而且有利于提高训练后神经网络的准确率；由服务器选择与各个客户端适配的神经网络，既避免了将所有神经网络模块发送给客户端，以减少对客户端存储资源的浪费；且避免了对客户端计算机资源的占用，有利于提高用户体验。

二、推理阶段

具体的，请参阅图 13，图 13 为本申请实施例提供的数据处理方法的一种流程示意图，方法可以包括：

1301、服务器获取与第二客户端存储的第二数据集合的数据特性对应的至少一个第三神经网络。

本申请实施例中，第二客户端可以为与服务器连接的多个客户端中的任一客户端，也可以为与服务器新建立连接关系的客户端。

具体的，在一种情况下，由第二客户端选择与第二数据集合的数据特性对应的至少一个第三神经网络，则服务器可以接收第二客户端发送的第二标识信息，第二标识信息为第三神经网络的标识信息，或者，第二标识信息为构建第三神经网络的神经网络模块的标识信息；对应的，服务器获取第二标识信息指向的一个或多个第三神经网络，或者，获取第二标识信息指向的用于构建一个或多个第一神经网络的神经网络模块。

在另一种情况下，由服务器选择与第二数据集合的数据特性对应的至少一个第三神经网络，则在一种实现方式中，服务器在得到第二客户端的标识信息之后，根据第二适配关系，获取与第二客户端的标识信息适配的至少一个第三神经网络。

在另一种情况下，第二客户端对第二数据集合执行聚类操作，得到至少一个第二数据子集合，生成与至少一个第二数据子集合对应的至少一个第二类中心，服务器接收到至少一个第二类中心后，将至少一个第二类中心分别输入选择器中，以得到用于构建至少一个第三神经网络的神经网络模块。

在另一种情况下，服务器根据第二客户端的标识信息和第二适配关系，从至少两个第二神经网络中选取至少一个第三神经网络，至少一个第三神经网络中包括与第二数据集合适配高的神经网络。

在另一种情况下，服务器从多个第二神经网络中随机选取至少一个第三神经网络。

1302、服务器向第二客户端发送至少一个第三神经网络。

本申请实施例中，服务器向第二客户端发送至少一个第三神经网络，或者，用于构建至少一个第三神经网络的神经网络模块。

1303、第二客户端通过至少一个第三神经网络，生成待处理数据的预测结果。

本申请实施例中，第二客户端可以从至少一个第三神经网络中随机选取一个第三神经网络，也可以根据第二数据集合，从至少一个第三神经网络中选取与第二数据集合适配度最高的等，进而通过选取出一个第三神经网络，生成待处理数据的预测结果。

本申请实施例中，不仅在训练阶段可以结合各个客户端中存储的数据集合的数据特性

来对神经网络执行训练操作，也即不仅在训练阶段可以实现神经网络的个性化定制，在推理阶段，也可以实现神经网络的个性化分配，保持了训练阶段和推理阶段的连贯性，且有利于提高推理阶段的准确率。

在上述各个实施例的基础上，由于需要利用客户端上的数据来训练神经网络，为了提高用户数据的安全度。本申请实施例还提供了在执行训练操作之前，对客户端上的数据进行加密的方法，请参阅如下描述。

实施例 1：基于梯度的模块并进行打包处理的预训练方案

客户端 A 拥有特征集合 $F_A = \{f_1, f_2, \dots, f_N\}$ ，客户端 B 拥有特征集合 $F_B = \{f_{N+1}, f_{N+2}, \dots, f_{N+M}\}$ ，客户 A 的数据为 $D_A = \{d_{1A}, d_{2A}, d_{3A}, \dots, d_{pA}\}$ ，客户端 B 拥有数据 $D_B = \{d_{1B}, d_{2B}, d_{3B}, \dots, d_{pB}\}$ 。 d_{pA} 的数据的特征为 F_A ， d_{pB} 的数据的特征为 F_B ， $d_p = [d_{pA}, d_{pB}]$ 表示第 p 条数据的全部特征值，客户端 B 用户数据标签 $L = \{l_1, l_2, l_3, \dots, l_p\}$ ，客户端 A 拥有的模型参数为 W_A ，客户端 B 拥有的模型参数为 W_B ，客户端 A 对应的模型梯度为 G_A ，客户端 B 对应的模型梯度为 G_B 。

训练过程：

第 1 步：客户端 A 生成半全同态加密的公钥 pk_A 和私钥 sk_A 。

第 2 步：客户端 B 生成全同态加密的公钥 pk_B 和私钥 sk_B 。

第 3 步：客户端 A 将其公钥 pk_A 发送给客户端 B，客户端 B 将其公钥 pk_B 发送给客户端 A。

第 4 步：客户端 A 利用其拥有的模型参数 W_A 以及其拥有的数据 D_A 计算 U_A 。客户端 A 对 U_A 进行打包操作得到 DU_A 。户端 A 利用客户端 A 的公钥 pk_A 对 DU_A 进行同态加密，得到加密后的 $[[DU_A]]_{pk_A}$ ，并将其发送给客户端 B。

打包指的是将数据 $U_A = [u_{A1}, u_{A2}, u_{A3}, \dots, u_{AP}]$ 按指定包长 L 分割成小的数据包

$DU_A = [Du_{A1}, Du_{A2}, \dots, Du_{AP/L}]$ ，其中，

$Du_{A1} = [u_{A1}, u_{A2}, \dots, u_{AL}]$ ， $Du_{A2} = [u_{AL+1}, u_{AL+2}, \dots, u_{AL+L}]$ 。

公钥 pk_A 对 DU_A 进行同态加密指的是用公钥 pk_A 分别对 $Du_{A1}, Du_{A2}, \dots, Du_{AP/L}$ 进行加密。

第 5 步：客户端 B 利用其拥有的模型参数 W_B 以及其拥有的数据 D_B 以及标签 L，计算 $U_{B-L} = U_B - L$ 客户端 B 对 U_{B-L} 进行打包，得到 DU_{B-L} 客户端 B 用客户端 B 的公钥 pk_B 对打包后进行加密得到 $[[DU_{B-L}]]_{pk_B}$ ，并将其发送给客户端 A。

第 6 步：客户端 A 将自己 DU_A 的利用客户端 B 的公钥 pk_B 进行加密得到 $[[DU_A]]_{pk_B}$ 将其与从客户端 B 得到的利用客户端 B 公钥加密的 DU_{B-L} 即 $[[DU_{B-L}]]_{pk_B}$ 相加，与编码后的数据集 D_A 相乘得到经同态加密的模型对应的梯度 $[[G_A]]$ 。生成与 W_A 的维度*打包长度相同大小的 W_{A_Noise} 保存。对 W_{A_Noise} 进行打包得到 DW_{A_Noise} 并利用客户端 B 的公钥加密打包后的

— 35 —

DW_{A_Noise} 得到 $[[DW_{A_Noise}]]$ ，将前面得到的同态加密的模型梯度与 $[[DW_{A_Noise}]]$ 相加，得到带噪声且同态加密的模型梯度。将其发送给客户端 B 使用客户端 B 的私钥进行解密后发送回客户端 A，客户端 A 利用解密后的带噪声梯度值，减去保存的噪声 W_{A_Noise} 并沿打包维度进度累加得到真实的模型梯度，更新模型参数 W_A 。

- 5 第 7 步：客户端 B 将自己的 DU_{B-L} 利用客户端 A 的公钥 pk_A 进行加密得到 $[[DU_{B-L}]]_{pk_A}$ ，将其与从客户端 A 得到的利用客户端 A 公钥加密的 DU_A 即 $[[DU_A]]_{pk_A}$ 相加，与编码后的 D_B 相乘得到经同态加密的模型 W_B 对应的梯度。生成与 W_B 维度*打包长度相同大小的 W_{B_Noise} 保存。对 W_{B_Noise} 进行打包得到 DW_{B_Noise} ，并利用客户端 A 的公钥加密 DW_{A_Noise} 得到 $[[DW_{A_Noise}]]$ ，将前面得到的同态加密的模型参数与 $[[DW_{B_Noise}]]$ 相加，得到带噪声且同态加密的模型梯度 $[[G_B]]$ 。将其发送给客户端 A 使用客户端 A 的私钥进行解密后发送回客户端 B，客户端 B 利用解密后的带噪声梯度值，减去保存的噪声 W_{B_Noise} 并沿打包维度进度累加得到真实的模型梯度，更新模型参数 W_B 。

第 8 步：判断是否达到收敛条件，如果达到则结束训练过程，否则继续回到第 4 步继续执行。

- 15 推理过程：

对于拟合和分类问题可以使用：

客户端 A 和 B 分布计算 U_A 和 U_B 并有一方计算 U_A+U_B 的值。

对于分类问题可以使用：

客户端 A 和 B 各自计算 U_A 和 $-U_B$ 。

- 20 客户端 A 和 B 按预制的小数点前固定位数对数据进行补 0 得到： IU_A 和 $-IU_B$

例如 $U_A = 1234.5678$ ， $-U_B = 12.3456$ 我们预设的小数点前为 6 位，则

$IU_A = 001234.5678$ ， $-U_B = 000012.3456$ 。

- 客户端 A 和 B 分别每次从最高位开始按预制的位数取数比较，如果相等则在比较接下来的预制的位数的数，如果能够判断大小则停止比较并根据比较大小判断 U_A 和 $-U_B$ 的大小；
25 如果已经取了指定个数的数进行比较，则停止比较 $U_A=-U_B$ ，判定。如预设为 2 位，则客户端 A 取 IU_A 的 00 和客户端 B 取 $-IU_B$ 的 00 比较，因为相同，所以客户端 A 取 IU_A 的 12 和客户端 B 取 $-IU_B$ 的 00 比较，由于 12 比 00 大，所以 U_A 大于 $-U_B$ 。

取数比较的过程如下：

客户端 A 截取数据 I_a ，客户端 B 截取数据 I_b 。

- 30 客户端 B 生产公钥-私钥对并将公钥发送给客户端 A。

客户端 A 生产一个随机整数 $RIntX$ 并使用客户端 B 发来的公钥加密，得到 $[[RIntX]]$ ，客户端 A 将 $[[RIntX]]-I_a$ 发送给 B。

客户端 B 对收到的 $[[RIntX]]-I_a$ 分别加 0 到 99，然后解密得到

-36-

DRIntX = [DRIntX₀, DRIntX₁, DRIntX₂, ..., DRIntX₉₉], 然后客户端 B 对 DRIntX 第 I_b 位置的数据-1, 对 DRIntX 位置大于 I_b 的-2, 并对 DRIntX 中的每个数按预置的模数取余并将结果发送给 A。

客户端 A 对自己的 RIntX 按照预置的与 B 相同的模数取余, 然后跟收到的 DRIntX 的第 I_a 位置的数据相比较, 如果相等, 则说明 I_a < I_b, 如果取模相差为 1 则 I_a = I_b, 如果取模相差为 2 则 I_a > I_b。

如果 U_A 大于 -U_B 则 U_A + U_B > 0; 如果 U_A 小于 -U_B 则 U_A + U_B < 0; 如果 U_A 等于 -U_B 则 U_A + U_B = 0。

10 实施例 2: 基于分类树的模块预训练方案

模块可以采用预训练的方式实施, 同时可以进行多用户不同特征的联合学习。

客户端 A 拥有特征集合 F_A = {f₁, f₂, ..., f_N}, 客户端 B 拥有特征集合 F_B = {f_{N+1}, f_{N+2}, ..., f_{N+M}}, 客户 A 的数据为 D_A = {d_{1A}, d_{2A}, d_{3A}, ..., d_{pA}}, 客户端 B 拥有数据 D_B = {d_{1B}, d_{2B}, d_{3B}, ..., d_{pB}}。d_{pA} 的数据的特征为 F_A, d_{pB} 的数据的特征为 F_B, d_p =

15 [d_{pA}, d_{pB}] 表示第 p 条数据的全部特征值, 客户端 B 用户数据标签 L = {l₁, l₂, l₃, ..., l_p}, l_p = 0 表示类 0, l_p = 1 表示类 1。

训练过程:

第 1 步: 客户端生成半同态加密 (全同态加密) 的公钥 pk_B 和私钥 sk_B, 使用公钥 pk_B 加密数据标签 L 得到加密的数据 pk_B(L) = {pk_B(l₁), pk_B(l₂), pk_B(l₃), ..., pk_B(l_p)}。

20 第 2 步: 客户端 B 将公钥 pk_B 和加密后的标签 pk_B(L) 发送给客户端 A, 设置节点编号 h=0, 全部数据都属于节点 h。B 输出推理树为空, A 和 B 初始化分割树为空。

第 3 步: 客户端 A 根据本地数据生成特征切割方案集合 S_A = {s_{1A}, s_{2A}, s_{3A}, ..., s_{iA}}, 根据分割策略 s_{iA} 将属于节点 h 的数据分成左右两个子节点 2*h 和 2*h+1。计算子节点 2*h 和 2*h+1 的加密的数据标签的和:

$$25 \quad \text{sum_pk_label}_{s_{iA_2h}} = \left(\sum_{p \text{ 属于节点 } 2h} \text{pk}_B(l_p) \right) + \text{pk}_B(r_{s_{iA_2h}});$$

$$\text{sum_pk_label}_{s_{iA_2h+1}} = \left(\sum_{p \text{ 属于节点 } 2h+1} \text{pk}_B(l_p) \right) + \text{pk}_B(r_{s_{iA_2h+1}});$$

以及两个集合中数据的个数:

$$\text{sum_num}_{s_{iA_2h}} = \sum_{p \text{ 属于节点 } 2h} 1;$$

$$\text{sum_label}_{\text{set}_{s_{iA_2h+1}}} = \sum_{p \text{ 属于节点 } 2h+1} 1;$$

30 其中 r_{s_{iA}_2h} 表示一个随机数, pk_B(r_{s_{iA}_2h}) 表示用 pk_B 对随机数 r_{s_{iA}_2h} 加密。

客户端 B 根据本地数据生成特征切割方案集合 S_B = {s_{1B}, s_{2B}, s_{3B}, ..., s_{iB}}, 根据分割策

— 37 —

略 s_{iB} 将属于节点 h 的数据分成左右两个子节点 $2*h$ 和 $2*h+1$ 。计算子节点 $2*h$ 和 $2*h+1$ 的数据标签的和：

$$\text{sum_label}_{s_{iB_2h}} = \sum_{p \text{ 属于节点 } 2h} l_p;$$

$$\text{sum_label}_{s_{iB_2h+1}} = \sum_{p \text{ 属于节点 } 2h+1} l_p;$$

5 以及两个集合中数据的个数：

$$\text{sum_num}_{s_{iB_2h}} = \sum_{p \text{ 属于节点 } 2h} 1;$$

$$\text{sum_num}_{s_{iB_2h+1}} = \sum_{p \text{ 属于节点 } 2h+1} 1;$$

其中 $r_{s_{iA_2h}}$ 表示一个随机数， $\text{pk}_B(r_{s_{iA_2h}})$ 表示用 pk_B 对随机数 $r_{s_{iA_2h}}$ 加密。

