

(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 2 月 10 日 (10.02.2022)



(10) 国际公布号
WO 2022/028045 A1

- (51) 国际专利分类号:
G06N 20/00 (2019.01)
- (21) 国际申请号: PCT/CN2021/094938
- (22) 国际申请日: 2021 年 5 月 20 日 (20.05.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202010787792.7 2020年8月6日 (06.08.2020) CN
- (71) 申请人: 深圳前海微众银行股份有限公司
(**WEBANK CO., LTD**) [CN/CN]; 中国广东省深圳市南山区沙河西路1819号深圳湾科技生态园7栋A座, Guangdong 518000 (CN)。
- (72) 发明人: 高大山(**GAO, Dashan**); 中国广东省深圳市南山区沙河西路1819号深圳湾科技生态园7栋A座, Guangdong 518000 (CN)。 刘洋(**LIU, Yang**);

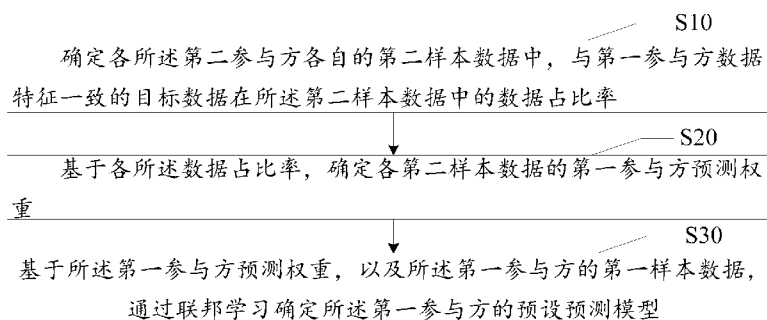
中国广东省深圳市南山区沙河西路1819号深圳湾科技生态园7栋A座, Guangdong 518000 (CN)。
杨强(**YANG, Qiang**); 中国广东省深圳市南山区沙河西路1819号深圳湾科技生态园7栋A座, Guangdong 518000 (CN)。 鞠策(**JU, Ce**); 中国广东省深圳市南山区沙河西路1819号深圳湾科技生态园7栋A座, Guangdong 518000 (CN)。

(74) 代理人: 深圳市世纪恒程知识产权代理事务所 (**CENFO INTELLECTUAL PROPERTY AGENCY**); 中国广东省深圳市南山区西丽街道松坪山社区松坪山路3号奥特迅电力大厦201, Guangdong 518052 (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, 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,

(54) **Title:** DATA PROCESSING METHOD, APPARATUS, AND DEVICE, AND MEDIUM

(54) 发明名称: 数据处理方法、装置、设备及介质



- S10 Determine a data proportion of target data, consistent with data characteristics of a first participant in second sample data of each second participant, in the second sample data
- S20 Determine a first participant prediction weight of each second sample data on the basis of each data proportion
- S30 Determine a preset prediction model of the first participant by means of federated learning on the basis of the first participant prediction weight and first sample data of the first participant

(57) **Abstract:** The present application discloses a data processing method, apparatus, and device, and a medium. The method comprises: determining a data proportion of target data, consistent with data characteristics of a first participant in second sample data of each second participant, in the second sample data; determining a first participant prediction weight of each second sample data on the basis of each data proportion; and determining a preset prediction model of the first participant by means of federated learning on the basis of the first participant prediction weight and first sample data of the first participant.

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: 本申请公开了一种数据处理方法、装置、设备和介质, 该方法包括: 确定各所述第二参与方各自的第二样本数据中, 与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率; 基于各所述数据占比率, 确定各第二样本数据的第一参与方预测权重; 基于所述第一参与方预测权重, 以及所述第一参与方的第一样本数据, 通过联邦学习确定所述第一参与方的预设预测模型。

数据处理方法、装置、设备及介质

本申请要求 2020 年 8 月 6 日申请的，申请号为 202010787792.7，名称为“数据处理方法、装置、设备及介质”的中国专利申请的优先权，在此将其全文引入作为参考。

技术领域

本申请涉及金融科技（Fintech）的人工智能技术领域，尤其涉及一种数据处理方法、装置、设备及介质。

背景技术

随着金融科技，尤其是互联网科技金融的不断发展，越来越多的技术（如分布式、区块链Blockchain、人工智能等）应用在金融领域，但金融业也对技术提出了更高的要求，如对金融业对数据处理也有更高的要求。

目前，一方面由于任何一个单一的参与方都不能收集足够的数据来准确构建有效的机器学习模型，另一方面由于隐私保护的需求，致使某一参与方通过联邦学习的方式联合不同的参与方合作训练适合自身任务的模型成为构建需要的目标机器学习模型的主要方式。

相关技术中，在某一参与方联合其他参与方联邦训练自身模型时，各个参与方是具有相同的训练比重或者预测比重的，各个参与方具有相同的训练比重或者预测比重，致使某一参与方数据中部分无效的训练数据占用训练比重或者预测比重，致使训练效率低下且得到的目标模型的预测性能低下。

发明内容

本申请的主要目的在于提供一种数据处理方法、装置、设备和介质，旨在解决相关技术中联邦训练效率低下且得到的联邦标模型的预测性能低下的技术问题。

为实现上述目的，本申请提供一种数据处理方法，应用于第一参与方，所述第一参与方与第二参与方进行通信连接，所述第二参与方为多个，所述数据处理方法包括：

确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

基于各所述数据占比率，确定各第二样本数据的第一参与方预测权重；

基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定所述第一参与方的预设预测模型。

5 可选地，所述确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率的步骤，包括：

接收每个第二参与方发送的数据占比率，其中，每个第二参与方将对应第二样本数据输入至相应预设领域分类模型中后，对所述第二样本数据进行预测处理，得到第二样本数据中，与第一参与方数据特征一致的目标数据在
10 所述第二样本数据中的数据占比率；

其中，所述预设领域分类模型是基于具有预设参与方来源标签的训练样本数据，通过执行第一预设联邦流程，对第一预设待训练预测模型进行迭代训练后，获得的预测各参与方数据来源占比率的第一目标模型。

15 可选地，第一参与方中的样本数据为第一样本数据，所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定所述第一参与方的预设预测模型的步骤，包括：

