(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
2021 年 6 月 17 日 (17.06.2021)

WIPO | PCT

(10) 国际公布号
WO 2021/114616 A1

(51) 国际专利分类号:
G06N 3/04 (2006.01) *G06N 20/00* (2019.01)
G06N 3/08 (2006.01)

(21) 国际申请号: PCT/CN2020/098884

(22) 国际申请日: 2020 年 6 月 29 日 (29.06.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

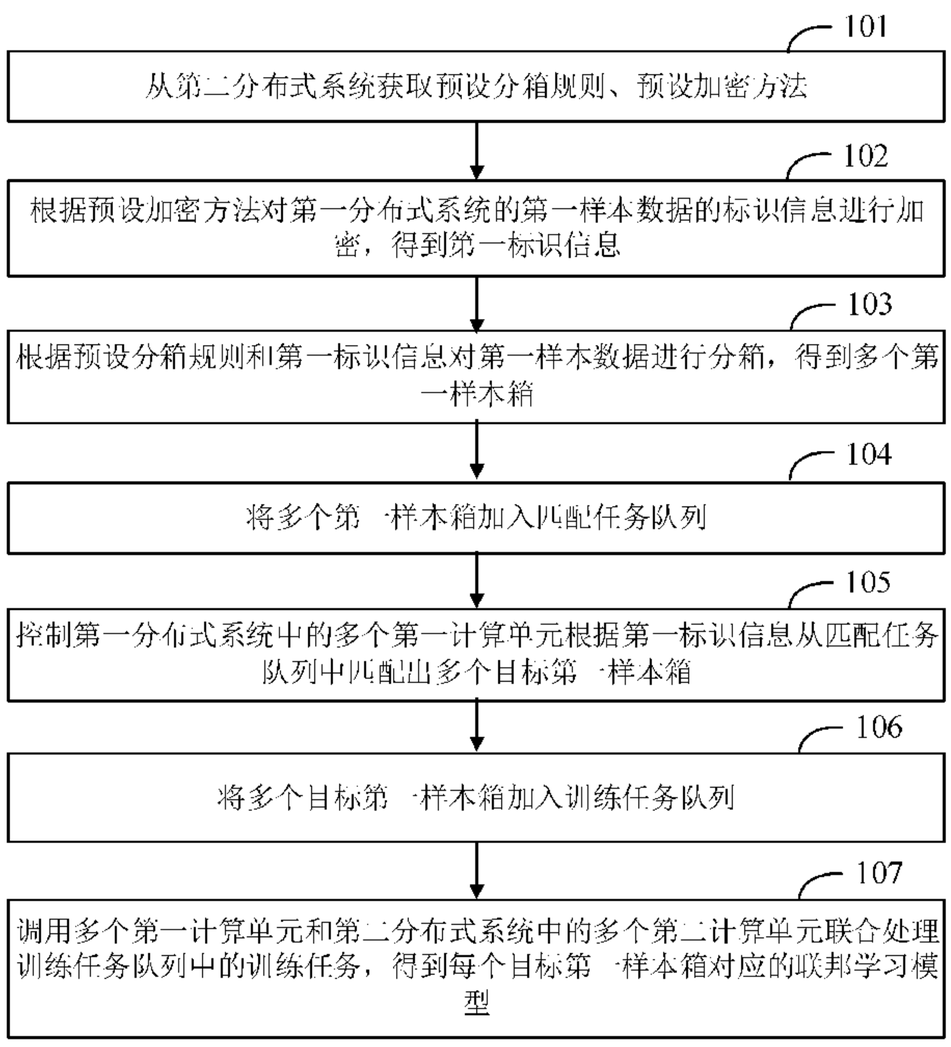
(30) 优先权:
202010409022.9 2020年5月14日 (14.05.2020) CN

(71) 申请人: **平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.)** [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(72) 发明人: **周学立 (ZHOU, Xueli)**; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 **陈玉 (CHEN, Yu)**; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 **杜均 (DU, Jun)**; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 **孙召元 (SUN, Zhaoyuan)**; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 **朱恩东 (ZHU, Endong)**; 中国广东省深圳市福田