第 4 步：客户端 A 将 $\text{sum_pk_label}_{s_{iA_2h}}$ 和 $\text{sum_pk_label}_{s_{iA_2h+1}}$ 发送给客户端 B。

10 第 5 步：客户端 B 使用 sk_B 解密 $\text{sum_pk_label}_{s_{iA_2h}}$ 和 $\text{sum_pk_label}_{s_{iA_2h+1}}$ 得到 $\text{sum_r_label}_{s_{iA_2h}}$ 和 $\text{sum_r_label}_{s_{iA_2h+1}}$ 并发送给客户端 A，客户端 A 收到 $\text{sum_r_label}_{s_{iA_2h}}$ 和 $\text{sum_r_label}_{s_{iA_2h+1}}$ 减去对应的随机数得到 $\text{sum_label}_{s_{iA_2h}}$ 和 $\text{sum_label}_{s_{iA_2h+1}}$ 。

第 6 步：客户端 A 使用 $\text{sum_label}_{s_{iA_2h}}$ 和 $\text{sum_label}_{s_{iA_2h+1}}$ 和 $\text{sum_num}_{s_{iA_2h}}$ 和 $\text{sum_num}_{s_{iA_2h+1}}$ 计算各种分割下的 gini 系数，并选择最小的 gini 系数对应的切割方案记为 $s_{\min A}$ 以及基尼系数值 $\text{gini}_{\min A}$ ，客户端 B 使用 $\text{sum_label}_{s_{iB_2h}}$ 和 $\text{sum_label}_{s_{iB_2h+1}}$ 和 $\text{sum_num}_{s_{iB_2h}}$ 和 $\text{sum_num}_{s_{iB_2h+1}}$ 计算各种分割下的 gini 系数，并选择最小的 gini 系数对应的切割方案记为 $s_{\min B}$ 以及基尼系数值 $\text{gini}_{\min B}$ 。

第 7 步：客户端 A 发送 $\text{gini}_{\min A}$ 给客户端 B，客户端 B 比较大小，并将比较结果返回给 A。B 推理树第 h 个节点标记为 gini 值小的一方的编号。

20 第 8 步：gini 小的一方根据对应的数据分割方案，把数据进行分割并将分割结果发送给对方。并将分割策略写入分割树的第 h 个节点。

第 9 步： $h = h+1$ 重复第 3 步到 7 步直到指定的重复步数。

第 10 步：B 统计叶子节点哪个类别多则标记该叶子节点为该类别。

推理过程：

25 步骤 1：根据推理树，选择处理方是 A 和 B，选择处理方。

步骤 2：根据分割树的分割策略，选择下一个节点的位置。

重复 1 和 2 直到叶子节点，分类结果为该类标记的类别。

实施例 3：基于回归树的模块预训练方案

— 38 —

模块可以采用预训练的方式实施，同时可以进行多用户不同特征的联合学习。

客户端 A 拥有特征集合 $F_A = \{f_1, f_2, \dots, f_N\}$ ，客户端 B 拥有特征集合 $F_B = \{f_{N+1}, f_{N+2}, \dots, f_{N+M}\}$ ，客户 A 的数据为 $D_A = \{d_{1A}, d_{2A}, d_{3A}, \dots, d_{pA}\}$ ，客户端 B 拥有数据 $D_B = \{d_{1B}, d_{2B}, d_{3B}, \dots, d_{pB}\}$ 。 d_{pA} 的数据的特征为 F_A ， d_{pB} 的数据的特征为 F_B ， $d_p =$

5 $[d_{pA}, d_{pB}]$ 表示第 p 条数据的全部特征值，客户端 B 用户数据标签 $L = \{l_1, l_2, l_3, \dots, l_p\}$ 。

训练过程：

第 1 步：客户端生成半同态加密（全同态加密）的公钥 pk_B 和私钥 sk_B ，使用公钥 pk_B 加密数据标签 L 得到加密的数据 $pk_B(L) = \{pk_B(l_1), pk_B(l_2), pk_B(l_3), \dots, pk_B(l_p)\}$ 和数据标签的平方值 $pk_B(L^2) = \{pk_B(l_1^2), pk_B(l_2^2), pk_B(l_3^2), \dots, pk_B(l_p^2)\}$ 。

10 第 2 步：客户端 B 将公钥 pk_B 和加密后的标签 $pk_B(L)$ 发送给客户端 A，设置节点编号 $h=0$ ，全部数据都属于节点 h 。B 输出推理树为空，A 和 B 初始化分割树为空。

第 3 步：客户端 A 根据本地数据生成特征切割方案集合 $S_A = \{s_{1A}, s_{2A}, s_{3A}, \dots, s_{IA}\}$ ，根据分割策略 s_{iA} 将属于节点 h 的数据分成左右两个子节点 $2*h$ 和 $2*h+1$ 。计算子节点 $2*h$ 和 $2*h+1$ 的加密的数据标签的和：

$$15 \quad \text{sum_pk_label}_{s_{iA_2h}} = \left(\sum_{p \text{ 属于节点 } 2h} pk_B(l_p) \right) + pk_B(r_{s_{iA_2h}});$$

$$\text{sum_pk_label}_{s_{iA_2h+1}} = \left(\sum_{p \text{ 属于节点 } 2h+1} pk_B(l_p) \right) + pk_B(r_{s_{iA_2h+1}});$$

以及两个集合中数据的个数：

$$\text{sum_num}_{s_{iA_2h}} = \sum_{p \text{ 属于节点 } 2h} 1;$$

$$\text{sum_label}_{\text{set}_{s_{iA_2h+1}}} = \sum_{p \text{ 属于节点 } 2h+1} 1;$$

20 其中 $r_{s_{iA_2h}}$ 表示一个随机数， $pk_B(r_{s_{iA_2h}})$ 表示用 pk_B 对随机数 $r_{s_{iA_2h}}$ 加密。

客户端 B 根据本地数据生成特征切割方案集合 $S_B = \{s_{1B}, s_{2B}, s_{3B}, \dots, s_{IB}\}$ ，根据分割策略 s_{iB} 将属于节点 h 的数据分成左右两个子节点 $2*h$ 和 $2*h+1$ 。计算子节点 $2*h$ 和 $2*h+1$ 的数据标签的和：

$$\text{sum_label}_{s_{iB_2h}} = \sum_{p \text{ 属于节点 } 2h} l_p;$$

$$25 \quad \text{sum_label}_{s_{iB_2h+1}} = \sum_{p \text{ 属于节点 } 2h+1} l_p;$$

以及两个集合中数据的个数：

$$\text{sum_num}_{s_{iB_2h}} = \sum_{p \text{ 属于节点 } 2h} 1;$$

$$\text{sum_num}_{s_{iB_2h+1}} = \sum_{p \text{ 属于节点 } 2h+1} 1;$$

其中 $r_{s_{iA_2h}}$ 表示一个随机数， $pk_B(r_{s_{iA_2h}})$ 表示用 pk_B 对随机数 $r_{s_{iA_2h}}$ 加密。

— 39 —

第 4 步：客户端 A 将 $\text{sum_pk_label}_{s_{iA-2h}}$ 和 $\text{sum_pk_label}_{s_{iA-2h-1}}$ 发送给客户端 B。

第 5 步：客户端 B 使用 sk_B 解密 $\text{sum_pk_label}_{s_{iA-2h}}$ 和 $\text{sum_pk_label}_{s_{iA-2h-1}}$ 得到 $\text{sum_r_label}_{s_{iA-2h}}$ 和 $\text{sum_r_label}_{s_{iA-2h-1}}$ 并发送给客户端 A。

客户端 A 收到 $\text{sum_r_label}_{s_{iA-2h}}$ 和 $\text{sum_r_label}_{s_{iA-2h-1}}$ 减去对应的随机数得到

5 $\text{sum_label}_{s_{iA-2h}}$ 和 $\text{sum_label}_{s_{iA-2h-1}}$ 。

第 6 步：客户端 A 使用 $\text{sum_label}_{s_{iA-2h}}$ 和 $\text{sum_label}_{s_{iA-2h-1}}$ 和 $\text{sum_num}_{s_{iA-2h}}$ 和 $\text{sum_num}_{s_{iA-2h-1}}$ 计算各种分割下的平均值。

$$\text{ave}_{s_{iA-2h}} = \frac{\text{sum_label}_{s_{iA-2h}}}{\text{sum_num}_{s_{iA-2h}}};$$

$$\text{ave}_{s_{iA-2h-1}} = \frac{\text{sum_label}_{s_{iA-2h-1}}}{\text{sum_num}_{s_{iA-2h-1}}};$$

10 使用 pk_B 对 $\text{ave}_{s_{iA-2h}}$ 和 $\text{ave}_{s_{iA-2h-1}}$ 进行加密得到 $\text{pk}_B(\text{ave}_{s_{iA-2h}})$ 和 $\text{pk}_B(\text{ave}_{s_{iA-2h-1}})$ ，并计算加密数据的方差值。

$$\text{var_pk}_{s_{iA-2h}} = \left(\sum_{p \text{ 属于节点 } 2h} \text{pk}_B(l_p^2) + \text{pk}_B(\text{ave}_{s_{iA-2h}}) - 2 * \text{pk}_B(l_p) * \text{ave}_{s_{iA-2h}} \right) + \text{pk}_B(r_{s_{iA-2h}});$$

$$\text{var_pk}_{s_{iA-2h-1}} = \left(\sum_{p \text{ 属于节点 } 2h} \text{pk}_B(l_p^2) + \text{pk}_B(\text{ave}_{s_{iA-2h-1}}) - 2 * \text{pk}_B(l_p) * \text{ave}_{s_{iA-2h-1}} \right) +$$

15 $\text{pk}_B(r_{s_{iA-2h-1}});$

客户端 B 使用 $\text{sum_label}_{s_{iB-2h}}$ 和 $\text{sum_label}_{s_{iB-2h-1}}$ 和 $\text{sum_num}_{s_{iB-2h}}$ 和 $\text{sum_num}_{s_{iB-2h-1}}$ 计算各种分割下的平均值：

$$\text{ave}_{s_{iB-2h}} = \frac{\text{sum_label}_{s_{iB-2h}}}{\text{sum_num}_{s_{iB-2h}}};$$

$$\text{ave}_{s_{iB-2h-1}} = \frac{\text{sum_label}_{s_{iB-2h-1}}}{\text{sum_num}_{s_{iB-2h-1}}};$$

20 并计算方差值：

$$\text{var}_{s_{iB-2h}} = \sum_{p \text{ 属于节点 } 2h} l_p^2 + \text{ave}_{s_{iB-2h}} - 2 * l_p * \text{ave}_{s_{iA-2h}};$$

$$\text{var}_{s_{iB-2h-1}} = \sum_{p \text{ 属于节点 } 2h+1} l_p^2 + \text{ave}_{s_{iB-2h-1}} - 2 * l_p * \text{ave}_{s_{iA-2h-1}};$$

第 7 步：客户端 A 将 $\text{var_pk}_{s_{iA-2h}}$ 和 $\text{var_pk}_{s_{iA-2h-1}}$ 发送给客户端 B。

第 8 步: 客户端 B 使用 sk_B 解密 $var_pk_{s_{iA-2h}}$ 和 $var_pk_{s_{iA-2h-1}}$ 得到 $var_r_{s_{iA-2h}}$ 和 $var_r_{s_{iA-2h-1}}$ 并发送给客户端 A。

客户端 A 收到 $var_r_{s_{iA-2h}}$ 和 $var_r_{s_{iA-2h-1}}$ 减去对应的随机数得到 $var_{s_{iA-2h}}$ 和 $var_{s_{iA-2h-1}}$ 。

第 10 步: 客户端 A 和 B 分别选择方差最小的划分记为 s_{minA} 以及方差值 var_{minA} 和 s_{minB} 以及方差值 var_{minB} 。

第 11 步: 客户端 A 发送 var_{minA} 给客户端 B, 客户端 B 比较大小, 并将比较结果返回给 A。B 推理树第 h 个节点标记为 gini 值小的一方的编号。

第 12 步: 方差小的一方根据对应的数据分割方案, 把数据进行分割并将分割结果发送给对方, 并将分割策略写入分割树的第 h 个节点。

第 13 步: $h = h+1$ 重复第 3 到 7 直到指定的重复步数。

第 14 步: B 统计叶子节点哪个类别多则标记该叶子节点为该类。

推理过程:

步骤 1: 根据推理树, 选择处理方是 A 和 B, 选择处理方。

步骤 2: 根据分割树的分割策略, 选择下一个节点的位置。

重复 1 和 2 直到叶子节点, 分类结果为该类标记的类别。

在图 1 至图 13 所对应的实施例的基础上, 为了更好的实施本申请实施例的上述方案, 下面还提供用于实施上述方案的相关设备。具体参阅图 14, 图 14 为本申请实施例提供的机器学习模型的训练装置的一种结构示意图。机器学习模型的训练装置 1400 应用于第一客户端, 多个客户端与服务器通信连接, 服务器中存储有多个模块, 多个模块用于构建机器学习模型, 第一客户端为多个客户端中的任一客户端, 机器学习模型的训练装置 1400 用于执行多轮迭代, 机器学习模型的训练装置 1400 包括: 获取单元 1401、训练单元 1402 和发送单元 1403, 在多轮迭代中的一轮迭代中, 获取单元 1401, 用于获取至少一个第一机器学习模型, 至少一个第一机器学习模型为根据机器学习模型的训练装置存储的第一训练数据集的数据特性选取出来的; 训练单元 1402, 用于利用第一数据集对至少一个第一机器学习模型执行训练操作, 得到至少一个训练后的第一机器学习模型; 发送单元 1403, 用于将至少一个训练后的第一机器学习模型包括的至少一个更新后的模块发送给服务器, 更新后的模块用于供服务器更新存储的模块的权重参数。

本申请实施例中, 能够对不同数据特性的训练数据分配不同的神经网络, 也即实现了神经网络与数据特性之间的个性化匹配; 此外, 由于第一客户端为多个客户端中的任一客户端, 对多个客户端中的每个客户端均根据客户端存储的训练数据集的数据特性分配并训练神经网络, 能够利用相同数据特性的训练数据训练相同的神经网络, 不同数据特性的训练数据训练不同的神经网络, 从而不仅实现了神经网络与数据特性之间的个性化匹配, 而且有利于提高训练后神经网络的准确率。

在一种可能的设计中, 多个模块用于构建至少两个第二机器学习模型, 至少一个第一机器学习模型为从至少两个第二机器学习模型中选取出来的; 或者, 用于构建至少一个第一机器学习模型的模块为从多个模块中选取出来的。

在一种可能的设计中，请参阅图 15，图 15 为本申请实施例提供的机器学习模型的训练装置的一种结构示意图。机器学习模型为神经网络，服务器中存储的多个模块为神经网络模块，机器学习模型的训练装置 1400 上存储有第一适配关系，第一适配关系包括多个适配值，适配值用于表示第一数据集合与第二神经网络之间的适配程度；装置 1400 还包括：
5 接收单元 1404，用于接收服务器发送的多个神经网络模块；获取单元 1401，具体用于根据第一适配关系，从至少两个第二神经网络中选取至少一个第一神经网络，至少一个第一神经网络中包括与第一数据集合的适配值高的至少一个第一神经网络。