基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型；

20 将所述第二目标模型设置为所述第一参与方的预设预测模型。

可选地，所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型的步骤，包括：

25 接收第二参与方发送的第二模型初始梯度，其中，所述第二模型初始梯度是各第二参与方在执行第二预设联邦流程过程中，基于对应第二样本数据确定的；

基于所述第二模型初始梯度以及对应所述第一参与方预测权重，获取各第二参与方的第二模型更新梯度；

30 基于所述第二模型更新梯度，以及第一参与方在执行第二预设联邦流程中，基于所述第一样本数据确定的第一模型更新梯度，确定替换更新模型参

数；

基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获得第二目标模型。

可选地，所述第一参与方与第二参与方通过预设中间方通信连接；

5 所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型的步骤，包括：

在执行第二预设联邦流程中，基于所述第一样本数据确定第一模型更新梯度，并将所述第一模型更新梯度发送给预设中间方；

10 接收预设中间方发送的基于第二模型初始梯度，对应所述第一参与方预测权重以所述及第一模型更新梯度，确定的替换更新模型参数；

其中，第二参与方基于所述第二样本数据，确定第二模型初始梯度，并将第二模型初始梯度加密发送给所述预设中间方；

15 基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获得第二目标模型。

可选地，所述接收预设中间方发送的基于第二模型初始梯度，对应所述第一参与方预测权重以及第一模型更新梯度，确定的替换更新模型参数的步骤，包括：

20 接收预设中间方加密发送的基于第二模型初始梯度，对应所述第一参与方预测权重以及第一模型更新梯度，确定的替换更新模型参数。

可选地，所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，确定所述第一参与方的预设预测模型的步骤之后，所述包括：

获取待处理数据，将所述待处理数据输入至所述预设预测模型中；

25 基于所述预设预测模型对所述待处理数据进行预测处理，得到目标预测结果。

本申请还提供一种数据处理装置，应用于第一参与方，所述第一参与方与第二参与方进行通信连接，所述第二参与方为多个，所述数据处理装置包括：

30 第一确定模块，用于确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

第二确定模块，用于基于各所述数据占比率，确定各第二样本数据的第一参与方预测权重；

第三确定模块，用于基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定所述第一参与方的预设预测模型。

5 可选地，所述第一确定模块包括：

接收单元，用于接收每个第二参与方发送的数据占比率，其中，每个第二参与方将对应第二样本数据输入至相应预设领域分类模型中后，对所述第二样本数据进行预测处理，得到第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

10 其中，所述预设领域分类模型是基于具有预设参与方来源标签的训练样本数据，通过执行第一预设联邦流程，对第一预设待训练预测模型进行迭代训练后，获得的预测各参与方数据来源占比率的第一目标模型。

可选地，第一参与方中的样本数据为第一样本数据，所述第三确定模块包括：

15 第一执行单元，用于基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型；

设置单元，用于将所述第二目标模型设置为所述第一参与方的预设预测模型。

20 可选地，所述第一执行单元包括：

第一接收子单元，用于接收第二参与方发送的第二模型初始梯度，其中，所述第二模型初始梯度是各第二参与方在执行第二预设联邦流程过程中，基于对应第二样本数据确定的；

25 获取子单元，用于基于所述第二模型初始梯度以及对应所述第一参与方预测权重，获取各第二参与方的第二模型更新梯度；

确定子单元，用于基于所述第二模型更新梯度，以及第一参与方在执行第二预设联邦流程中，基于所述第一样本数据确定的第一模型更新梯度，确定替换更新模型参数；

30 第一更新子单元，用于基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获得第二目标模型。

可选地，所述第一参与方与第二参与方通过预设中间方通信连接；

所述第一执行单元还包括：

发送子单元，用于在执行第二预设联邦流程中，基于所述第一样本数据确定第一模型更新梯度，并将所述第一模型更新梯度发送给预设中间方；

5 第二接收子单元，用于接收预设中间方发送的基于第二模型初始梯度，对应所述第一参与方预测权重以所述及第一模型更新梯度，确定的替换更新模型参数；

其中，第二参与方基于所述第二样本数据，确定第二模型初始梯度，并将第二模型初始梯度加密发送给所述预设中间方；

10 第二更新子单元，用于基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获得第二目标模型。

可选地，所述第二接收子单元用于实现：

接收预设中间方加密发送的基于第二模型初始梯度，对应所述第一参与方预测权重以及第一模型更新梯度，确定的替换更新模型参数。

15 可选地，所述数据处理装置还包括：

第一获取模块，用于获取待处理数据，将所述待处理数据输入至所述预设预测模型中；

第二获取模块，用于基于所述预设预测模型对所述待处理数据进行预测处理，得到目标预测结果。

20 本申请还提供一种数据处理设备，所述数据处理设备为实体设备，所述数据处理设备包括：存储器、处理器以及存储在所述存储器上并可在所述处理器上运行的所述数据处理方法的程序，所述数据处理方法的程序被处理器执行时可实现如上述的数据处理方法的步骤。

本申请还提供一种介质，所述介质上存储有实现上述数据处理方法的程序，所述数据处理方法的程序被处理器执行时实现如上述的数据处理方法的步骤。

25 本申请通过确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；基于各所述数据占比率，确定各第二样本数据的第一参与方预测权重；基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定

所述第一参与方的预设预测模型。本申请通过确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率，确定各第二样本数据的第一参与方预测权重，以联邦确定所述第一参与方的预设预测模型，即本申请通过加入数据占比率的方式，
5 确定各不同分布的第二样本数据中，针对第一参与方具有实际有效训练功能的预测权重，以联邦得到第一参与方的预设预测模型，与相关技术中各个参与方具有相同的训练比重或者预测比重，导致某一参与方数据中部分无效的训练数据，占用训练比重或者预测比重相比，本申请基于各个参与方实际有效的（占用训练比重或者预测比重的）训练数据联合训练联邦模型，进而，
10 避免无效数据所造成的效率干扰，提升联邦训练效率，且由于训练数据是有效的，因而提升联邦模型的模型准确度，以提升模型的预测性能。

附图说明

此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本申请的实施例，并与说明书一起用于解释本申请的原理。

15 为了更清楚地说明本申请实施例或相关技术中的技术方案，下面将对实施例或相关技术描述中所需要使用的附图作简单地介绍，显而易见地，对于本领域普通技术人员而言，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图1为本申请数据处理方法第一实施例的流程示意图；

20 图2为本申请数据处理方法第一实施例中对待识别语音数据进行语音识别得到所述待识别语音数据的各候选结果的步骤细化流程示意图；

图3为本申请实施例方案涉及的硬件运行环境的设备结构示意图；

图4为本申请数据处理方法中的第一场景示意图；

图5为本申请数据处理方法中的第二场景示意图；

25 图6为本申请数据处理方法中的第三场景示意图。

本申请目的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

具体实施方式

应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。
30

本申请实施例提供一种数据处理方法，在本申请数据处理方法的第一实施例中，参照图 1，应用于第一参与方，所述第一参与方与第二参与方进行通信连接，所述第二参与方为多个，所述数据处理方法包括：

步骤 S10，确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

步骤 S20，基于各所述数据占比率，确定各第二样本数据的第一参与方预测权重；

步骤 S30，基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定所述第一参与方的预设预测模型。

具体步骤如下：

步骤 S10，确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

整体地，所述预设预测模型是基于多个带有第一参与方预测权重的第二参与方数据（第二样本数据）以及第一参与方的第一样本数据，通过联邦学习，对第一预设待训练预测模型进行迭代训练，获得的第一目标模型；