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

(54) **发明名称:** 联邦学习模型训练方法及相关设备



101 ACQUIRE A PRESET BINNING RULE AND A PRESET ENCRYPTION METHOD FROM A SECOND DISTRIBUTED SYSTEM

102 ON THE BASIS OF THE PRESET ENCRYPTION METHOD, ENCRYPT IDENTIFICATION INFORMATION OF FIRST SAMPLE DATA OF A FIRST DISTRIBUTED SYSTEM TO OBTAIN FIRST IDENTIFICATION INFORMATION

103 ON THE BASIS OF THE PRESET BINNING RULE AND THE FIRST IDENTIFICATION INFORMATION, BIN THE FIRST SAMPLE DATA TO OBTAIN A PLURALITY OF FIRST SAMPLE BINS

104 ADD THE PLURALITY OF FIRST SAMPLE BINS TO A MATCHING TASK QUEUE

105 CONTROL A PLURALITY OF FIRST COMPUTATION UNITS IN THE FIRST DISTRIBUTED SYSTEM TO MATCH A PLURALITY OF TARGET FIRST SAMPLE BINS FROM THE MATCHING TASK QUEUE ON THE BASIS OF THE FIRST IDENTIFICATION INFORMATION

106 ADD THE PLURALITY OF TARGET FIRST SAMPLE BINS TO A TRAINING TASK QUEUE

107 INVOKE THE PLURALITY OF FIRST COMPUTATION UNITS AND A PLURALITY OF SECOND COMPUTATION UNITS IN THE SECOND DISTRIBUTED SYSTEM TO JOINTLY PROCESS THE TRAINING TASKS IN THE TRAINING TASK QUEUE TO OBTAIN A FEDERATED LEARNING MODEL CORRESPONDING TO EACH TARGET FIRST SAMPLE BIN

图 1

(57) Abstract: A federated learning model training method and a related device, relating to artificial intelligence. The federated learning model training method comprises: acquiring a preset binning rule and a preset encryption method from a second distributed system (101); on the basis of the preset encryption method, encrypting identification information of first sample data of a first distributed system to obtain first identification information (102); on the basis of the preset binning rule and the first identification information, binning the first sample data to obtain a plurality of first sample bins (103); adding the plurality of first sample bins to a matching task

[见续页]

WO 2021/114616 A1

区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市赛恩倍吉知识产权代理有限公司(SHENZHEN SCIENBIZIP INTELLECTUAL PROPERTY AGENCY CO.,LTD.); 中国广东省深圳市龙华新区龙观东路83号荣群大厦9楼, Guangdong 518109 (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, 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))。
— 在修改权利要求的期限届满之前进行，在收到该修改后将重新公布(细则48.2(h))。
— 根据申请人的请求，在条约第21条(2)(a)所规定的期限届满之前进行。

queue (104); controlling a plurality of first computation units in the first distributed system to match a plurality of target first sample bins from the matching task queue on the basis of the first identification information (105); adding the plurality of target first sample bins to a training task queue (106); and invoking the plurality of first computation units and a plurality of second computation units in the second distributed system to jointly process the training tasks in the training task queue to obtain a federated learning model corresponding to each target first sample bin (107). The efficiency of federated learning model training is thereby increased.

(57) 摘要: 一种联邦学习模型训练方法及相关设备，涉及人工智能。所述联邦学习模型训练方法包括：从第二分布式系统获取预设分箱规则、预设加密方法（101）；根据预设加密方法对第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息（102）；根据预设分箱规则和第一标识信息对第一样本数据进行分箱，得到多个第一样本箱（103）；将所述多个第一样本箱加入匹配任务队列（104）；控制第一分布式系统中的多个第一计算单元根据第一标识信息从匹配任务队列中匹配出多个目标第一样本箱（105）；将多个目标第一样本箱加入训练任务队列（106）；调用多个第一计算单元和第二分布式系统中的多个第二计算单元联合处理训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型（107）。提升了训练联邦学习模型的效率。

联邦学习模型训练方法及相关设备

本申请要求于 2020 年 05 月 14 日提交中国专利局，申请号为 202010409022.9 申请名称为“联邦学习模型训练方法及相关设备”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技术领域，具体涉及一种联邦学习模型训练方法、装置、计算机设备及计算机可读存储介质。