在一种可能的设计中，第一数据集合与一个第二神经网络之间的适配值与第一损失函数的函数值对应，第一损失函数的函数值越小，第一数据集合与一个第二神经网络之间的
10 适配值越大；其中，第一损失函数指示第一训练数据的预测结果与第一数据的正确结果之间的相似度，第一数据的预测结果通过一个第二神经网络得到，第一数据和第一数据的正确结果基于第一数据集合得到。

在一种可能的设计中，第一数据集合与一个第二神经网络之间的适配值与第一相似度对应，第一相似度越大，第一数据集合与一个第二神经网络之间的适配值越大；其中，
15 第一相似度指的是一个第二神经网络和第三神经网络之间的相似度，第三神经网络为上一轮迭代中输出预测结果的准确率最高的神经网络。

在一种可能的设计中，一个第二神经网络和第三神经网络之间的相似度通过以下任一种方式确定：将相同数据分别输入至一个第二神经网络和第三神经网络，并对比一个第二神经网络的输出数据与第三神经网络的输出数据之间的相似度；或者，计算一个第二神经
20 网络的权重参数矩阵和第三神经网络的权重参数矩阵之间的相似度。

在一种可能的设计中，请参阅图 15，机器学习模型为神经网络，装置 1400 还包括：
接收单元 1404 和输入单元 1405；接收单元 1404，用于接收服务器发送的选择器，选择器为用于从多个神经网络模块中选取与第一数据集合的数据特征匹配的至少一个神经网络模块的神经网络；输入单元 1405，用于根据第一数据集合，将训练数据输入至选择器，得到
25 选择器输出的指示信息，指示信息包括多个神经网络模块中每个神经网络模块被选中的概率，用于指示构建至少一个第一神经网络的神经网络模块；接收单元 1404，还用于从服务端接收用于构建至少一个第一神经网络的神经网络模块。

在一种可能的设计中，请参阅图 15，机器学习模型为神经网络，服务器中存储的多个模块为神经网络模块，装置 1400 还包括：计算单元 1406，用于计算第一数据集合与至少
30 一个第一神经网络中每个第一神经网络之间的适配值；其中，第一数据集合包括多个第一训练数据，第一训练数据与第一神经网络之间的适配值越高，在利用第一训练数据对第一神经网络进行一次训练的过程中，对第一神经网络的权重参数的修改程度越大。

在一种可能的设计中，请参阅图 15，计算单元 1406，具体用于：对第一数据集合进行聚类，得到至少两个数据子集合，第一数据子集合为第一数据集合的子集，第一数据子集合为至少两个数据子集合中的任一个；根据第一数据子集合和第一损失函数，生成第一数据子集合与一个第一神经网络之间的适配值，第一损失函数的函数值越小，第一数据子集合与一个第一神经网络之间的适配值越大；其中，第一损失函数指示第一训练数据的预测
35

结果与第一数据的正确结果之间的相似度，第一数据的预测结果通过一个第一神经网络得到，第一数据和第一数据的正确结果基于第一数据子集合得到，第一数据子集合与一个第一神经网络之间的适配值被确定为第一数据子集合中每个数据与一个第一神经网络之间的适配值。

5 在一种可能的设计中，机器学习模型为神经网络，服务器中存储的多个模块为神经网络模块；训练单元 1402，具体用于根据第二损失函数，利用第一数据集合对第一神经网络执行训练操作；其中，第一数据集合包括多个第一训练数据，第二损失函数指示第一预测结果与第一训练数据的正确结果之间的相似度，还指示第一预测结果与第二预测结果之间的相似度，第一预测结果为将第一训练数据输入第一神经网络后，由第一神经网络输出的
10 第一训练数据的预测结果，第二预测结果为将第一训练数据输入第四神经网络后，由第四神经网络输出的第一训练数据的预测结果，第四神经网络为未执行过训练操作的第一神经网络。

在一种可能的设计中，请参阅图 15，第一数据集合包括多个第一训练数据和每个第一训练数据的正确结果；接收单元 1404，还用于接收服务器发送的选择器，选择器为用于从
15 多个神经网络模块中选取与第一数据集合的数据特征匹配的至少一个第一神经网络模块的神经网络；训练单元 1402，具体用于：将第一训练数据输入选择器，得到选择器输出的指示信息，指示信息包括多个神经网络模块中每个神经网络模块被选中的概率，用于指示构建第一神经网络的神经网络模块；根据多个神经网络模块，指示信息和第一训练数据，得到第一神经网络输出的第一训练数据的预测结果；根据第三损失函数，对第一神经网络和
20 选择器执行训练操作，其中，第三损失函数指示第一训练数据的预测结果与正确结果之间的相似度，还指示该指示信息的离散程度；发送单元 1403，还用于向服务器发送训练后的选择器。

需要说明的是，机器学习模型的训练装置 1400 中各模块/单元之间的信息交互、执行过程等内容，与本申请中图 4 至图 13 对应的各个方法实施例基于同一构思，具体内容可参
25 见本申请前述所示的方法实施例中的叙述，此处不再赘述。

本申请实施例还提供一种机器学习模型的训练装置，具体参阅图 16，图 16 为本申请实施例提供的机器学习模型的训练装置的一种结构示意图。机器学习模型的训练装置 1600 应用于服务器，服务器与多个客户端通信连接，服务器中存储有多个模块，多个模块用于构建机器学习模型，第一客户端为多个客户端中的任一客户端，机器学习模型的训练装置
30 1600 用于执行多轮迭代，机器学习模型的训练装置 1600 包括：获取单元 1601、发送单元 1602 和更新单元 1603，在多轮迭代中的一轮迭代中，获取单元 1601，用于获取与第一客户端对应的至少一个第一机器学习模型，第一客户端为多个客户端中的一个客户端，至少一个第一机器学习模型与第一客户端存储的第一数据集合的数据特性对应；发送单元 1602，用于将至少一个第一机器学习模型发送给第一客户端，至少一个第一机器学习模型指示第一客户端利用第一数据集合对至少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型；更新单元 1603，用于从第一客户端接收至少一个训练后的第一机器学习模型包括的至少一个更新后的神经网络模块，并根据至少一个更新后的神经网络
35

模块更新存储的神经网络模块的权重参数。

本申请实施例中，通过上述方式，能够对不同数据特性的训练数据分配不同的神经网络，也即实现了神经网络与数据特性之间的个性化匹配；由于第一客户端为多个客户端中的任一客户端，对多个客户端中的每个客户端均根据客户端存储的训练数据集合的数据特性分配并训练神经网络，能够利用相同数据特性的训练数据训练相同的神经网络，不同数据特性的训练数据训练不同的神经网络，从而不仅实现了神经网络与数据特性之间的个性化匹配，而且有利于提高训练后神经网络的准确率；由服务器选择与各个客户端适配的神经网络，既避免了将所有神经网络模块发送给客户端，以减少对客户端存储资源的浪费；且避免了对客户端计算机资源的占用，有利于提高用户体验。

在一种可能的设计中，多个模块用于构建至少两个第二机器学习模型，至少一个第一机器学习模型为从至少两个第二机器学习模型中选取出来的；或者，用于构建至少一个第一机器学习模型的模块为从多个模块中选取出来的。

在一种可能的设计中，请参阅图 17，图 17 为本申请实施例提供的机器学习模型的训练装置的一种结构示意图。机器学习模型为神经网络，机器学习模型的训练装置 1600 中存储的多个模块为神经网络模块，机器学习模型的训练装置 1600 上存储有第二适配关系，第二适配关系中包括多个适配值，适配值用于表示客户端中存储的训练数据与第二神经网络之间的适配程度，装置 1600 还包括：接收单元 1604，用于接收第一客户端发送的第一数据集合与至少一个第二神经网络之间的适配值，并更新第二适配关系；获取单元 1601，具体用于根据第二适配关系，从多个第二神经网络中选取至少一个第一神经网络，至少一个第一神经网络包括与第一数据集合的适配值高的神经网络。

在一种可能的设计中，请参阅图 17，机器学习模型为神经网络，机器学习模型的训练装置 1600 中存储的多个模块为神经网络模块，装置 1600 还包括：接收单元 1604，用于接收第一客户端发送的第一标识信息，第一标识信息为第一神经网络的标识信息，或者，第一标识信息为构建第一神经网络的神经网络模块的标识信息；发送单元 1602，具体用于向第一客户端发送第一标识信息指向的第一神经网络，或者，向第一客户端发送第一标识信息指向的构建第一神经网络的神经网络模块。

在一种可能的设计中，请参阅图 17，机器学习模型为神经网络，机器学习模型的训练装置 1600 中存储的多个模块为神经网络模块，机器学习模型的训练装置 1600 还配置有选择器，装置还包括：接收单元 1604，用于接收第一客户端发送的至少一个类中心，对第一数据集合执行聚类操作后，得到至少一个数据子集合，至少一个类中心中的一个类中心为至少一个数据子集合中一个数据子集合的类中心；获取单元 1601，具体用于将类中心分别输入选择器，得到选择器输出的指示信息，并根据指示信息，确定构建至少一个第一神经网络的神经网络模块，指示信息包括多个神经网络模块中每个神经网络模块被选中的概率；发送单元 1602，具体用于将构建至少一个第一神经网络的神经网络模块发送给第一客户端。

在一种可能的设计中，请参阅图 17，机器学习模型为神经网络，机器学习模型的训练装置 1600 中存储的多个模块为神经网络模块，一个神经网络被分为至少两个子模块，机器学习模型的训练装置 1600 存储的神经网络模块被分为与至少两个子模块对应的至少两个

组，同一组中不同的神经网络模块的功能相同，根据至少一个更新后的神经网络模块更新存储的神经网络模块的权重参数之后，装置 1600 还包括：计算单元 1605，用于计算同一组包括的至少两个神经网络模块中不同的神经网络模块之间的相似度，并将相似度大于预设阈值的两个神经网络模块进行合并。

5 在一种可能的设计中，不同的神经网络模块包括第二神经网络模块和第一神经网络模块，第二神经网络模块和第一神经网络模块之间的相似度通过以下任一种方式确定：将相同数据分别输入至第二神经网络模块和第一神经网络模块，并对比第二神经网络模块的输出数据与第一神经网络模块的输出数据之间的相似度；或者，计算第二神经网络模块的权重参数矩阵和第一神经网络模块的权重参数矩阵之间的相似度。

10 需要说明的是，机器学习模型的训练装置 1600 中各模块/单元之间的信息交互、执行过程等内容，与本申请中图 4 至图 13 对应的各个方法实施例基于同一构思，具体内容可参见本申请前述所示的方法实施例中的叙述，此处不再赘述。

本申请实施例还提供一种服务器，请参阅图 18，图 18 为本申请实施例提供的服务器的一种结构示意图。服务器 1800 上可以部署有图 14 和图 15 对应实施例中所描述的机器学习模型的训练装置 1400，用于实现图 4 至图 13 对应实施例中第一客户端的功能。或者，
15 在第一客户端配置于服务器形态的设备中时，服务器 1800 上可以部署有图 16 和图 17 对应实施例中所描述的机器学习模型的训练装置 1600，用于实现图 4 至图 13 对应实施例中服务器的功能。具体的，服务器 1800 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（central processing units，CPU）1822（例如，一个或一个以上处
20 理器）和存储器 1832，一个或一个以上存储应用程序 1842 或数据 1844 的存储介质 1830（例如一个或一个以上海量存储设备）。其中，存储器 1832 和存储介质 1830 可以是短暂存储或持久存储。存储在存储介质 1830 的程序可以包括一个或一个以上模块（图示没标出），每个模块可以包括对服务器中的一系列指令操作。更进一步地，中央处理器 1822 可以设置为与存储介质 1830 通信，在服务器 1800 上执行存储介质 1830 中的一系列指令操作。

25 服务器 1800 还可以包括一个或一个以上电源 1826，一个或一个以上有线或无线网络接口 1850，一个或一个以上输入输出接口 1858，和/或，一个或一个以上操作系统 1841，例如 Windows Server™，Mac OS X™，Unix™，Linux™，FreeBSD™ 等等。

在一种情况下，本申请实施例中，在中央处理器 1822 用于执行图 4 至图 13 对应实施例中的第一客户端执行的机器学习模型的训练方法，具体的，机器学习模型的训练包括多
30 轮迭代，在多层迭代中的一轮迭代中，中央处理器 1822 具体用于：

获取至少一个第一机器学习模型，至少一个第一机器学习模型为根据第一客户端存储的第一数据集的数据特性选取出来的；利用第一数据集对至少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型；将至少一个训练后的第一机器学习模型包括的至少一个更新后的模块发送给服务器，更新后的模块用于供所述服务器更
35 新存储的模块的权重参数。

中央处理器 1822 还用于执行图 4 至图 13 中第一客户端执行的其他步骤，对于中央处理器 1422 执行图 4 至图 13 对应实施例中的第一客户端执行的步骤的具体实现方式以及带

来的有益效果，均可以参考图 4 至图 13 对应的各个方法实施例中的叙述，此处不再一一赘述。

在另一种情况下，本申请实施例中，在中央处理器 1822 用于执行图 4 至图 13 对应实施例中的服务器执行的机器学习模型的训练方法，具体的，机器学习模型的训练包括多轮迭代，在多轮迭代中的一轮迭代中，中央处理器 1822 具体用于：

获取与第一客户端对应的至少一个第一机器学习模型，第一客户端为多个客户端中的一个客户端，至少一个第一机器学习模型与第一客户端存储的第一数据集合的数据特性对应；将至少一个第一机器学习模型发送给第一客户端，至少一个第一机器学习模型指示第一客户端利用第一数据集合对至少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型；从第一客户端接收至少一个训练后的第一机器学习模型包括的至少一个更新后的神经网络模块，并根据至少一个更新后的神经网络模块更新存储的神经网络模块的权重参数。

中央处理器 1822 还用于执行图 4 至图 13 中服务器执行的其他步骤，对于中央处理器 1422 执行图 4 至图 13 对应实施例中的服务器执行的步骤的具体实现方式以及带来的有益效果，均可以参考图 4 至图 13 对应的各个方法实施例中的叙述，此处不再一一赘述。

本申请实施例还提供一种终端设备，请参阅图 19，图 19 为本申请实施例提供的终端设备的一种结构示意图。其中，在客户端配置于移动设备形态的设备上时，终端设备 1900 上可以部署有图 14 和图 15 对应实施例中所描述的机器学习模型的训练装置 1400，用于实现图 4 至图 13 对应实施例中第一客户端的功能。具体的，终端设备 1900 包括：接收器 1901、发射器 1902、处理器 1903 和存储器 1904（其中终端设备 1900 中的处理器 1903 的数量可以一个或多个，图 19 中以一个处理器为例），其中，处理器 1903 可以包括应用处理器 19031 和通信处理器 19032。在本申请的一些实施例中，接收器 1901、发射器 1902、处理器 1903 和存储器 1904 可通过总线或其它方式连接。