在本实施例中，在得到预设预测模型后，获取待处理数据，将所述待处理数据输入至预设预测模型中，其中，预设预测模型是已经训练好的，能够对待处理数据进行准确预测的模型，例如，若待处理数据中包含有用户的年龄，体重，血糖含量，胰岛素水平数据，则将待处理数据输入至预设预测模型中后，可以确定输出用户是否具有糖尿病的结果。

需要说明的是，在本实施例中，联邦学习包括横向联邦学习或者纵向联邦学习，联邦流程包括横向联邦流程以及纵向联邦流程，在本实施例中以横向联邦学习应用为例进行具体说明。

需要说明的是，第一参与方中包括所有标注（有标签）数据与无标注数据的数据并集， i 为 P_i 的编号，对于确定的参与方（包括第一参与方与第二参与方）而言， X 对 Y 服从相同的条件分布，即

$P(Y|X_i) = P(Y|X_j), \forall i, j \in \{1, \dots, n\}, i \neq j$ ，而来自不同的参与方的数据边缘分布不同。具体地，例如，医院积累的数据往往来自患病人群，养老院积累的数据为老年人群，身体疾病状况差异较大，而体检中心的机器学习任务多数

针对青壮年在职成人，现有医院在联合养老院以及体检中心等共同训练机器

模型时，医院，养老院以及体检中心等参与方具有相同的训练比重或者预测比重，即默认医院，养老院以及体检中心等参与方的数据特征与医院数据的特征相同，而实际上不同参与方数据是异质的，基于相同的训练比重或者预测比重致使医院机器模型的训练数据的有效性低下，致使联邦学习训练得到
5 的模型的预测性能低下。

在本实施例中，可以通过确定每个第二样本数据中，与第一参与方数据特征一致的目标数据，以进一步确定目标数据在所述第二样本数据中的数据占比率，或者直接确定目标数据在所述第二样本数据中的数据占比率，其中，第一参与方基于本地的训练数据，训练对应本地的模型外，还确定其他第二
10 参与方的第二样本数据中的目标数据，以确定目标（有效数据）在所述第二样本数据中的数据占比率，以提升该第一参与方本地模型的训练有效性，具体地，例如，多个第二参与方如参与方 A 和参与方 C，共同训练参与方 B（第一参与方）的模型，为了确保用户的隐私，且由于初始模型是相同的，在本实施例中，参与方 A 和参与方 C 共同训练参与方 B 的模型的方式为：选取参
15 与方 A 中对参与方 B 有效的目标数据，得到参与方 A 中对参与方 B 有效的数据占比率，即来自参与方 A 的有标注（预设特征标签）数据被判定为参与方 B（或者具有第一参与方 B 数据特征）的有标注数据的概率，进而求得来自参与方 A 的有标注数据用于训练参与方 B 的模型过程中所对应的第一权重，选取参与方 C 中对参与方 B 有效的目标数据，以确定来自参与方 C 的有标注数
20 据被判定为参与方 B 的有标注数据（或者具有第一参与方 B 数据特征）的概率，求得来自参与方 C 的第二样本数据用于训练参与方 B 的模型过程中所对应的数据占比率，基于该参与方 A 的数据占比率以及参与方 C 的数据占比率，可以联合计算得到第一参与方的需要聚合更新的模型参数，例如，分别获取参与方 A 自身标注数据乘以 A 对参与方 B 的数据占比率，参与方 C 自身标注
25 数据乘以对参与方 B 的数据占比率，进而通过预设计算公式可以联合计算得到第一参与方 B 的需要聚合更新的模型参数，第二参与方 A 的需要聚合更新的模型参数，第二参与方 C 的需要聚合更新的模型参数，具体如图 4，图 5，图 6 所示。

所述确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率的步骤，包括：
30

步骤 S11，接收每个第二参与方发送的数据占比率，其中，每个第二参与方将对应第二样本数据输入至相应预设领域分类模型中后，对所述第二样本数据进行预测处理，得到第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

5 其中，所述预设领域分类模型是基于具有预设参与方来源标签的训练样本数据，通过执行第一预设联邦流程，对第一预设待训练预测模型进行迭代训练后，获得的预测各参与方数据来源占比率的第一目标模型。

10 在本实施例中，是通过预设领域分类模型确定来自第二参与方的有标注数据被判定为第一参与方的有标注数据（与第一参与方数据特征一致）的概率即数据占比率。

整体地，预设领域分类模型是已经通过预设 n 个参与方的训练数据进行横向联邦得到的，其中，预设领域分类器中的输出层经过预设 Sigmoid 激活函数处理后，输出一个维度为 n 的向量，每一个维度输出值在 $(0,1)$ 区间，输出第 i 个维度的输出值用于表示输入数据为第 i 个参与方训练的概率即数据占比率。

15 在本实施例中，所述预设领域分类模型是已经训练好的，即是基于具有预设参与方来源标签的训练样本数据，通过执行第一预设联邦流程，对第一预设待训练预测模型进行迭代训练后，获得的预测各参与方数据来源占比率的第一目标模型，因而，每个第二参与方将对应第二样本数据输入至相应预设领域分类模型中后，对所述第二样本数据进行预测处理，得到第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率。

20 具体地，在本实施例中，基于具有预设参与方来源标签的参与方数据，通过执行第一预设联邦流程，对第一预设待训练预测模型进行迭代训练，得到预测参与方数据来源数据占比率的第一目标模型，例如，第一参与方发起第一联邦学习任务，第一参与方将第一联邦学习任务的第一预设待训练预测模型发送给各个第二参与方，第一参与方基于本地的训练数据，对第一预设待训练预测模型进行训练，得到第一参与方训练后的模型参数，在得到第一参与方训练后的模型参数后，执行第一预设横向联邦流程，获取得到第一联合模型参数（包括其他第二参与方的部分联合模型参数），在得到第一联合模型参数后，持续进行第一横向联邦流程，以供第一参与方得到第二目标联

25

30

合模型参数，以基于第二目标联合模型参数继续执行第一横向两帮流程，以最终得到第一目标模型，在得到第一目标模型的过程中，是学习到了各个参与方数据的隐含特征的，因而，在预设领域分类模型中设置预设 Sigmoid 激活函数，可以得到并输出第*i*个维度的输出值用于表示输入数据为第*i*个参与方的概率或者数据占比率。