背景技术

随着人工智能技术的发展，通过联合不同的参与者(participant，或者 party，也称为数据拥有者(data owner)、或者客户(client))进行机器学习建模成为发展趋势，即联邦学习。

在联邦学习中，发明人意识到目前训练联邦学习模型的效率较低。如何提升训练联邦学习模型的效率，成为亟待解决的技术问题。

发明内容

鉴于以上内容，有必要提出一种联邦学习模型训练方法、装置、计算机设备及计算机可读存储介质，其可以提高训练联邦学习模型的效率。

本申请的第一方面提供一种联邦学习模型训练方法，所述联邦学习模型训练方法应用于第一分布式系统中，所述联邦学习模型训练方法包括：

从第二分布式系统获取预设分箱规则、预设加密方法；

根据所述预设加密方法对所述第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

将所述多个第一样本箱加入匹配任务队列；

控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

将所述多个目标第一样本箱加入训练任务队列；

调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

本申请的第二方面提供一种联邦学习模型训练装置，所述联邦学习模型训练装置运行于第一分布式系统中，所述联邦学习模型训练装置包括：

获取模块，用于从第二分布式系统获取预设分箱规则、预设加密方法；

加密模块，用于根据所述预设加密方法对所述第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

分箱模块，用于根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

第一加入模块，用于将所述多个第一样本箱加入匹配任务队列；

匹配模块，用于控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

第二加入模块，用于将所述多个目标第一样本箱加入训练任务队列；

处理模块，用于调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

本申请的第三方面提供一种计算机设备，多个所述计算机设备组成第一分布式系统，其中，每个计算机设备包括处理器，所述处理器用于执行存储器中存储的计算机可读指令以实现以下步骤：

从第二分布式系统获取预设分箱规则、预设加密方法；

根据所述预设加密方法对所述第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

将所述多个第一样本箱加入匹配任务队列；

控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

将所述多个目标第一样本箱加入训练任务队列；

调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型；

所述第一分布式系统是数据请求方，所述第二分布式系统是数据提供方，数据请求方需要联合数据提供方进行联邦学习。

本申请的第四方面提供一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机可读指令，所述计算机可读指令被处理器执行时实现以下步骤：

从第二分布式系统获取预设分箱规则、预设加密方法；

根据所述预设加密方法对第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

将所述多个第一样本箱加入匹配任务队列；

控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

将所述多个目标第一样本箱加入训练任务队列；

调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

本申请通过分布式系统的多个计算单元对待处理匹配任务和训练任务进行异步处理，避免待处理任务在单位时间点对计算资源的需求压力，同时提高容错性，提升了训练联邦学习模型的效率。

附图说明

图 1 是本申请实施例提供的联邦学习模型训练方法的流程图。

图 2 是本申请实施例提供的联邦学习模型训练装置的结构图。

图 3 是本申请实施例提供的计算机设备的示意图。

具体实施方式

为了能够更清楚地理解本申请的上述目的、特征和优点，下面结合附图和具体实施例对本申请进行详细描述。需要说明的是，在不冲突的情况下，本申请的实施例及实施例中的特征可以相互组合。

在下面的描述中阐述了很多具体细节以便于充分理解本申请，所描述的实施例仅仅

是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

除非另有定义，本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中在本申请的说明书中所使用的术语只是为了描述具体的实施例的目的，不是旨在于限制本申请。

优选地，本申请的联邦学习模型训练方法应用在一个或者多个计算机设备中。所述计算机设备是一种能够按照事先设定或存储的指令，自动进行数值计算和/或信息处理的设备，其硬件包括但不限于微处理器、专用集成电路(Application Specific Integrated Circuit，ASIC)、可编程门阵列(Field－Programmable Gate Array，FPGA)、数字处理器(Digital Signal Processor，DSP)、嵌入式设备等。

本申请可用于众多通用或专用的计算机系统环境或配置中。例如：个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、置顶盒、可编程的消费电子设备、网络 PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

所述计算机设备可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述计算机设备可以与用户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互。

实施例一

图 1 是本申请实施例一提供的联邦学习模型训练方法的流程图。所述联邦学习模型训练方法应用于第一分布式系统中，所述第一分布式系统由多个计算机设备构成，用于训练联邦学习模型。