存储器 1904 可以包括只读存储器和随机存取存储器，并向处理器 1903 提供指令和数据。存储器 1904 的一部分还可以包括非易失性随机存取存储器（non-volatile random access memory, NVRAM）。存储器 1904 存储有处理器和操作指令、可执行模块或者数据结构，或者它们的子集，或者它们的扩展集，其中，操作指令可包括各种操作指令，用于实现各种操作。

处理器 1903 控制终端设备的操作。具体的应用中，终端设备的各个组件通过总线系统耦合在一起，其中总线系统除包括数据总线之外，还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见，在图中将各种总线都称为总线系统。

上述本申请实施例揭示的方法可以应用于处理器 1903 中，或者由处理器 1903 实现。处理器 1903 可以是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器 1903 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 1903 可以是通用处理器、数字信号处理器（digital signal processing, DSP）、微处理器或微控制器，还可进一步包括专用集成电路（application specific integrated circuit, ASIC）、现场可编程门阵列（field-programmable gate array, FPGA）或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。该处理器 1903 可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于
5 随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 1904，处理器 1903 读取存储器 1904 中的信息，结合其硬件完成上述方法的步骤。

接收器 1901 可用于接收输入的数字或字符信息，以及产生与终端设备的相关设置以及功能控制有关的信号输入。发射器 1902 可用于通过第一接口输出数字或字符信息；发射器
10 1902 还可用于通过第一接口向磁盘组发送指令，以修改磁盘组中的数据；发射器 1902 还可以包括显示屏等显示设备。

本申请实施例中，应用处理器 19031，用于执行图 4 至图 13 对应实施例中第一客户端的功能。需要说明的是，对于应用处理器 19031 执行图 4 至图 13 对应实施例中第一客户端的功能的具体实现方式以及带来的有益效果，均可以参考图 4 至图 13 对应的各个方法实施
15 例中的叙述，此处不再一一赘述。

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 至图 13 所示实施例描述的方法中第一客户端所执行的步骤；或者，使得计算机执行如前述图 4 至图 13 所示实施例描述的方法中服务器所执行的步骤。

本申请实施例中还提供一种包括计算机程序产品，当其在计算机上运行时，使得计算机执行如前述图 4 至图 13 所示实施例描述的方法中第一客户端所执行的步骤，或者，使得
20 计算机执行如前述图 4 至图 13 所示实施例描述的方法中服务器所执行的步骤。

本申请实施例中还提供一种电路系统，所述电路系统包括处理电路，所述处理电路配置为执行如前述图 4 至图 13 所示实施例描述的方法中第一客户端所执行的步骤，或者，所
25 述处理电路配置为执行如前述图 4 至图 13 所示实施例描述的方法中服务器所执行的步骤。

本申请实施例提供的机器学习模型的训练装置、客户端和服务器的具体可以为芯片，芯片包括：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，以使芯片执行上述图 4 至图 13 所示实施例描述的神经网络的训练的方法。可选地，所述存储
30 单元为所述芯片内的存储单元，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（random access memory，RAM）等。

具体的，请参阅图 20，图 20 为本申请实施例提供的芯片的一种结构示意图，所述芯片可以表现为神经网络处理器 NPU 200，NPU 200 作为协处理器挂载到主 CPU（Host CPU）
35 上，由 Host CPU 分配任务。NPU 的核心部分为运算电路 2003，通过控制器 2004 控制运算电路 2003 提取存储器中的矩阵数据并进行乘法运算。

在一些实现中，运算电路 2003 内部包括多个处理单元（Process Engine，PE）。在一些实现中，运算电路 2003 是二维脉动阵列。运算电路 2003 还可以是一维脉动阵列或者能够执行例如乘法和加法这样的数学运算的其它电子线路。在一些实现中，运算电路 2003 是通用的矩阵处理器。

5 举例来说，假设有输入矩阵 A，权重矩阵 B，输出矩阵 C。运算电路从权重存储器 2002 中取矩阵 B 相应的数据，并缓存在运算电路中每一个 PE 上。运算电路从输入存储器 2001 中取矩阵 A 数据与矩阵 B 进行矩阵运算，得到的矩阵的部分结果或最终结果，保存在累加器（accumulator）2008 中。

10 统一存储器 2006 用于存放输入数据以及输出数据。权重数据直接通过存储单元访问控制器（Direct Memory Access Controller，DMAC）2005，DMAC 被搬运到权重存储器 2002 中。输入数据也通过 DMAC 被搬运到统一存储器 2006 中。

BIU 为 Bus Interface Unit 即，总线接口单元 2010，用于 AXI 总线与 DMAC 和取指存储器（Instruction Fetch Buffer，IFB）2009 的交互。

15 总线接口单元 2010（Bus Interface Unit，简称 BIU），用于取指存储器 2009 从外部存储器获取指令，还用于存储单元访问控制器 2005 从外部存储器获取输入矩阵 A 或者权重矩阵 B 的原数据。

DMAC 主要用于将外部存储器 DDR 中的输入数据搬运到统一存储器 2006 或将权重数据搬运到权重存储器 2002 中或将输入数据数据搬运到输入存储器 2001 中。

20 向量计算单元 2007 包括多个运算处理单元，在需要的情况下，对运算电路的输出做进一步处理，如向量乘，向量加，指数运算，对数运算，大小比较等等。主要用于神经网络中非卷积/全连接层网络计算，如 Batch Normalization（批归一化），像素级求和，对特征平面进行上采样等。

25 在一些实现中，向量计算单元 2007 能将经处理的输出的向量存储到统一存储器 2006。例如，向量计算单元 2007 可以将线性函数和/或非线性函数应用到运算电路 2003 的输出，例如对卷积层提取的特征平面进行线性插值，再例如累加值的向量，用以生成激活值。在一些实现中，向量计算单元 2007 生成归一化的值、像素级求和的值，或二者均有。在一些实现中，处理过的输出的向量能够用作到运算电路 2003 的激活输入，例如用于在神经网络中的后续层中的使用。

30 控制器 2004 连接的取指存储器（instruction fetch buffer）2009，用于存储控制器 2004 使用的指令；统一存储器 2006，输入存储器 2001，权重存储器 2002 以及取指存储器 2009 均为 On-Chip 存储器。外部存储器私有于该 NPU 硬件架构。

其中，循环神经网络中各层的运算可以由运算电路 2003 或向量计算单元 2007 执行。

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第一方面方法的程序执行的集成电路。

35 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5 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 CLU、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，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10 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、ROM、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，服务器，或者网络设备等）执行本申请各个实施15 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。15 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20 输，例如，所述计算机指令可以从一个网站站点、计算机、服务器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、服务器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器、数据中心等数25 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。

权利要求

1、一种机器学习模型的训练方法，其特征在于，所述方法应用于第一客户端，多个客户端与服务器通信连接，所述服务器中存储有多个模块，所述多个模块用于构建机器学习模型，所述第一客户端为所述多个客户端中的任一客户端，所述机器学习模型的训练包括多轮迭代，所述多轮迭代中的一轮迭代包括：

获取至少一个第一机器学习模型，所述至少一个第一机器学习模型为根据所述第一客户端存储的第一数据集合的数据特性选取出来的；

利用所述第一数据集合对所述至少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型；

将所述至少一个训练后的第一机器学习模型包括的至少一个更新后的模块发送给所述服务器，所述更新后的模块用于供所述服务器更新存储的模块的权重参数。

2、根据权利要求1所述的方法，其特征在于，

所述多个模块用于构建至少两个第二机器学习模型，所述至少一个第一机器学习模型为从所述至少两个第二机器学习模型中选取出来的；或者，

用于构建所述至少一个第一机器学习模型的模块为从所述多个模块中选取出来的。

3、根据权利要求2所述的方法，其特征在于，机器学习模型为神经网络，所述服务器中存储的多个模块为神经网络模块，所述第一客户端上存储有第一适配关系，所述第一适配关系包括多个适配值，所述适配值用于表示所述第一数据集合与第二神经网络之间的适配程度；

所述获取至少一个第二机器学习模块之前，所述方法还包括：

接收所述服务器发送的所述多个神经网络模块；

所述获取至少一个机器学习模型包括：

根据所述第一适配关系，从所述至少两个第二神经网络中选取至少一个第一神经网络，所述至少一个第一神经网络中包括与所述第一数据集合的适配值高的神经网络。

4、根据权利要求3所述的方法，其特征在于，所述第一数据集合与一个第二神经网络之间的适配值与第一损失函数的函数值对应，所述第一损失函数的函数值越小，所述第一数据集合与所述一个第二神经网络之间的适配值越大；

其中，所述第一损失函数指示第一数据的预测结果与所述第一数据的正确结果之间的相似度，所述第一数据的预测结果通过所述一个第二神经网络得到，所述第一数据和所述第一数据的正确结果基于所述第一数据集合得到。

5、根据权利要求3所述的方法，其特征在于，所述第一数据集合与一个第二神经网络之间的适配值与第一相似度对应，所述第一相似度越大，所述第一数据集合与所述一个第二神经网络之间的适配值越大；

其中，所述第一相似度指的是所述一个第二神经网络和第三神经网络之间的相似度，所述第三神经网络为上一轮迭代中输出预测结果的准确率最高的神经网络。

6、根据权利要求5所述的方法，其特征在于，所述一个第二神经网络和第三神经网络之间的相似度通过以下任一种方式确定：

将相同数据分别输入至所述一个第二神经网络和所述第三神经网络，并对比所述一个第二神经网络的输出数据与所述第三神经网络的输出数据之间的相似度；或者，

计算所述一个第二神经网络的权重参数矩阵和所述第三神经网络的权重参数矩阵之间的相似度。

5 7、根据权利要求 1 所述的方法，其特征在于，机器学习模型为神经网络，所述方法还包括：

接收所述服务器发送的选择器，所述选择器为用于从所述多个神经网络模块中选取与
所述第一数据集合的数据特征匹配的至少一个神经网络模块的神经网络；

10 根据所述第一数据集合，将训练数据输入至所述选择器，得到所述选择器输出的指示信息，所述指示信息包括所述多个神经网络模块中每个神经网络模块被选中的概率，用于指示构建所述至少一个第一神经网络的神经网络模块；

从所述服务端接收用于构建所述至少一个第一神经网络的神经网络模块。

15 8、根据权利要求 1 至 7 任一项所述的方法，其特征在于，机器学习模型为神经网络，所述服务器中存储的多个模块为神经网络模块，所述获取至少一个第一机器学习模型之后，所述方法还包括：

计算所述第一数据集合与所述至少一个第一神经网络中每个第一神经网络之间的适配值；

20 其中，所述第一数据集合包括多个第一训练数据，所述第一训练数据与所述第一神经网络之间的适配值越高，在利用所述第一训练数据对所述第一神经网络进行一次训练的过程中，对所述第一神经网络的权重参数的修改程度越大。

9、根据权利要求 8 所述的方法，其特征在于，所述计算所述第一数据集合与所述至少一个第一神经网络中每个第一神经网络之间的适配值，包括：

对所述第一数据集合进行聚类，得到至少两个数据子集合，第一数据子集合为所述第一数据集合的子集，所述第一数据子集合为所述至少两个数据子集合中的任一个；

25 根据所述第一数据子集合和第一损失函数，生成所述第一数据子集合与一个第一神经网络之间的适配值，所述第一损失函数的函数值越小，所述第一数据子集合与所述一个第一神经网络之间的适配值越大；

30 其中，所述第一损失函数指示第一数据的预测结果与所述第一数据的正确结果之间的相似度，所述第一数据的预测结果通过所述一个第一神经网络得到，所述第一数据和所述第一数据的正确结果基于所述第一数据子集合得到，所述第一数据子集合与一个第一神经网络之间的适配值被确定为所述第一数据子集合中每个数据与所述一个第一神经网络之间的适配值。

35 10、根据权利要求 1 至 7 任一项所述的方法，其特征在于，机器学习模型为神经网络，所述服务器中存储的多个模块为神经网络模块，所述利用所述第一数据集合对所述至少一个第一机器学习模型执行训练操作，包括：

根据第二损失函数，利用所述第一数据集合对所述第一神经网络执行训练操作；

其中，所述第一数据集合包括多个第一训练数据，所述第二损失函数指示第一预测结

果与所述第一训练数据的正确结果之间的相似度，还指示所述第一预测结果与第二预测结果之间的相似度，所述第一预测结果为将所述第一训练数据输入所述第一神经网络后，由所述第一神经网络输出的所述第一训练数据的预测结果，所述第二预测结果为将所述第一训练数据输入第四神经网络后，由所述第四神经网络输出的所述第一训练数据的预测结果，所述第四神经网络为未执行过训练操作的第一神经网络。

11、根据权利要求 7 所述的方法，其特征在于，所述第一数据集合包括多个第一训练数据和每个第一训练数据的正确结果，所述方法还包括：

接收所述服务器发送的选择器，所述选择器为用于从所述多个神经网络模块中选取与所述第一数据集合的数据特征匹配的至少一个第一神经网络模块的神经网络；

所述利用所述第一数据集合对所述至少一个第一机器学习模型执行训练操作，包括：

将所述第一训练数据输入所述选择器，得到所述选择器输出的指示信息，所述指示信息包括所述多个神经网络模块中每个神经网络模块被选中的概率，用于指示构建所述第一神经网络的神经网络模块；

根据所述多个神经网络模块，所述指示信息和所述第一训练数据，得到所述第一神经网络输出的所述第一训练数据的预测结果；

根据第三损失函数，对所述第一神经网络和所述选择器执行训练操作，其中，所述第三损失函数指示所述第一训练数据的预测结果与正确结果之间的相似度，还指示所述指示信息的离散程度；

所述方法还包括：向所述服务器发送训练后的选择器。

12、一种机器学习模型的训练方法，其特征在于，所述方法应用于服务器，所述服务器与多个客户端通信连接，所述服务器中存储有多个模块，所述多个模块用于构建机器学习模型，所述第一客户端为所述多个客户端中的任一客户端，所述机器学习模型的训练包括多轮迭代，所述多轮迭代中的一轮迭代包括：

获取与第一客户端对应的至少一个第一机器学习模型，所述第一客户端为所述多个客户端中的一个客户端，所述至少一个第一机器学习模型与所述第一客户端存储的第一数据集合的数据特性对应；

将所述至少一个第一机器学习模型发送给所述第一客户端，所述至少一个第一机器学习模型指示所述第一客户端利用所述第一数据集合对所述至少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型；

从所述第一客户端接收所述至少一个训练后的第一机器学习模型包括的至少一个更新后的神经网络模块，并根据所述至少一个更新后的神经网络模块更新存储的神经网络模块的权重参数。

13、根据权利要求 12 所述的方法，其特征在于，

所述多个模块用于构建至少两个第二机器学习模型，所述至少一个第一机器学习模型为从所述至少两个第二机器学习模型中选取出来的；或者，

用于构建所述至少一个第一机器学习模型的模块为从所述多个模块中选取出来的。

14、根据权利要求 13 所述的方法，其特征在于，机器学习模型为神经网络，所述服务

器中存储的多个模块为神经网络模块，所述服务器上存储有第二适配关系，所述第二适配关系中包括多个适配值，所述适配值用于表示客户端中存储的训练数据与第二神经网络之间的适配程度，所述方法还包括：