具体地，例如，多个第二参与方为参与方 A 和参与方 C，第一参与方为参与方 B，那么对于参与方 B 而言，需要经过预设领域分类模型获取参与方 A 中第二样本数据用于训练参与方 B 模型的数据占比率，并需要经过预设领域分类模型获取参与方 C 第二样本数据中用于训练参与方 B 模型的数据占比率，对于参与方 A 而言，需要经过预设领域分类模型获取参与方 B 中第一样本数据用于训练参与方 A 模型的数据占比率，并需要经过预设领域分类模型获取参与方 C 第二样本数据中用于训练参与方 A 模型的数据占比率，对于参与方 C 而言，需要经过预设领域分类模型获取参与方 A 第二样本数据中用于训练参与方 B 模型的数据占比率，并需要经过预设领域分类模型获取参与方 C 第二样本数据中用于训练参与方 B 模型的数据占比率，也即，对于每一个参与方而言，将本地所有标注数据输入至预设领域分类模型中，求得第*i*个输出值 $P(x_{ik} \in_j) = f_{\theta}(x_k)_j, \forall x_k \in X_i, \forall i, j \in \{1, \dots, n\}, i \neq j$ ，表示来自参与方 P_i 的有标注样本 x_{ik} 被判定为参与方 P_j 的有标注数据的概率。

步骤 S20，基于各所述数据占比率，确定各第二样本数据的第一参与方预测权重；

在得到数据占比率即概率后，根据预设领域分类模型中的如下计算公式计算参与方 P_i 的有标注样本 x_{ik} 用于训练参与方 P_j 的模型过程中所对应的预测

$$\text{权重 } w_{ikj} = \frac{(N - N_j)P(x_{ik} \in X_j)}{N_j(1 - P(x_{ik} \in X_j))}。$$

步骤 S30，基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定所述第一参与方的预设预测模型。

基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习，具体地，通过横向联邦学习确定所述第一参与方的预设预测模型。

所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，确定所述第一参与方的预设预测模型的步骤之后，所述包括：

步骤 S40，获取待处理数据，将所述待处理数据输入至所述预设预测模型中；

步骤 S50，基于所述预设预测模型对所述待处理数据进行预测处理，得到目标预测结果。

5 在本实施例中，在得到预设预测模型后，获取待处理数据，将所述待处理数据输入至所述预设预测模型中，得到目标预测结果。

本申请通过确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；基于各所述数据占比率，确定各第二样本数据的第一参与方预测权重；基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定
10 所述第一参与方的预设预测模型。本申请通过确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率，确定各第二样本数据的第一参与方预测权重，以联邦确定所述第一参与方的预设预测模型，即本申请通过加入数据占比率的方式，
15 确定各不同分布的第二样本数据中，针对第一参与方具有实际有效训练功能的预测权重，以联邦得到第一参与方的预设预测模型，与相关技术中各个参与方具有相同的训练比重或者预测比重，导致某一参与方数据中部分无效的训练数据，占用训练比重或者预测比重相比，本申请基于各个参与方实际有效的（占用训练比重或者预测比重的）训练数据联合训练联邦模型，进而，
20 避免无效数据所造成的效率干扰，提升联邦训练效率，且由于训练数据是有效的，因而提升联邦模型的模型准确度，以提升模型的预测性能。

本申请实施例提供一种数据处理方法，在本申请数据处理方法的另一实施例中，参照图 2，第一参与方中的样本数据为第一样本数据，所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习
25 确定所述第一参与方的预设预测模型的步骤，包括：

步骤 S31，基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型；

步骤 S32，将所述第二目标模型设置为所述第一参与方的预设预测模型。

30 需要说明的是，在本实施例中，基于所述第一参与方预测权重，以及所

述第一参与方的第一样本数据，在中间方或者是在某一参与方通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型，将所述第二目标模型设置为所述第一参与方的预设预测模型。

具体地，在本实施例中，存在协调方，该协调方分别与第一参与方和第二参与方进行通信，在联邦模型训练前，可以是协调方（coordinator）发起第二联邦学习任务，协调方将第二联邦学习任务的初始模型（第二预设待训练预测模型）发送给各个参与方，对于第一参与方而言，第一参与方基于第二预设待训练预测模型，第一参与方本地的训练数据对第一参与方中的初始模型进行训练，得到第一参与方训练后的模型参数，在得到第一参与方训练后的模型参数后，执行第二预设横向联邦流程（其中，各个第二参与方基于初始模型，各个第二参与方得到的第一参与方预测权重的目标数据，对第二预设待训练预测模型进行训练，得到每个第二参与方训练后的模型参数），具体地，基于协调方联合各个第二参与方的训练后的模型参数（加密的）得到的第一联合模型参数，第一参与方继续进行训练并执行第二预设横向联邦流程直至达到预设结束条件，得到第二目标模型。

其中，执行第一预设横向联邦流程的过程具体可以为：对所述第二预设待训练预测模型进行迭代训练，判断迭代训练的所述第二预设待训练预测模型是否达到预设替换更新条件，如是否迭代训练 500 次，若所述第二预设待训练预测模型达到预设替换更新条件，则通过执行所述第二预设横向联邦流程，对训练更新的所述第二预设待训练预测模型的模型变量进行替换更新，获得替换更新的所述第二预设待训练预测模型。

将训练更新的所述第二预设待训练预测模型的模型变量加密发送至与所述第一参与方关联的中间方，以确保安全性，以供所述中间方对多个其他所述第一参与方发送的模型变量进行聚合处理，获得聚合模型变量，并将所述模型变量加密反馈至各所述第一参与方，第一参与方接收所述中间方反馈的所述聚合模型变量，并将训练更新的所述第二预设待训练预测模型的模型变量替换更新为所述聚合模型变量，获得替换更新的所述第二预设待训练预测模型。

持续对替换更新的所述第二预设待训练预测模型进行迭代训练和替换更新，直至所述第二预设待训练模型满足预设训练完成条件，预设训练完成条

件可以是训练次数达到确定数据，如是 1 万次或者是 5000 次，或者预设训练完成条件可以是对应损失函数收敛。以得到第二目标模型。

另外，在本实施例中，还可以是第一参与方发起第二联邦学习任务，第一参与方将第二联邦学习任务的初始模型发送给各个第二参与方，第一参与方基于本地的第一样本数据，分别对第一参与方中的初始模型进行训练（其中，各个第二参与方基于初始模型，以及目标数据，对初始模型进行训练，得到每个第二参与方训练后的模型参数），得到第一参与方训练后的模型参数，在得到第一参与方训练后的模型参数后，执行第一预设横向联邦流程，具体地，基于第一参与方联合各个其他第二参与方的训练后的模型参数得到的第一联合模型参数，继续进行训练并执行第二预设横向联邦流程直至达到预设结束条件，得到第二目标模型。

在本实施例中，通过基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型；将所述第二目标模型设置为所述第一参与方的预设预测模型。本实施例中实现准确得到预设预测模型。

本申请实施例提供一种数据处理方法，在本申请数据处理方法的另一实施例中，所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型的步骤，包括：