所述联邦学习模型训练方法具体包括以下步骤，根据不同的需求，该流程图中步骤的顺序可以改变，某些可以省略。

101，从第二分布式系统获取预设分箱规则、预设加密方法。

所述第一分布式系统可以是数据请求方，所述第二分布式系统可以是数据提供方，数据请求方需要联合数据提供方进行联邦学习，并不需要获取数据提供方的数据。例如，数据请求方为保险公司，数据提供方为金融公司；保险公司需要在不获取金融公司的数据的情况下，与金融公司进行联邦学习，以获得对客户进行可靠性评分的评分模型。

所述第二分布式系统由多个计算机设备构成。

所述第一分布式系统可以使用所述预设分箱规则对样本数据进行分箱。所述预设分箱规则包括：根据数值范围或除数取余结果将样本数据分入不同的样本箱。所述第二分布式系统也是用所述预设分箱规则对样本数据进行分箱，使两个系统中相同的样本数据被分入相对应的样本箱。

所述预设加密方法可以是不可回溯的均匀加密算法。所述预设加密方法可以是哈希加密方法、对称加密方法、非对称加密方法。所述预设加密方法可以用于对样本数据的识别信息或所述第一分布式系统与所述第二分布式系统之间的通信数据进行加密。所述预设加密方法可以使所述第一分布式系统和所述第二分布式系统不获取对方的明文数据。所述第一分布式系统和所述第二分布式系统使用相同的预设加密方法，当所述第一分布式相同中的样本数据的标识信息与所述第二分布式系统中的样本数据的标识信息相同时，

加密后的两个标识信息也相同。

102，根据所述预设加密方法对所述第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息。

所述第一样本数据可以包括多个用户的基本信息和一些特征数据等，所述第一标识信息可以是加密后的用户的身份证号码或手机号。

例如，保险公司中的第一样本数据包括 3 条样本数据，分别为{用户 1 的标识信息，年龄，保险特征数据}、{用户 2 的标识信息，年龄，保险特征数据}、{用户 3 的标识信息，年龄，保险特征数据}。可以根据哈希算法（预设加密方法）对每个用户的识别信息进行加密。

103，根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱。

在一具体实施例中，在所述根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱之前，所述联邦学习模型训练方法还包括：

对所述第一样本数据进行数据清洗、格式转换、特征工程等预处理。

还可以对所述第一样本数据进行数据混淆、数据编码等预处理。

在一具体实施例中，所述根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱包括：

获取所述预设分箱规则中的预设除数；

对每个所述第一标识信息与所述预设除数进行相除并获取余数；

根据所述第一标识信息对应的余数对所述第一样本数据进行分箱，得到多个第一样本箱；

判断每个第一样本箱中的第一样本数量是否大于预设数量阈值；

对第一样本数量大于所述预设数量阈值的第一样本箱进行均匀分割，得到多个子箱。

例如，用 30 个第一标识信息分别除以 6（预设除数），得到各个第一标识信息的余数。将第一标识信息的余数为 1 的第一样本数据分入第 1 个第一样本箱；将第一标识信息的余数为 2 的第一样本数据分入第 2 个第一样本箱；将第一标识信息的余数为 3 的第一样本数据分入第 3 个第一样本箱；将第一标识信息的余数为 4 的第一样本数据分入第 4 个第一样本箱；将第一标识信息的余数为 5 的第一样本数据分入第 5 个第一样本箱；将第一标识信息的余数为 0 的第一样本数据分入第 6 个第一样本箱。如果判断第 2 个第一样本箱中的第一样本数量大于预设数量阈值，将第 2 个第一样本箱均匀分割为 2 个子箱（或其它数量）。