接收所述第一客户端发送的所述第一数据集合与至少一个第二神经网络之间的适配值，并更新所述第二适配关系；

所述获取至少一个第一神经网络包括：

根据所述第二适配关系，从所述多个第二神经网络中选取所述至少一个第一神经网络，所述至少一个第一神经网络包括与所述第一数据集合的适配值高的神经网络。

15、根据权利要求 12 所述的方法，其特征在于，机器学习模型为神经网络，所述服务器中存储的多个模块为神经网络模块，所述方法还包括：

接收所述第一客户端发送的第一标识信息，所述第一标识信息为所述第一神经网络的标识信息，或者，所述第一标识信息为构建所述第一神经网络的神经网络模块的标识信息；

所述将所述至少一个第一机器学习模型发送给所述第一客户端，包括：

向所述第一客户端发送所述第一标识信息指向的所述第一神经网络，或者，向所述第一客户端发送所述第一标识信息指向的构建所述第一神经网络的神经网络模块。

16、根据权利要求 12 所述的方法，其特征在于，机器学习模型为神经网络，所述服务器中存储的多个模块为神经网络模块，所述服务器还配置有选择器，所述方法还包括：

接收所述第一客户端发送的至少一个类中心，对所述第一数据集合执行聚类操作后，得到至少一个数据子集合，所述至少一个类中心中的一个类中心为所述至少一个数据子集合中一个数据子集合的类中心；

所述获取与第一客户端对应的至少一个第一机器学习模型，包括：

将所述类中心分别输入所述选择器，得到所述选择器输出的指示信息，并根据所述指示信息，确定构建所述至少一个第一神经网络的神经网络模块，所述指示信息包括所述多个神经网络模块中每个神经网络模块被选中的概率；

所述将所述至少一个第一机器学习模型发送给所述第一客户端，包括：

将构建所述至少一个第一神经网络的神经网络模块发送给所述第一客户端。

17、根据权利要求 12 或 14 所述的方法，其特征在于，机器学习模型为神经网络，所述服务器中存储的多个模块为神经网络模块，一个神经网络被分为至少两个子模块，所述服务器存储的神经网络模块被分为与所述至少两个子模块对应的至少两个组，同一组中不同的神经网络模块的功能相同，所述根据所述至少一个更新后的神经网络模块更新存储的神经网络模块的权重参数之后，所述方法还包括：

计算同一组包括的至少两个神经网络模块中不同的神经网络模块之间的相似度，并将相似度大于预设阈值的两个神经网络模块进行合并。

18、根据权利要求 17 所述的方法，其特征在于，所述不同的神经网络模块包括第二神经网络模块和第一神经网络模块，所述第二神经网络模块和所述第一神经网络模块之间的相似度通过以下任一种方式确定：

将相同数据分别输入至所述第二神经网络模块和所述第一神经网络模块，并对比所述

第二神经网络模块的输出数据与所述第一神经网络模块的输出数据之间的相似度；或者，
计算所述第二神经网络模块的权重参数矩阵和所述第一神经网络模块的权重参数矩阵之间的相似度。

19、一种机器学习模型的训练装置，其特征在于，所述装置应用于第一客户端，多个
5 客户端与服务器通信连接，所述服务器中存储有多个模块，所述多个模块用于构建机器学习模型，所述第一客户端为所述多个客户端中的任一客户端，所述机器学习模型的训练装置用于执行多轮迭代，所述机器学习模型的训练装置包括：获取单元、训练单元和发送单元，在所述多轮迭代中的一轮迭代中，

所述获取单元，用于获取至少一个第一机器学习模型，所述至少一个第一机器学习模型为根据所述机器学习模型的训练装置存储的第一训练数据集的数据特性选取出来的；

所述训练单元，用于利用所述第一数据集对所述至少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型；

所述发送单元，用于将所述至少一个训练后的第一机器学习模型包括的至少一个更新后的模块发送给所述服务器，所述更新后的模块用于供所述服务器更新存储的模块的权重
15 参数。

20、根据权利要求 19 所述的装置，其特征在于，

所述多个模块用于构建至少两个第二机器学习模型，所述至少一个第一机器学习模型为从所述至少两个第二机器学习模型中选取出来的；或者，

用于构建所述至少一个第一机器学习模型的模块为从所述多个模块中选取出来的。

21、根据权利要求 20 所述的装置，其特征在于，机器学习模型为神经网络，所述服务器中存储的多个模块为神经网络模块，所述机器学习模型的训练装置上存储有第一适配关系，所述第一适配关系包括多个适配值，所述适配值用于表示所述第一数据集与第二神经网络之间的适配程度；

所述装置还包括：接收单元，用于接收所述服务器发送的所述多个神经网络模块；

所述获取单元，具体用于根据所述第一适配关系，从所述至少两个第二神经网络中选取至少一个第一神经网络，所述至少一个第一神经网络中包括与所述第一数据集的适配值高的至少一个第一神经网络。

22、根据权利要求 21 所述的装置，其特征在于，所述第一数据集与一个第二神经网络之间的适配值与第一损失函数的函数值对应，所述第一损失函数的函数值越小，所述第一数据集与所述一个第二神经网络之间的适配值越大；

其中，所述第一损失函数指示第一训练数据的预测结果与所述第一数据的正确结果之间的相似度，所述第一数据的预测结果通过所述一个第二神经网络得到，所述第一数据和所述第一数据的正确结果基于所述第一数据集得到。

23、根据权利要求 21 所述的装置，其特征在于，所述第一数据集与一个第二神经网络之间的适配值与第一相似度对应，所述第一相似度越大，所述第一数据集与所述一个第二神经网络之间的适配值越大；

其中，所述第一相似度指的是所述一个第二神经网络和第三神经网络之间的相似度，

所述第三神经网络为上一轮迭代中输出预测结果的准确率最高的神经网络。

24、根据权利要求 23 所述的装置，其特征在于，所述一个第二神经网络和第三神经网络之间的相似度通过以下任一种方式确定：

5 将相同数据分别输入至所述一个第二神经网络和所述第三神经网络，并对比所述一个第二神经网络的输出数据与所述第三神经网络的输出数据之间的相似度；或者，

计算所述一个第二神经网络的权重参数矩阵和所述第三神经网络的权重参数矩阵之间的相似度。

25、根据权利要求 19 所述的装置，其特征在于，机器学习模型为神经网络，所述装置还包括：接收单元和输入单元；

10 所述接收单元，用于接收所述服务器发送的选择器，所述选择器为用于从所述多个神经网络模块中选取与所述第一数据集合的数据特征匹配的至少一个神经网络模块的神经网络；

15 所述输入单元，用于根据所述第一数据集合，将训练数据输入至所述选择器，得到所述选择器输出的指示信息，所述指示信息包括所述多个神经网络模块中每个神经网络模块被选中的概率，用于指示构建所述至少一个第一神经网络的神经网络模块；

所述接收单元，还用于从所述服务端接收用于构建所述至少一个第一神经网络的神经网络模块。

26、根据权利要求 19 至 25 任一项所述的装置，其特征在于，机器学习模型为神经网络，所述服务器中存储的多个模块为神经网络模块，所述装置还包括：

20 计算单元，用于计算所述第一数据集合与所述至少一个第一神经网络中每个第一神经网络之间的适配值；

其中，所述第一数据集合包括多个第一训练数据，所述第一训练数据与所述第一神经网络之间的适配值越高，在利用所述第一训练数据对所述第一神经网络进行一次训练的过程中，对所述第一神经网络的权重参数的修改程度越大。

25 27、根据权利要求 26 所述的装置，其特征在于，所述计算单元，具体用于：

对所述第一数据集合进行聚类，得到至少两个数据子集合，第一数据子集合为所述第一数据集合的子集，所述第一数据子集合为所述至少两个数据子集合中的任一个；

30 根据所述第一数据子集合和第一损失函数，生成所述第一数据子集合与一个第一神经网络之间的适配值，所述第一损失函数的函数值越小，所述第一数据子集合与所述一个第一神经网络之间的适配值越大；

35 其中，所述第一损失函数指示第一训练数据的预测结果与所述第一数据的正确结果之间的相似度，所述第一数据的预测结果通过所述一个第一神经网络得到，所述第一数据和所述第一数据的正确结果基于所述第一数据子集合得到，所述第一数据子集合与一个第一神经网络之间的适配值被确定为所述第一数据子集合中每个数据与所述一个第一神经网络之间的适配值。

28、根据权利要求 19 至 25 任一项所述的装置，其特征在于，机器学习模型为神经网络，所述服务器中存储的多个模块为神经网络模块；

所述训练单元，具体用于根据第二损失函数，利用所述第一数据集合对所述第一神经网络执行训练操作；

其中，所述第一数据集合包括多个第一训练数据，所述第二损失函数指示第一预测结果与所述第一训练数据的正确结果之间的相似度，还指示所述第一预测结果与第二预测结果之间的相似度，所述第一预测结果为将所述第一训练数据输入所述第一神经网络后，由所述第一神经网络输出的所述第一训练数据的预测结果，所述第二预测结果为将所述第一训练数据输入第四神经网络后，由所述第四神经网络输出的所述第一训练数据的预测结果，所述第四神经网络为未执行过训练操作的第一神经网络。

29、根据权利要求 25 所述的装置，其特征在于，所述第一数据集合包括多个第一训练数据和每个第一训练数据的正确结果；

所述接收单元，还用于接收所述服务器发送的选择器，所述选择器为用于从所述多个神经网络模块中选取与所述第一数据集合的数据特征匹配的至少一个第一神经网络模块的神经网络；

所述训练单元，具体用于：

将所述第一训练数据输入所述选择器，得到所述选择器输出的指示信息，所述指示信息包括所述多个神经网络模块中每个神经网络模块被选中的概率，用于指示构建所述第一神经网络的神经网络模块；

根据所述多个神经网络模块，所述指示信息和所述第一训练数据，得到所述第一神经网络输出的所述第一训练数据的预测结果；

根据第三损失函数，对所述第一神经网络和所述选择器执行训练操作，其中，所述第三损失函数指示所述第一训练数据的预测结果与正确结果之间的相似度，还指示所述指示信息的离散程度；

所述发送单元，还用于向所述服务器发送训练后的选择器。

30、一种机器学习模型的训练装置，其特征在于，所述装置应用于服务器，所述服务器与多个客户端通信连接，所述服务器中存储有多个模块，所述多个模块用于构建机器学习模型，所述第一客户端为所述多个客户端中的任一客户端，所述机器学习模型的训练装置用于执行多轮迭代，所述机器学习模型的训练装置包括：获取单元、发送单元和更新单元，在所述多轮迭代中的一轮迭代中，

所述获取单元，用于获取与第一客户端对应的至少一个第一机器学习模型，所述第一客户端为所述多个客户端中的一个客户端，所述至少一个第一机器学习模型与所述第一客户端存储的第一数据集合的数据特性对应；

所述发送单元，用于将所述至少一个第一机器学习模型发送给所述第一客户端，所述至少一个第一机器学习模型指示所述第一客户端利用所述第一数据集合对所述至少一个第一机器学习模型执行训练操作，得到至少一个训练后的第一机器学习模型；

所述更新单元，用于从所述第一客户端接收所述至少一个训练后的第一机器学习模型包括的至少一个更新后的神经网络模块，并根据所述至少一个更新后的神经网络模块更新存储的神经网络模块的权重参数。

31、根据权利要求 30 所述的装置，其特征在于，

所述多个模块用于构建至少两个第二机器学习模型，所述至少一个第一机器学习模型为从所述至少两个第二机器学习模型中选取出来的；或者，

用于构建所述至少一个第一机器学习模型的模块为从所述多个模块中选取出来的。

5 32、根据权利要求 31 所述的装置，其特征在于，机器学习模型为神经网络，所述机器学习模型的训练装置中存储的多个模块为神经网络模块，所述机器学习模型的训练装置上存储有第二适配关系，所述第二适配关系中包括多个适配值，所述适配值用于表示客户端中存储的训练数据与第二神经网络之间的适配程度，所述装置还包括：

10 接收单元，用于接收所述第一客户端发送的所述第一数据集合与至少一个第二神经网络之间的适配值，并更新所述第二适配关系；

所述获取单元，具体用于根据所述第二适配关系，从所述多个第二神经网络中选取所述至少一个第一神经网络，所述至少一个第一神经网络包括与所述第一数据集合的适配值高的神经网络。

15 33、根据权利要求 30 所述的装置，其特征在于，机器学习模型为神经网络，所述机器学习模型的训练装置中存储的多个模块为神经网络模块，所述装置还包括：

接收单元，用于接收所述第一客户端发送的第一标识信息，所述第一标识信息为所述第一神经网络的标识信息，或者，所述第一标识信息为构建所述第一神经网络的神经网络模块的标识信息；

20 所述发送单元，具体用于向所述第一客户端发送所述第一标识信息指向的所述第一神经网络，或者，向所述第一客户端发送所述第一标识信息指向的构建所述第一神经网络的神经网络模块。

34、根据权利要求 30 所述的装置，其特征在于，机器学习模型为神经网络，所述机器学习模型的训练装置中存储的多个模块为神经网络模块，所述机器学习模型的训练装置还配置有选择器，所述装置还包括：

25 接收单元，用于接收所述第一客户端发送的至少一个类中心，对所述第一数据集合执行聚类操作后，得到至少一个数据子集合，所述至少一个类中心中的一个类中心为所述至少一个数据子集合中一个数据子集合的类中心；

30 所述获取单元，具体用于将所述类中心分别输入所述选择器，得到所述选择器输出的指示信息，并根据所述指示信息，确定构建所述至少一个第一神经网络的神经网络模块，所述指示信息包括所述多个神经网络模块中每个神经网络模块被选中的概率；

所述发送单元，具体用于将构建所述至少一个第一神经网络的神经网络模块发送给所述第一客户端。

35 35、根据权利要求 30 或 32 所述的装置，其特征在于，机器学习模型为神经网络，所述机器学习模型的训练装置中存储的多个模块为神经网络模块，一个神经网络被分为至少两个子模块，所述机器学习模型的训练装置存储的神经网络模块被分为与所述至少两个子模块对应的至少两个组，同一组中不同的神经网络模块的功能相同，所述装置还包括：

计算单元，用于计算同一组包括的至少两个神经网络模块中不同的神经网络模块之间

的相似度，并将相似度大于预设阈值的两个神经网络模块进行合并。

36、根据权利要求 35 所述的装置，其特征在于，所述不同的神经网络模块包括第二神经网络模块和第一神经网络模块，所述第二神经网络模块和所述第一神经网络模块之间的相似度通过以下任一种方式确定：

- 5 将相同数据分别输入至所述第二神经网络模块和所述第一神经网络模块，并对比所述第二神经网络模块的输出数据与所述第一神经网络模块的输出数据之间的相似度；或者，
 计算所述第二神经网络模块的权重参数矩阵和所述第一神经网络模块的权重参数矩阵之间的相似度。

- 10 37、一种训练设备，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指令，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 至 11 中任一项所述的方法，或者，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 12 至 18 中任一项所述的方法。