步骤 B1，接收第二参与方发送的第二模型初始梯度，其中，所述第二模型初始梯度是各第二参与方在执行第二预设联邦流程过程中，基于对应第二样本数据确定的；

步骤 B2 基于所述第二模型初始梯度以及对应所述第一参与方预测权重，获取各第二参与方的第二模型更新梯度；

步骤 B3，基于所述第二模型更新梯度，以及第一参与方在执行第二预设联邦流程中，基于所述第一样本数据确定的第一模型更新梯度，确定替换更新模型参数；

步骤 B4，基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获得第二目标模型。

在本实施例中，模型参数可以通过梯度得到的，首先接收第二参与方发送的第二模型初始梯度，其中，所述第二模型初始梯度是各第二参与方在执行第二预设联邦流程过程中，基于对应第二样本数据确定的，在得到第二模型初始梯度后，基于所述第二模型初始梯度以及对应所述第一参与方预测权重，获取各第二参与方的第二模型更新梯度，具体地，上标 lb 是标注的意思，每一个参与方将本地所有标注数据（第二样本数据） $\{X_i^{lb}\}_{i=1}^n$ 输入领域分类模型 θ 中，求得第 i 个输出值

$P(x_{ik} \in j) = f_{\theta}(x_k)_j, \forall x_k \in X_i, \forall i, j \in \{1, \dots, n\}, i \neq j$ ，表示来自参与方 P_i 的有标注样本 x_{ik} 被判定为参与方 P_j 的有标注数据集 X_j^{lb} 的概率，根据领域分类模型中如下计算公式计算参与方 P_i 的有标注样本 x_{ik} 用于训练参与方 P_j 的模型过

程中所对应的权重 $w_{ikj} = \frac{(N - N_j)P(x_{ik} \in X_j)}{N_j(1 - P(x_{ik} \in X_j))}$ 即第一参与方预测权重，其中 N 是所有参与方有标注数据的总量， N_j 是 P_j 的有标注训练数据量，得到第一参与方预测权重后，乘以相应的第二模型更新梯度，得到第二模型更新梯度。

基于所述第二模型更新梯度，以及第一参与方在执行第二预设联邦流程中，基于所述第一样本数据确定的第一模型更新梯度，确定替换更新模型参数，基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新如迭代更新 500 次，以获得第二目标模型。

所述第一参与方与第二参与方通过预设中间方通信连接；

所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型的步骤，包括：

步骤 C1，在执行第二预设联邦流程中，基于所述第一样本数据确定第一模型更新梯度，并将所述第一模型更新梯度发送给预设中间方；

步骤 C2，接收预设中间方发送的基于第二模型初始梯度，对应所述第一参与方预测权重以所述及第一模型更新梯度，确定的替换更新模型参数；

其中，第二参与方基于所述第二样本数据，确定第二模型初始梯度，并将第二模型初始梯度加密发送给所述预设中间方；

步骤 C3，基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获得第二目标模型。

在本实施例中，第一参与方以及第二参与方可以通过中间方进行梯度

的聚合，以确定的替换更新模型参数，以实现最终得到目标模型。

所述接收预设中间方发送的基于第二模型初始梯度，对应所述第一参与方预测权重以及第一模型更新梯度，确定的替换更新模型参数的步骤，包括：

5 步骤 D1，接收预设中间方加密发送的基于第二模型初始梯度，对应所述第一参与方预测权重以及第一模型更新梯度，确定的替换更新模型参数。

在本实施例中，第一参与方以及第二参与方可以通过中间方进行梯度的聚合，且在聚合后，中间方加密反馈的替换更新模型参数，以提升横向联邦过程中的安全性。

在本实施例中，通过接收第二参与方发送的第二模型初始梯度，其中，
10 所述第二模型初始梯度是各第二参与方在执行第二预设联邦流程过程中，基于对应第二样本数据确定的；基于所述第二模型初始梯度以及对应所述第一参与方预测权重，获取各第二参与方的第二模型更新梯度；基于所述第二模型更新梯度，以及第一参与方在执行第二预设联邦流程中，基于所述第一样本数据确定的第一模型更新梯度，确定替换更新模型参数；基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获
15 得第二目标模型。在本实施例中，实现准确得到第二目标模型。

参照图 3，图 3 是本申请实施例方案涉及的硬件运行环境的设备结构示意图。

如图 3 所示，该数据处理设备可以包括：处理器 1001，例如 CPU，存储
20 器 1005，通信总线 1002。其中，通信总线 1002 用于实现处理器 1001 和存储器 1005 之间的连接通信。存储器 1005 可以是高速 RAM 存储器，也可以是稳定的存储器（non-volatile memory），例如磁盘存储器。存储器 1005 可选的还可以是独立于前述处理器 1001 的存储设备。

可选地，该数据处理设备还可以包括矩形用户接口、网络接口、摄像头、
25 RF（Radio Frequency，射频）电路，传感器、音频电路、WiFi 模块等等。矩形用户接口可以包括显示屏（Display）、输入子模块比如键盘（Keyboard），可选矩形用户接口还可以包括标准的有线接口、无线接口。网络接口可选的可以包括标准的有线接口、无线接口（如 WI-FI 接口）。

本领域技术人员可以理解，图 3 中示出的数据处理设备结构并不构成对数
30 据处理设备的限定，可以包括比图示更多或更少的部件，或者组合某些部件，

或者不同的部件布置。

如图3所示，作为一种计算机介质的存储器1005中可以包括操作系统、网络通信模块以及数据处理程序。操作系统是管理和控制数据处理设备硬件和软件资源的程序，支持数据处理程序以及其它软件和/或程序的运行。网络通信模块用于实现存储器1005内部各组件之间的通信，以及与数据处理系统中其它硬件和软件之间通信。

在图3所示的数据处理设备中，处理器1001用于执行存储器1005中存储的数据处理程序，实现上述任一项所述的数据处理方法的步骤。

本申请数据处理设备具体实施方式与上述数据处理方法各实施例基本相同，在此不再赘述。

本申请还提供一种数据处理装置，应用于第一参与方，所述第一参与方与第二参与方进行通信连接，所述第二参与方为多个，所述数据处理装置包括：

第一确定模块，用于确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

第二确定模块，用于基于各所述数据占比率，确定各第二样本数据的第一参与方预测权重；

第三确定模块，用于基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定所述第一参与方的预设预测模型。

可选地，所述第一确定模块包括：

接收单元，用于接收每个第二参与方发送的数据占比率，其中，每个第二参与方将对应第二样本数据输入至相应预设领域分类模型中后，对所述第二样本数据进行预测处理，得到第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

其中，所述预设领域分类模型是基于具有预设参与方来源标签的训练样本数据，通过执行第一预设联邦流程，对第一预设待训练预测模型进行迭代训练后，获得的预测各参与方数据来源占比率的第一目标模型。