或者，所述根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱包括：

获取所述预设分箱规则中的多个预设取值范围，每个预设取值范围对应一个第一样本箱；

截取每个所述第一标识信息的预设位置的数值，得到每个所述第一标识信息的截取数值；

确定所述第一标识信息的截取数值所属的目标预设取值范围；

根据所述第一标识信息的截取数值所属的目标预设取值范围，将所述第一样本数据分入目标预设取值范围对应的第一样本箱；

判断每个第一样本箱中的第一样本数量是否大于预设数量阈值；

对第一样本数量大于所述预设数量阈值的第一样本箱进行均匀分割，得到多个子箱。

例如，获取预设分箱规则中的 5 个预设取值范围（如 0-1，2-3，4-5，6-7，8-9），5 个预设取值范围分别对应 5 个第一样本箱。截取各个第一标识信息的最后一位数值（如十进制 0-9），得到每个第一标识信息的截取数值。确定每个第一标识信息的截取数值所属的目标预设取值范围，如截取数值 0、1 属于取值范围 0-1（对应第 1 个第一样本箱），

截取数值 2、3 属于取值范围 2-3（对应第 2 个第一样本箱），截取数值 4、5 属于取值范围 4-5（对应第 3 个第一样本箱），截取数值 6、7 属于取值范围 6-7（对应第 4 个第一样本箱），截取数值 8、9 属于取值范围 8-9（对应第 5 个第一样本箱）。将第一标识信息的截取数值为 0 的第一样本数据分入第 1 个第一样本箱；将第一标识信息的截取数值为 1 的第一样本数据分入第 1 个第一样本箱；将第一标识信息的截取数值为 2 的第一样本数据分入第 2 个第一样本箱；以此类推。如果判断第 3 个第一样本箱中的第一样本数量大于预设数量阈值，将第 3 个第一样本箱均匀分割为 2 个子箱（或其它数量）。

通过分箱对需要计算的样本数据进行合理的分配。可以理解为把样本数据均衡分块成多个任务，交给不同的计算单元进行处理。缩小了遍历的计算量和访问压力。

104，将所述多个第一样本箱加入匹配任务队列。

可以将每个第一样本箱打包为待处理的匹配任务，将多个匹配任务加入匹配任务队列。所述第一分布式系统中的处于空闲状态的计算单元可以依次读取并处理所述匹配任务队列中的匹配任务，以达到异步处理匹配任务的目的，提升处理效率。

105，控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱。

可以从所述第一分布式系统获取执行任务需要的多个第一计算单元。对所述多个第一计算单元进行配置。为保证配置的一致性和易操作性，可以使用 docker 向预设代码仓库拉取镜像。

在一具体实施例中，所述控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱包括：

每次控制一个第一计算单元读取所述匹配任务队列中的一个匹配任务，获取所述匹配任务中的第一样本箱；

控制所述第一计算单元从所述第二分布式系统中获取与所述第一样本箱关联的第二标识信息；

控制所述第一计算单元从所述第一样本箱中匹配出与所述第二标识信息一致的第一标识信息对应的样本数据，得到多个目标样本数据；

通过所述第一计算单元，删除所述第一样本箱中除所述多个目标样本数据外的样本数据，得到目标第一样本箱。

在另一实施例中，所述第一样本箱包括多个子箱，所述控制所述第一计算单元从所述第一样本箱中匹配出与所述第二标识信息一致的第一标识信息对应的样本数据包括：

控制所述第一计算单元从所述第一样本箱的多个子箱中匹配出与所述第二标识信息一致的第一标识信息对应的样本数据，得到多个目标样本数据。将第一样本箱分为多个子箱，可以减少所述第一计算单元在短时间内对高运行资源的需求。避免卡顿或宕机，从而提升计算效率。

在另一实施例中，所述控制所述第一计算单元从所述第二分布式系统中获取与所述第一样本箱关联的第二标识信息包括：

为所述第一计算单元及所述第二分布式系统中的与所述第一计算单元关联的第二计算单元建立通信通道；

控制所述第一计算单元通过所述通信通道从所述关联的第二计算单元中获取与所述第一样本箱关联的第二样本箱，从所述关联的第二样本箱中获取第二样本数据的第二标识信息。

所述多个第一样本箱与所述第二分布式系统中的多个第二样本箱一一关联，处理两个关联样本箱的两个计算单元一一关联。例如第一样本箱 A 与第二样本箱 B 关联，第一计算单元 a 处理第一样本箱 A，第二计算单元 b 处理第二样本箱 B，第一计算单元 a 与第二计算单元 b 关联。