- 15 38、一种计算机可读存储介质，其特征在于，包括程序，当其在计算机上运行时，使得计算机执行如权利要求 1 至 11 中任一项所述的方法，或者，使得计算机执行如权利要求 12 至 18 中任一项所述的方法。

39、一种电路系统，其特征在于，所述电路系统包括处理电路，所述处理电路配置为执行如权利要求 1 至 11 中任一项所述的方法，或者，所述处理电路配置为执行如权利要求 12 至 18 中任一项所述的方法。

- 20 40、一种包含指令的计算机程序产品，当其在计算机上运行时，使得计算机执行如权利要求 1 至 18 中任一项所述的方法。

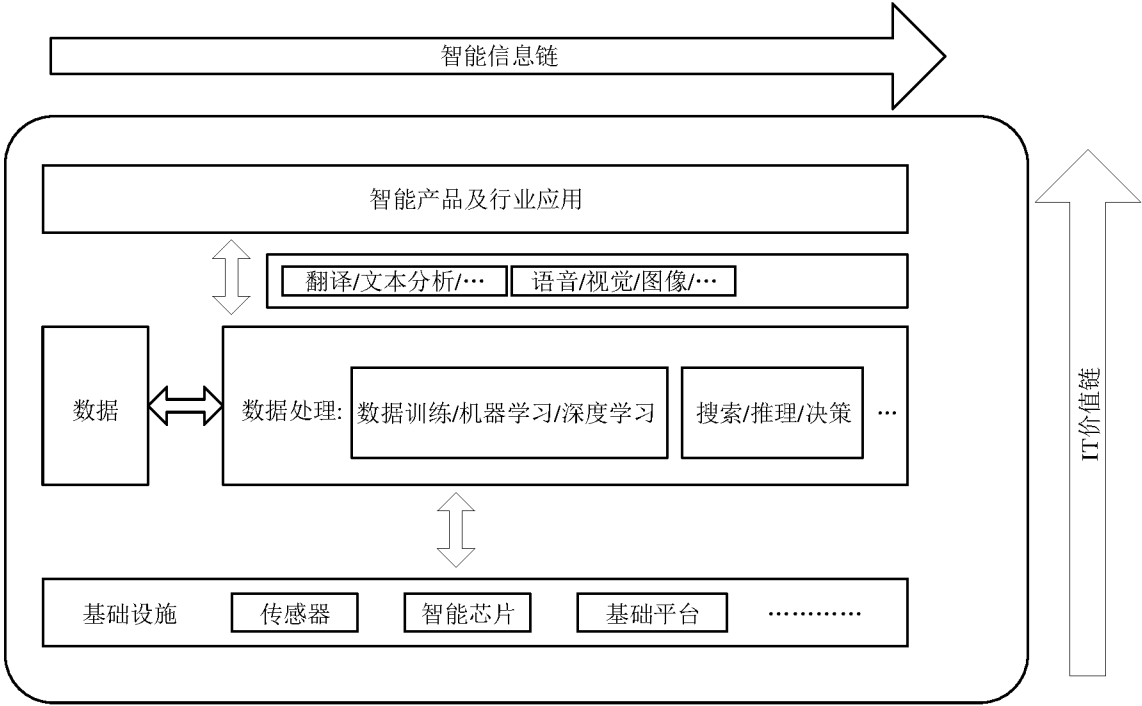


图 1

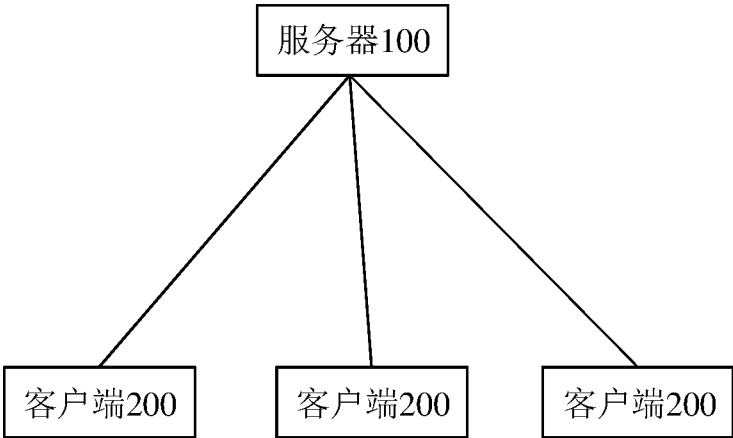


图 2

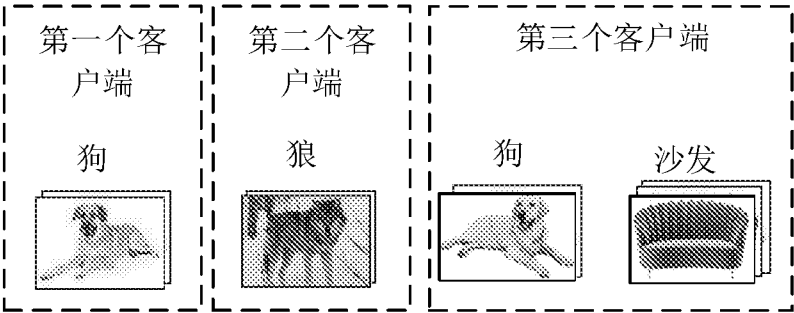


图 3

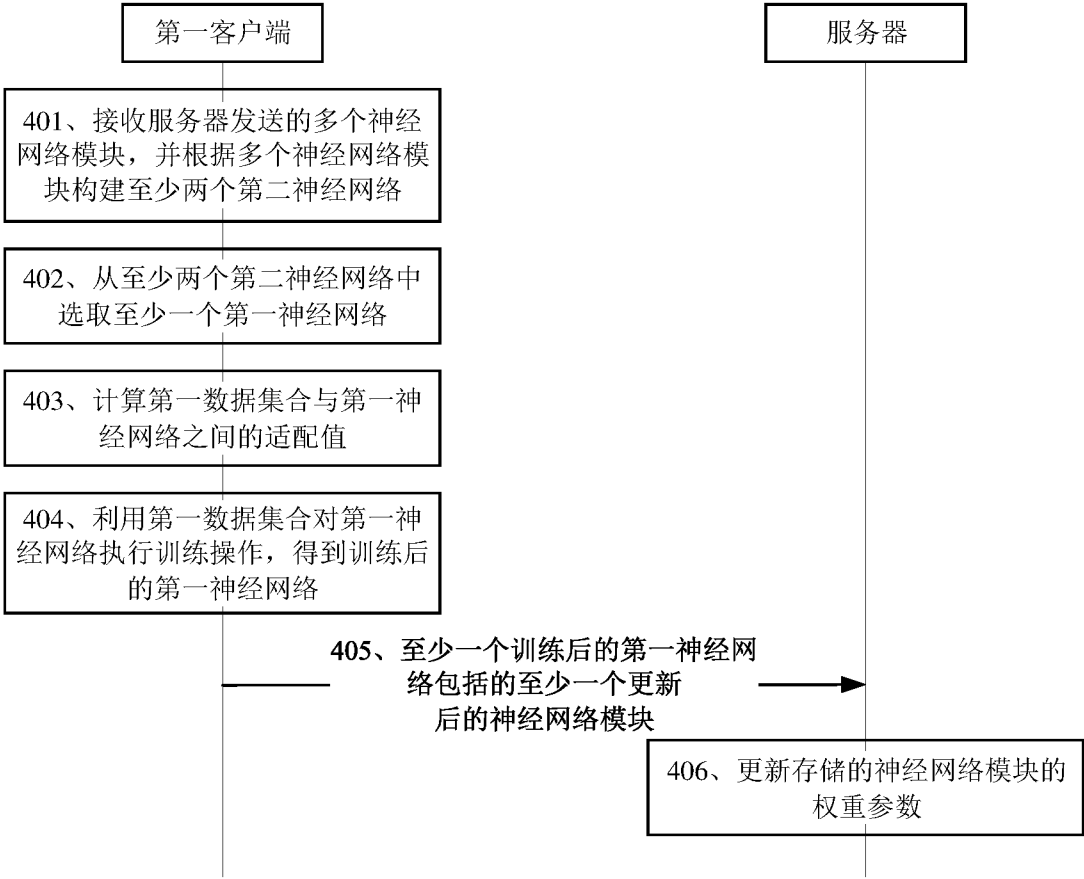


图 4

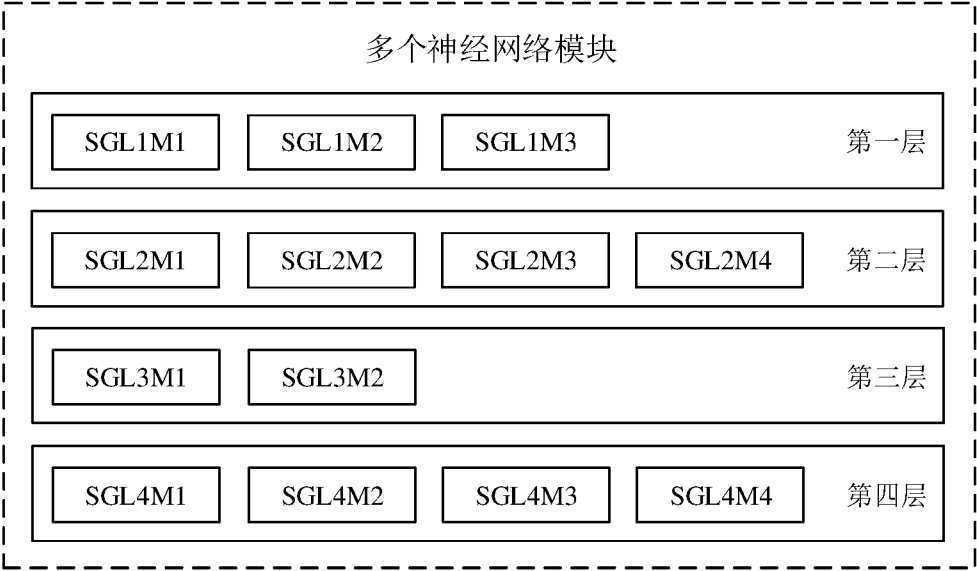


图 5

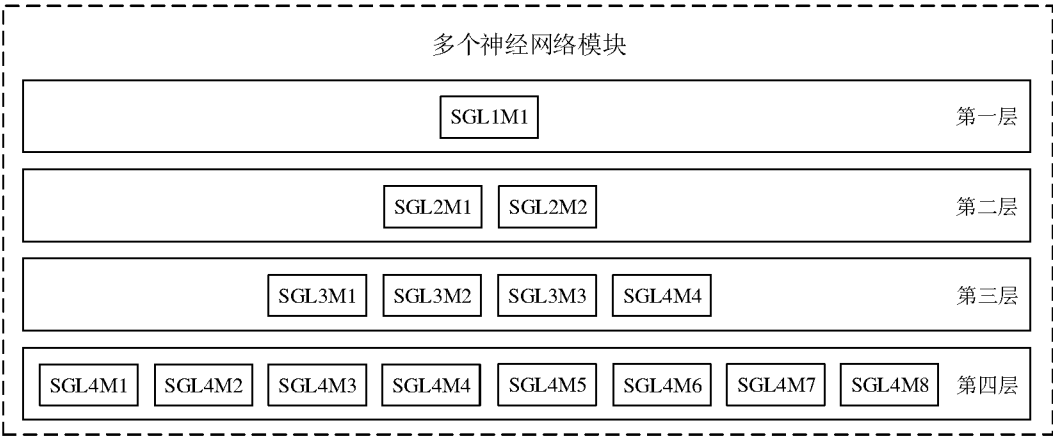


图 6

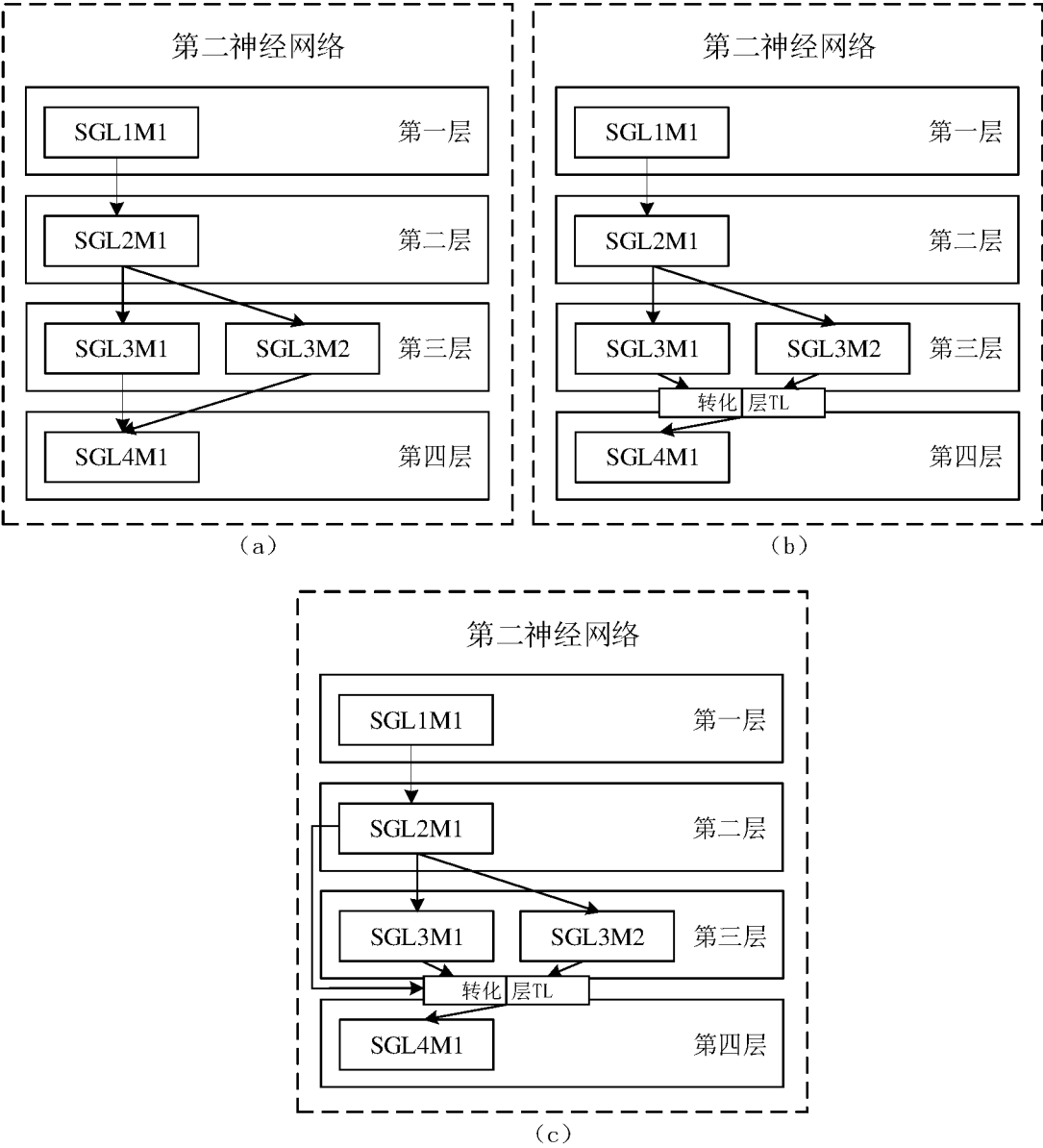


图 7

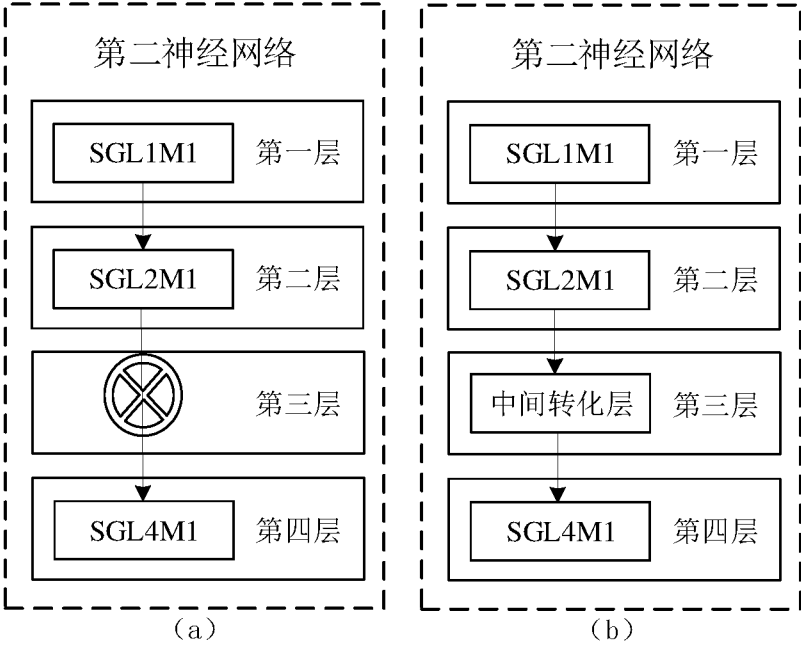


图 8

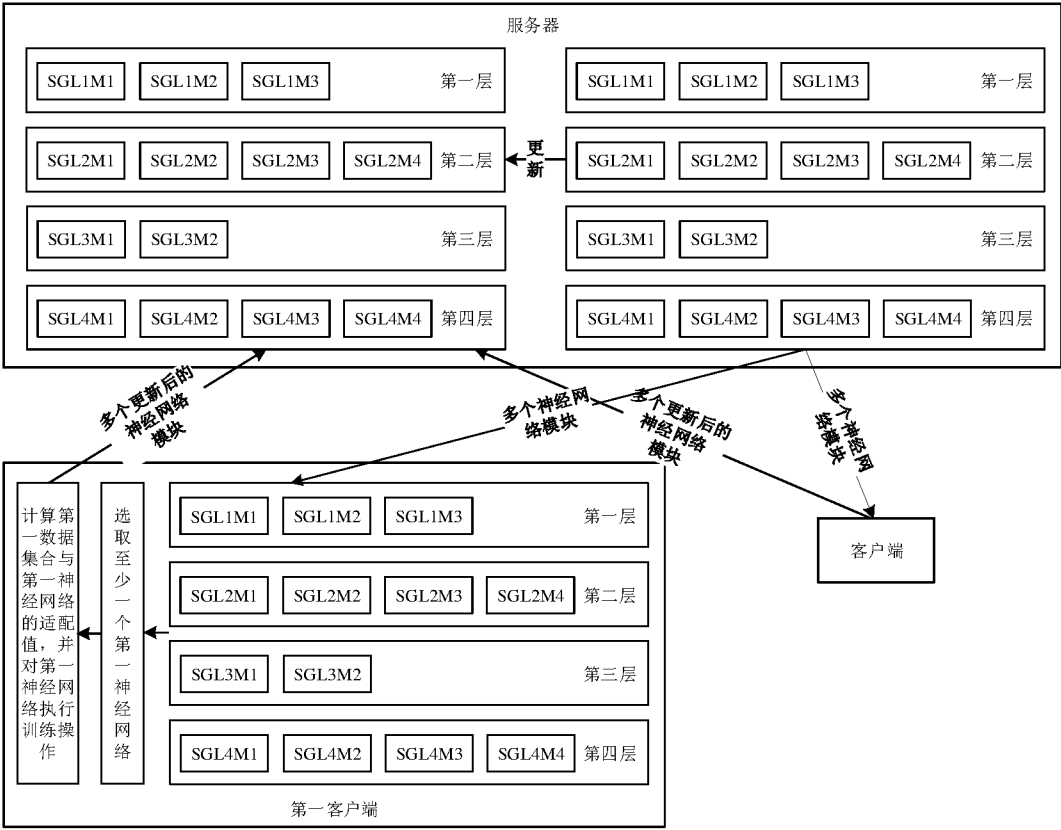


图 9

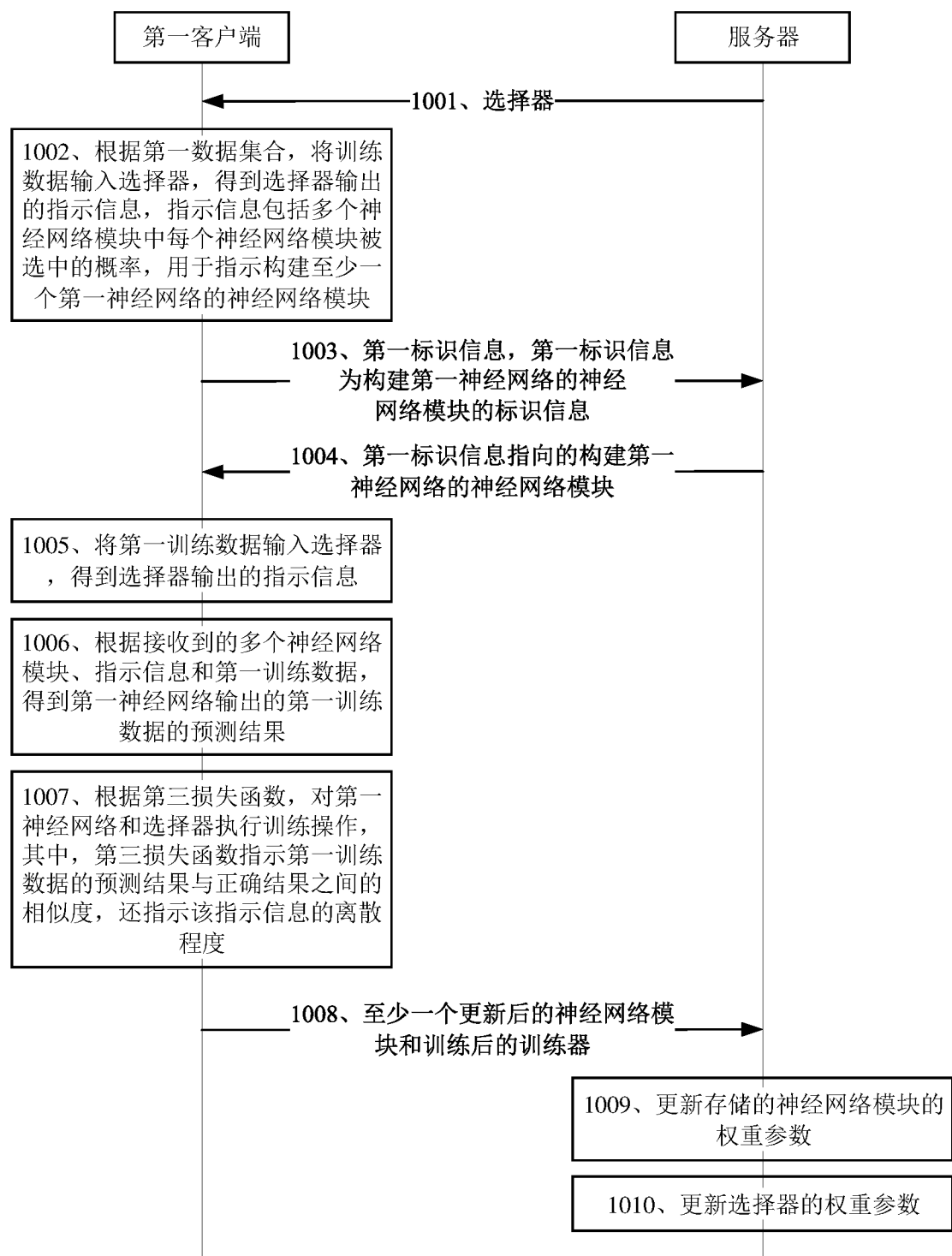


图 10

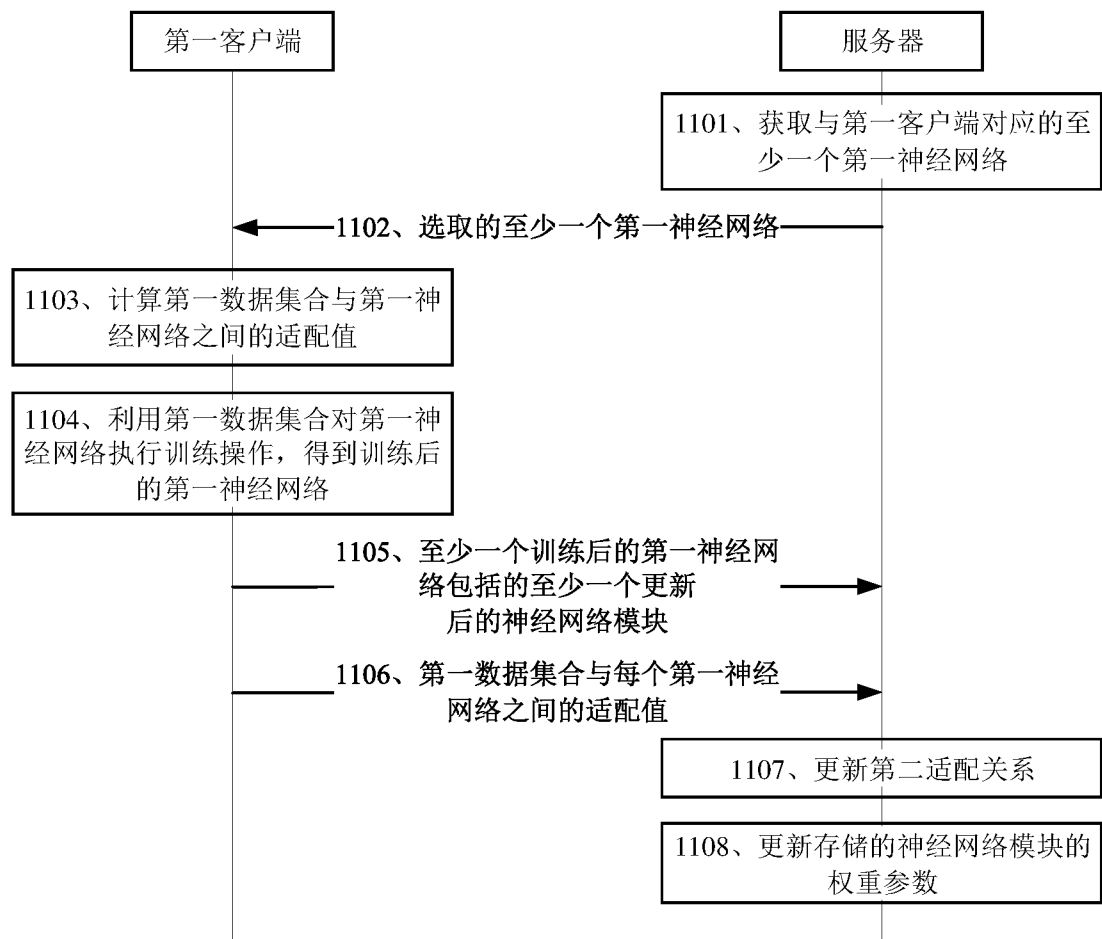


图 11

—8/11—

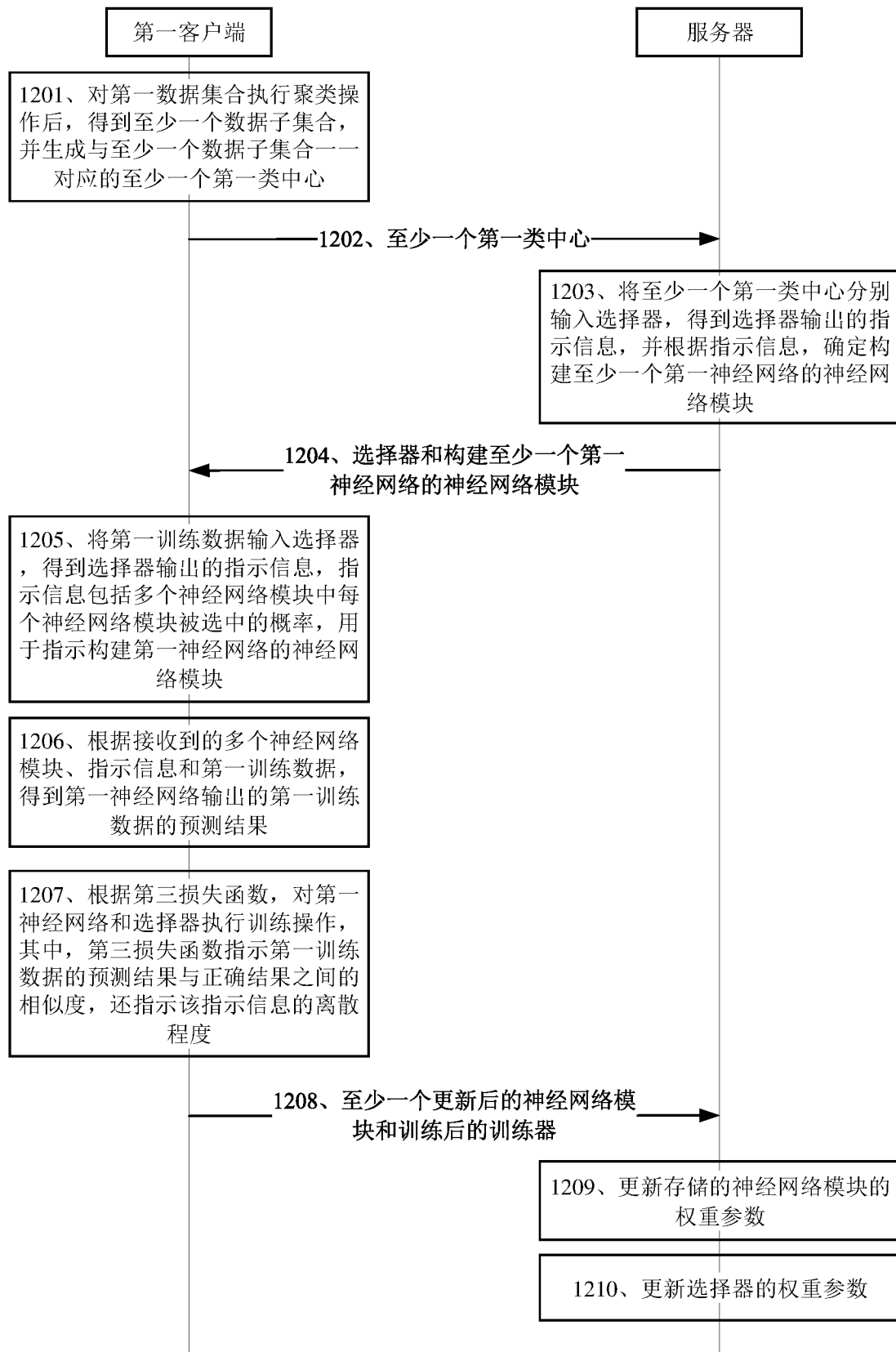


图 12

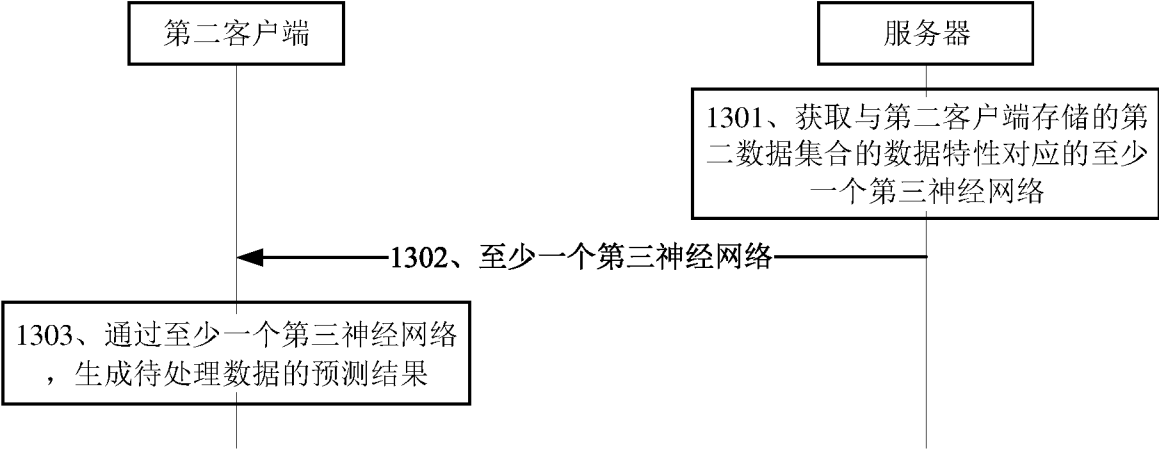


图 13

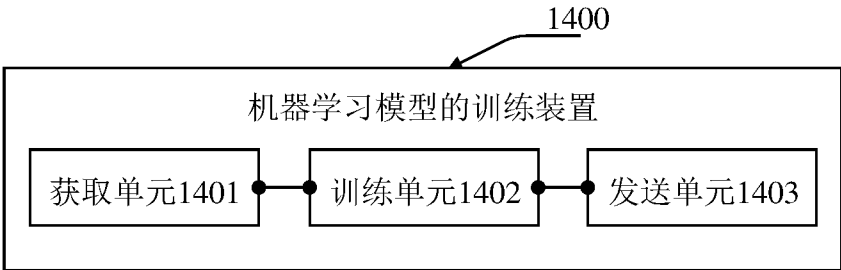


图 14

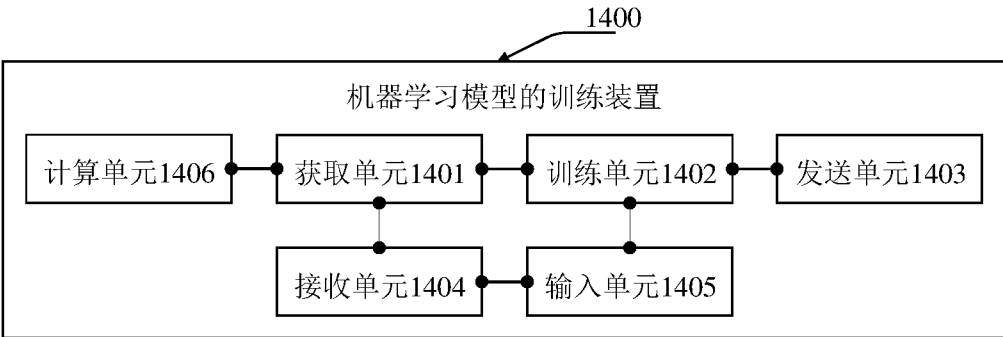


图 15

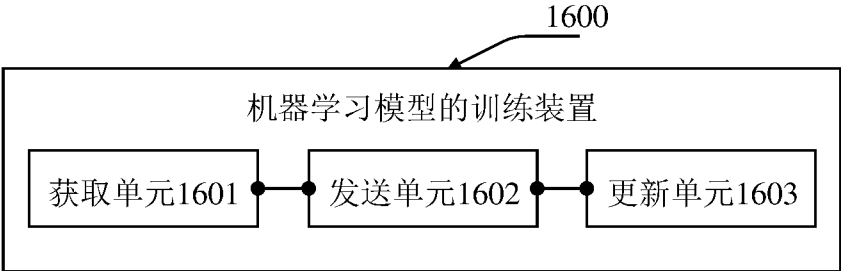


图 16

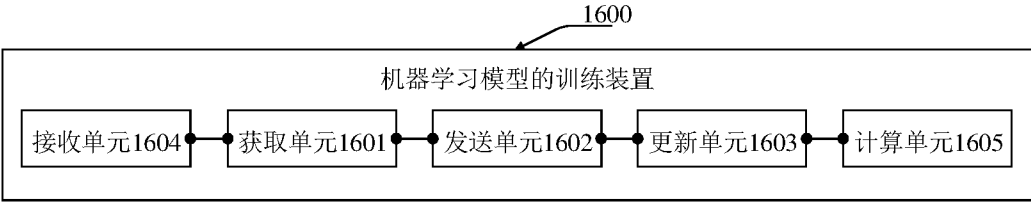


图 17

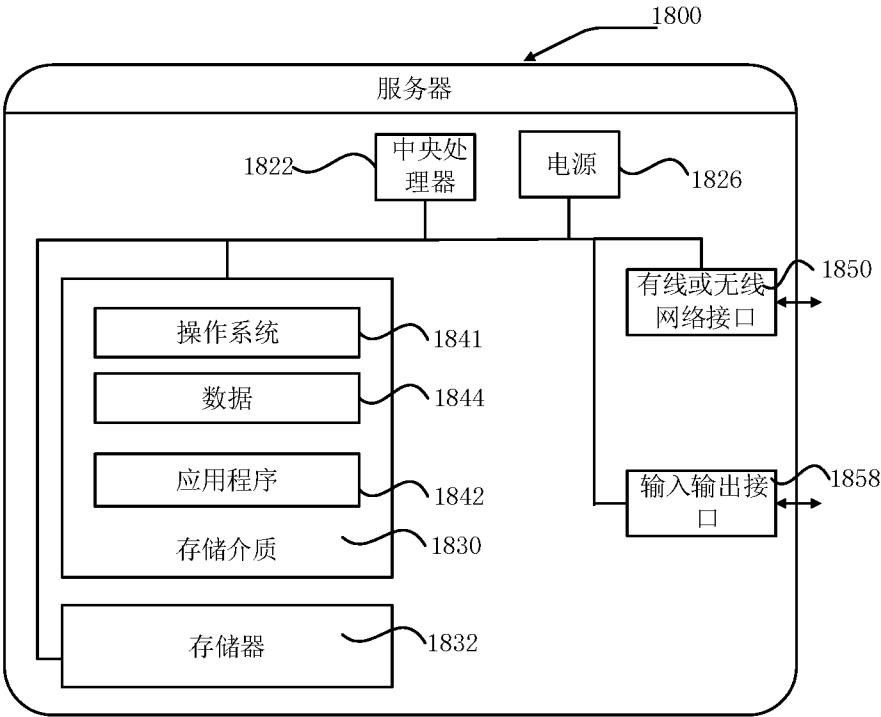


图 18

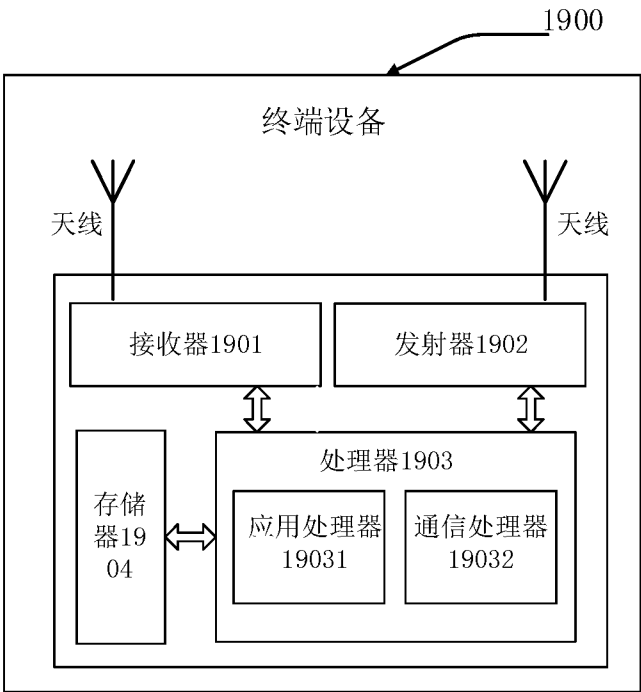


图 19

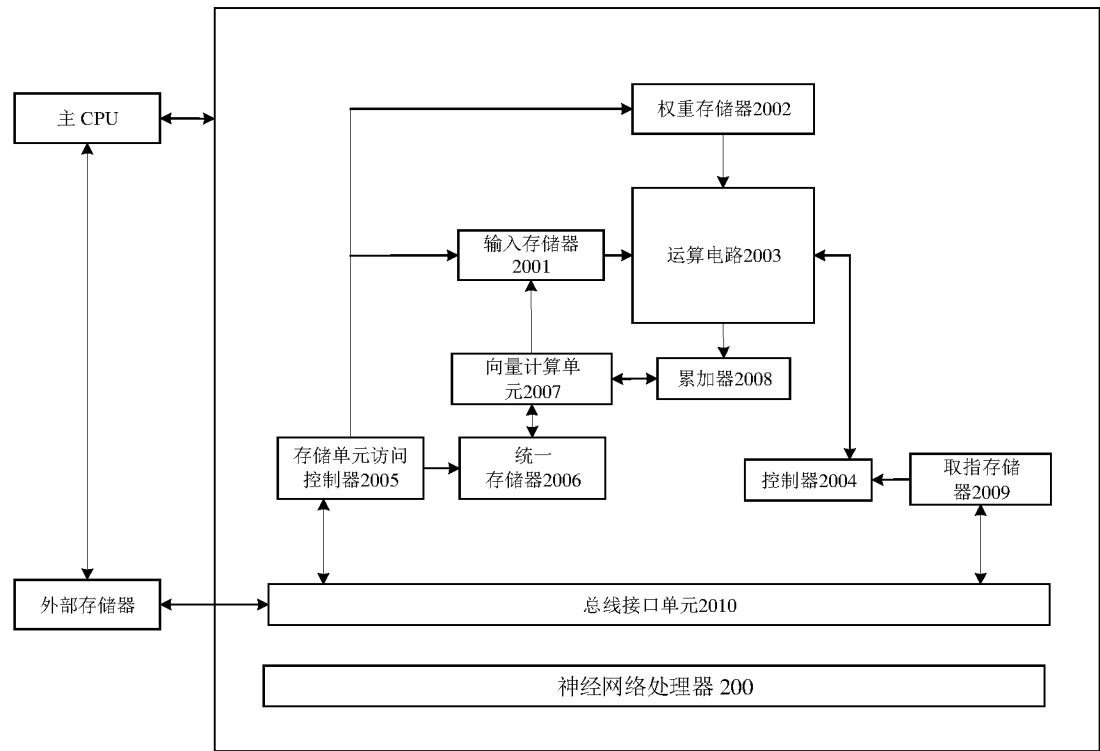


图 20

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/107391