可选地，第一参与方中的样本数据为第一样本数据，所述第三确定模块包括：

第一执行单元，用于基于所述第一参与方预测权重，以及所述第一参与

方的第一样本数据，通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型；

设置单元，用于将所述第二目标模型设置为所述第一参与方的预设预测模型。

5 可选地，所述第一执行单元包括：

第一接收子单元，用于接收第二参与方发送的第二模型初始梯度，其中，所述第二模型初始梯度是各第二参与方在执行第二预设联邦流程过程中，基于对应第二样本数据确定的；

10 获取子单元，用于基于所述第二模型初始梯度以及对应所述第一参与方预测权重，获取各第二参与方的第二模型更新梯度；

确定子单元，用于基于所述第二模型更新梯度，以及第一参与方在执行第二预设联邦流程中，基于所述第一样本数据确定的第一模型更新梯度，确定替换更新模型参数；

15 第一更新子单元，用于基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获得第二目标模型。

可选地，所述第一参与方与第二参与方通过预设中间方通信连接；

所述第一执行单元还包括：

发送子单元，用于在执行第二预设联邦流程中，基于所述第一样本数据确定第一模型更新梯度，并将所述第一模型更新梯度发送给预设中间方；

20 第二接收子单元，用于接收预设中间方发送的基于第二模型初始梯度，对应所述第一参与方预测权重以所述及第一模型更新梯度，确定的替换更新模型参数；

其中，第二参与方基于所述第二样本数据，确定第二模型初始梯度，并将第二模型初始梯度加密发送给所述预设中间方；

25 第二更新子单元，用于基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获得第二目标模型。

可选地，所述第二接收子单元用于实现：

接收预设中间方加密发送的基于第二模型初始梯度，对应所述第一参与方预测权重以及第一模型更新梯度，确定的替换更新模型参数。

30 可选地，所述数据处理装置还包括：

第一获取模块，用于获取待处理数据，将所述待处理数据输入至所述预设预测模型中；

第二获取模块，用于基于所述预设预测模型对所述待处理数据进行预测处理，得到目标预测结果。

5 本申请数据处理装置的具体实施方式与上述数据处理方法各实施例基本相同，在此不再赘述。

本申请实施例提供了一种介质，且所述介质存储有一个或者一个以上程序，所述一个或者一个以上程序还可被一个或者一个以上的处理器执行以用于实现上述任一项所述的数据处理方法的步骤。

10 本申请介质具体实施方式与上述数据处理方法各实施例基本相同，在此不再赘述。

以上仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利处理范围内。

权 利 要 求 书

1、一种数据处理方法，其中，应用于第一参与方，所述第一参与方与第二参与方进行通信连接，所述第二参与方为多个，所述数据处理方法包括：

5 确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

基于各所述数据占比率，确定各第二样本数据的第一参与方预测权重；

基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定所述第一参与方的预设预测模型。

10 2、如权利要求1所述数据处理方法，其中，所述确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率的步骤，包括：

接收每个第二参与方发送的数据占比率，其中，每个第二参与方将对应第二样本数据输入至相应预设领域分类模型中后，对所述第二样本数据进行
15 预测处理，得到第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

其中，所述预设领域分类模型是基于具有预设参与方来源标签的训练样本数据，通过执行第一预设联邦流程，对第一预设待训练预测模型进行迭代训练后，获得的预测各参与方数据来源占比率的第一目标模型。

20 3、如权利要求1所述数据处理方法，其中，第一参与方中的样本数据为第一样本数据，所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定所述第一参与方的预设预测模型的步骤，包括：

基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，
25 通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型；

将所述第二目标模型设置为所述第一参与方的预设预测模型。

4、如权利要求3所述数据处理方法，其中，所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过执行第二预设联邦流程，
30 对第二预设待训练预测模型进行迭代训练，获得第二目标模型的步骤，包括：

接收第二参与方发送的第二模型初始梯度，其中，所述第二模型初始梯度是各第二参与方在执行第二预设联邦流程过程中，基于对应第二样本数据确定的；

5 基于所述第二模型初始梯度以及对应所述第一参与方预测权重，获取各第二参与方的第二模型更新梯度；

基于所述第二模型更新梯度，以及第一参与方在执行第二预设联邦流程中，基于所述第一样本数据确定的第一模型更新梯度，确定替换更新模型参数；

10 基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获得第二目标模型。

5、如权利要求4所述数据处理方法，其中，所述第一参与方与第二参与方通过预设中间方通信连接；

15 所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过执行第二预设联邦流程，对第二预设待训练预测模型进行迭代训练，获得第二目标模型的步骤，包括：

在执行第二预设联邦流程中，基于所述第一样本数据确定第一模型更新梯度，并将所述第一模型更新梯度发送给预设中间方；

接收预设中间方发送的基于第二模型初始梯度，对应所述第一参与方预测权重以所述及第一模型更新梯度，确定的替换更新模型参数；

20 其中，第二参与方基于所述第二样本数据，确定第二模型初始梯度，并将第二模型初始梯度加密发送给所述预设中间方；

基于所述替换更新模型参数，对第二预设待训练预测模型中的模型参数进行迭代更新，以获得第二目标模型。

25 6、如权利要求5所述数据处理方法，其中，所述接收预设中间方发送的基于第二模型初始梯度，对应所述第一参与方预测权重以及第一模型更新梯度，确定的替换更新模型参数的步骤，包括：

接收预设中间方加密发送的基于第二模型初始梯度，对应所述第一参与方预测权重以及第一模型更新梯度，确定的替换更新模型参数。

30 7、如权利要求1所述数据处理方法，其中，所述基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，确定所述第一参与方的预设

预测模型的步骤之后，所述包括：

获取待处理数据，将所述待处理数据输入至所述预设预测模型中；

基于所述预设预测模型对所述待处理数据进行预测处理，得到目标预测结果。

5 8、一种数据处理装置，其中，应用于第一参与方，所述第一参与方与第二参与方进行通信连接，所述第二参与方为多个，所述数据处理装置包括：

第一确定模块，用于确定各所述第二参与方各自的第二样本数据中，与第一参与方数据特征一致的目标数据在所述第二样本数据中的数据占比率；

10 第二确定模块，用于基于各所述数据占比率，确定各第二样本数据的第一参与方预测权重；

第三确定模块，用于基于所述第一参与方预测权重，以及所述第一参与方的第一样本数据，通过联邦学习确定所述第一参与方的预设预测模型。

9、一种数据处理设备，其中，所述数据处理设备包括：存储器、处理器以及存储在存储器上的用于实现所述数据处理方法的程序，

15 所述存储器用于存储实现数据处理方法的程序；

所述处理器用于执行实现所述数据处理方法的程序，以实现如权利要求 1 所述数据处理方法的步骤。

10、一种数据处理设备，其中，所述数据处理设备包括：存储器、处理器以及存储在存储器上的用于实现所述数据处理方法的程序，

20 所述存储器用于存储实现数据处理方法的程序；

所述处理器用于执行实现所述数据处理方法的程序，以实现如权利要求 2 所述数据处理方法的步骤。

11、一种数据处理设备，其中，所述数据处理设备包括：存储器、处理器以及存储在存储器上的用于实现所述数据处理方法的程序，

25 所述存储器用于存储实现数据处理方法的程序；

所述处理器用于执行实现所述数据处理方法的程序，以实现如权利要求 3 所述数据处理方法的步骤。

12、一种数据处理设备，其中，所述数据处理设备包括：存储器、处理器以及存储在存储器上的用于实现所述数据处理方法的程序，

30 所述存储器用于存储实现数据处理方法的程序；

所述处理器用于执行实现所述数据处理方法的程序,以实现如权利要求 4 所述数据处理方法的步骤。

13、一种数据处理设备,其中,所述数据处理设备包括:存储器、处理器以及存储在存储器上的用于实现所述数据处理方法的程序,

5 所述存储器用于存储实现数据处理方法的程序;

所述处理器用于执行实现所述数据处理方法的程序,以实现如权利要求 5 所述数据处理方法的步骤。

14、一种数据处理设备,其中,所述数据处理设备包括:存储器、处理器以及存储在存储器上的用于实现所述数据处理方法的程序,

10 所述存储器用于存储实现数据处理方法的程序;