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/04(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, EPODOC, WPI, CNKI, ISI: 机器, 学习, 模型, 客户端, 服务器, 训练, 数据, 特性, 对应, 选取, 更新, 参数, 权重,
machine, learning, model, client, server, training, data, character, corresponding, select, update, parameter, weight**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 110442457 A (PEKING UNIVERSITY SHENZHEN GRADUATE SCHOOL) 12 November 2019 (2019-11-12) description paragraphs [0049]-[0103], [0119]-[0126] and figures 1-3	1-40
A	CN 111477290 A (SHANGHAI JIAO TONG UNIVERSITY) 31 July 2020 (2020-07-31) entire document	1-40
A	CN 109117953 A (BEIJING MEGVII TECHNOLOGY CO., LTD.) 01 January 2019 (2019-01-01) entire document	1-40
A	CN 110717589 A (BEIJING KUANGSHI TECHNOLOGY CO., LTD.) 21 January 2020 (2020-01-21) entire document	1-40
A	US 2019385043 A1 (ADOBE INC.) 19 December 2019 (2019-12-19) entire document	1-40



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

08 October 2021

Date of mailing of the international search report

18 October 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/107391

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	110442457	A	12 November 2019	None	
CN	111477290	A	31 July 2020	None	
CN	109117953	A	01 January 2019	None	
CN	110717589	A	21 January 2020	None	
US	2019385043	A1	19 December 2019	None	

国际检索报告