所述处理器用于执行实现所述数据处理方法的程序,以实现如权利要求 6 所述数据处理方法的步骤。

15、一种介质,其中,所述介质上存储有实现数据处理方法的程序,所述实现数据处理方法的程序被处理器执行以实现如权利要求 1 所述数据处理方法的步骤。

16、一种介质,其中,所述介质上存储有实现数据处理方法的程序,所述实现数据处理方法的程序被处理器执行以实现如权利要求 2 所述数据处理方法的步骤。

17、一种介质,其中,所述介质上存储有实现数据处理方法的程序,所述实现数据处理方法的程序被处理器执行以实现如权利要求 3 所述数据处理方法的步骤。

18、一种介质,其中,所述介质上存储有实现数据处理方法的程序,所述实现数据处理方法的程序被处理器执行以实现如权利要求 4 所述数据处理方法的步骤。

19、一种介质,其中,所述介质上存储有实现数据处理方法的程序,所述实现数据处理方法的程序被处理器执行以实现如权利要求 5 所述数据处理方法的步骤。

20、一种介质,其中,所述介质上存储有实现数据处理方法的程序,所述实现数据处理方法的程序被处理器执行以实现如权利要求 6 所述数据处理方法的步骤。

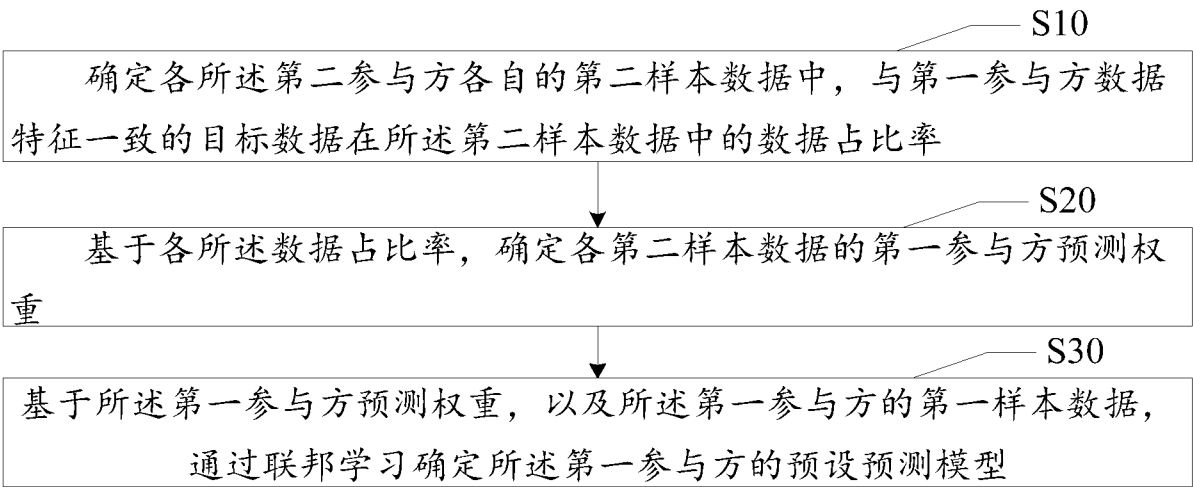


图 1

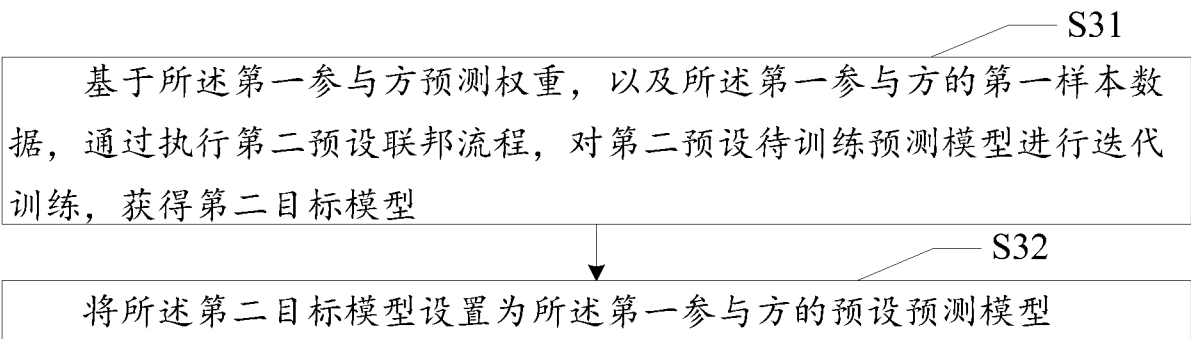


图 2

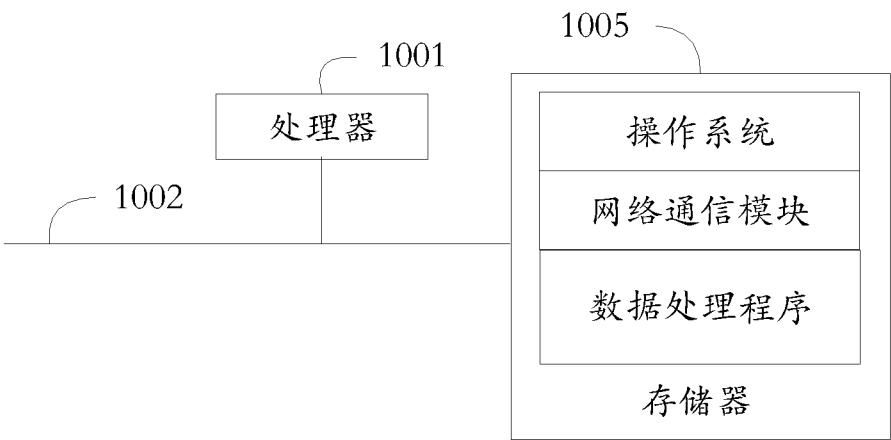


图 3

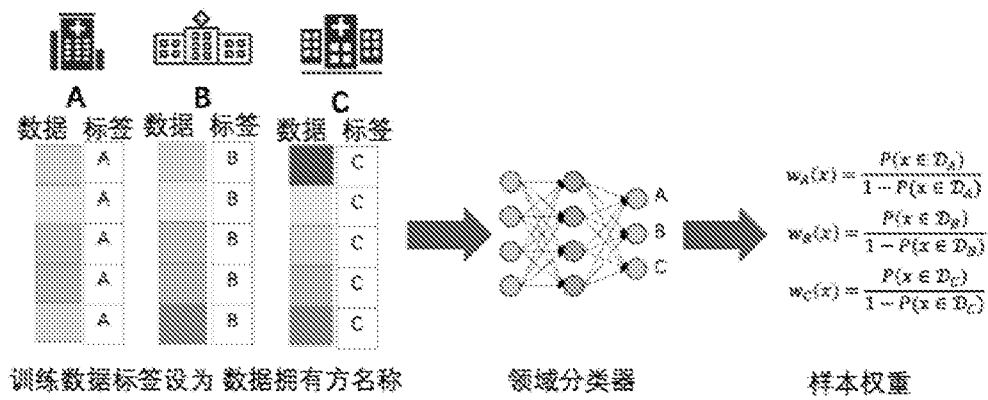


图 4

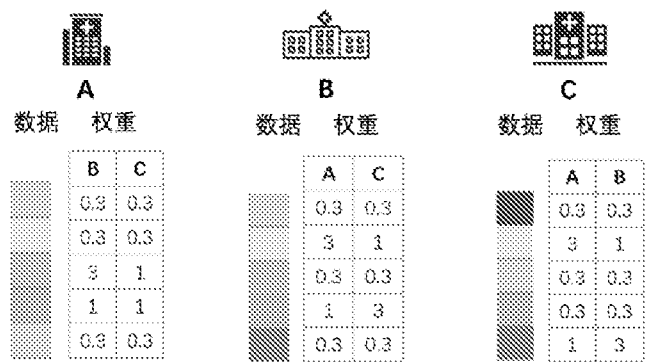


图 5

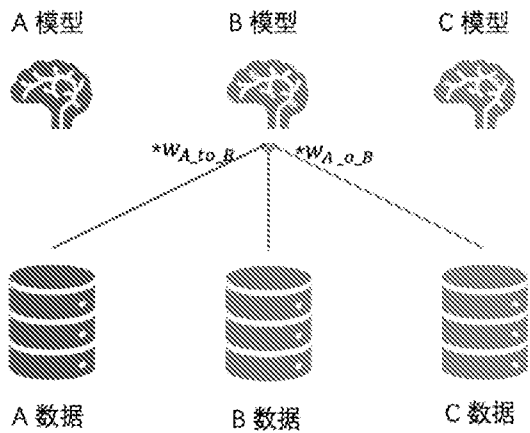


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/094938

A. CLASSIFICATION OF SUBJECT MATTER

G06N 20/00(2019.01)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,G06F 17/-,G06F 16/-

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)

DWPI, CNABS, SIPOABS, CNTXT, USTXT, CNKI: 微众, 联邦, 学习, 模型, 比例, 比率, 权重, 预测, 分类, 样本, 迭代, 更新, federal, federated, learning, models, proportion, ratio, weight, prediction, classification, sample, iteration, update

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 111898768 A (SHENZHEN QIANHAI WEBANK CO., LTD.) 06 November 2020 (2020-11-06) description paragraphs 71-163	1-20
A	CN 110490335 A (SHENZHEN QIANHAI WEBANK CO., LTD.) 22 November 2019 (2019-11-22) entire document	1-20
A	CN 111091200 A (SHENZHEN QIANHAI WEBANK CO., LTD.) 01 May 2020 (2020-05-01) entire document	1-20
A	CN 111355739 A (SHENZHEN QIANHAI WEBANK CO., LTD.) 30 June 2020 (2020-06-30) entire document	1-20
A	US 2020027019 A1 (LG ELECTRONICS INC.) 23 January 2020 (2020-01-23) entire document	1-20



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

20 August 2021

Date of mailing of the international search report

26 August 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/CN2021/094938

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111898768	A	06 November 2020	None			
CN	110490335	A	22 November 2019	None			
CN	111091200	A	01 May 2020	CN	111091200	B	19 March 2021
CN	111355739	A	30 June 2020	None			
US	2020027019	A1	23 January 2020	KR	20190103090	A	04 September 2019

国际检索报告

国际申请号

PCT/CN2021/094938

A. 主题的分类

G06N 20/00 (2019.01) 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, G06F 17/-, G06F 16/-

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))

DWPI, CNABS, SIPOABS, CNTXT, USTXT, CNKI: 微众, 联邦, 学习, 模型, 比例, 比率, 权重, 预测, 分类, 样本, 迭代, 更新, federal, federated, learning, models, proportion, ratio, weight, prediction, classification, sample, iteration, update

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
PX	CN 111898768 A (深圳前海微众银行股份有限公司) 2020年 11月 6日 (2020 - 11 - 06) 说明书第71-163段	1-20
A	CN 110490335 A (深圳前海微众银行股份有限公司) 2019年 11月 22日 (2019 - 11 - 22) 全文	1-20
A	CN 111091200 A (深圳前海微众银行股份有限公司) 2020年 5月 1日 (2020 - 05 - 01) 全文	1-20
A	CN 111355739 A (深圳前海微众银行股份有限公司) 2020年 6月 30日 (2020 - 06 - 30) 全文	1-20
A	US 2020027019 A1 (LG ELECTRONICS INC) 2020年 1月 23日 (2020 - 01 - 23) 全文	1-20

☐ 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 8月 20日

国际检索报告邮寄日期

2021年 8月 26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

田竞

电话号码 86-10-62411749

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/094938

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	111898768	A	2020年 11月 6日	无	
CN	110490335	A	2019年 11月 22日	无	
CN	111091200	A	2020年 5月 1日	CN 111091200 B	2021年 3月 19日
CN	111355739	A	2020年 6月 30日	无	
US	2020027019	A1	2020年 1月 23日	KR 20190103090 A	2019年 9月 4日