国际申请号

PCT/CN2021/107391

A. 主题的分类 G06N 3/04 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类		
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06N 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, EPDOC, WPI, CNKI, ISI: 机器, 学习, 模型, 客户端, 服务器, 训练, 数据, 特性, 对应, 选取, 更新, 参数, 权重, machine, learning, model, client, server, training, data, character, corresponding, select, update, parameter, weight		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 110442457 A (北京大学深圳研究生院) 2019年 11月 12日 (2019 - 11 - 12) 说明书第[0049]-[0103]段、第[0119]-[0126]及图1-3	1-40
A	CN 111477290 A (上海交通大学) 2020年 7月 31日 (2020 - 07 - 31) 全文	1-40
A	CN 109117953 A (北京迈格威科技有限公司) 2019年 1月 1日 (2019 - 01 - 01) 全文	1-40
A	CN 110717589 A (北京旷视科技有限公司) 2020年 1月 21日 (2020 - 01 - 21) 全文	1-40
A	US 2019385043 A1 (ADOBE INC.) 2019年 12月 19日 (2019 - 12 - 19) 全文	1-40
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 2021年 10月 8日		国际检索报告邮寄日期 2021年 10月 18日
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国 北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 李婉怡 电话号码 86-(10)-53961408

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/107391

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	110442457	A	2019年 11月 12日	无	
CN	111477290	A	2020年 7月 31日	无	
CN	109117953	A	2019年 1月 1日	无	
CN	110717589	A	2020年 1月 21日	无	
US	2019385043	A1	2019年 12月 19日	无	