所述第一分布式系统中的一个第一计算单元与所述第二分布式系统中的一个第二

计算单元一一关联，两个计算单元分别处理的两个样本箱也一一关联。

所述第二分布式系统根据所述预设加密方法对所述第二分布式系统的第二样本数据的标识信息进行加密，得到所述第二标识信息。所述第二分布式系统根据所述预设分箱规则对所述第二样本数据进行分箱，得到多个第二样本箱。

在一具体实施例中，所述控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱包括：

获取所述多个第一计算单元的第一数量和所述匹配任务队列中的匹配任务的第二数量；

判断所述第一数量是否小于所述第二数量；

当所述第一数量大于或等于所述第二数量时，用所述多个第一计算单元根据所述第一标识信息处理所述匹配任务队列中的全部匹配任务；

当所述第一数量小于所述第二数量时，用所述多个第一计算单元根据所述第一标识信息分批处理所述匹配任务队列中的匹配任务。

106，将所述多个目标第一样本箱加入训练任务队列。

可以将各个目标第一样本箱打包为待处理的训练任务，将各个训练任务加入训练任务队列。所述第一分布式系统中的处于空闲状态的计算单元可以依次读取并处理所述训练任务队列中的训练任务，以达到异步处理训练任务的目的，提升处理效率。

在一具体实施例中，所述将所述多个目标第一样本箱加入训练任务队列包括：

获取每个目标第一样本箱中的第二样本数量；

当所述第二样本数量小于预设第一样本数量阈值时，合并第二样本数量小于所述预设第一样本数量阈值的目标第一样本箱，将所述合并后的第一样本箱加入所述训练任务队列；

当所述第二样本数量大于所述预设第二样本数量阈值时，拆分第二样本数量大于所述预设第二样本数量阈值的目标第一样本箱，将所述拆分后的第一样本箱加入所述训练任务队列；

其中，所述预设第二样本数量阈值大于所述预设第一样本数量阈值，所述合并后的目标第一样本箱的第二样本数量小于或等于所述预设第二样本数量阈值，且大于或等于所述预设第一样本数量阈值，所述拆分后的第一样本箱的第二样本数量小于或等于预设第二样本数量阈值，且大于或等于所述预设第一样本数量阈值。

107，调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

在一具体实施例中，所述调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务包括：

每次控制一个第一计算单元读取所述训练任务队列中的一个训练任务，获取所述训练任务中的目标第一样本箱，为所述第一计算单元及所述第二分布式系统中的与所述第一计算单元关联的第二计算单元建立通信通道；

控制所述第一计算单元通过所述通信通道与所述关联的第二计算单元基于所述目标第一样本箱联合训练联邦学习模型，得到所述目标第一样本箱对应的联邦学习模型。

在一具体实施例中，所述联邦学习模型为神经网络模型（人工智能模型），所述控制所述第一计算单元通过所述通信通道与所述关联的第二计算单元基于所述目标第一样本箱联合训练联邦学习模型包括：

控制所述第一计算单元以所述目标第一样本箱为输入，通过初始化的第一神经网络子模型计算第一中间输出；

控制所述第一计算单元通过所述通信通道将加密后的第一中间输出传输给所述关联的第二计算单元，使所述关联的第二计算单元以所述第一中间输出为输入，通过初始化的第二神经网络子模型计算第二中间输出，将所述第二中间输出通过所述通信通道传

输给所述第一计算单元；

控制所述第一计算单元以所述第二中间输出为输入，通过所述第一神经网络子模型计算最终输出，根据所述最终输出和所述最终输出对应的标签通过反向传播算法优化所述第一神经网络子模型的参数；

控制所述第一计算单元通过反向传播算法计算中间误差，将所述中间误差通过所述通信通道传输给所述关联的第二计算单元，使所述关联的第二计算单元根据所述中间误差和所述第二中间输出通过反向传播算法优化所述第二神经网络子模型的参数；

组合所述第一神经网络子模型和所述第二神经网络子模型，得到所述联邦学习模型。

在另一实施例中，所述第一计算单元与所述关联的第二计算单元通过所述通信通道进行加密通信。用所述预设加密方法加密所述第一计算单元与所述关联的第二计算单元通过所述通信通道传输的数据。

在另一实施例中，所述联邦学习模型可以是树模型。

在一具体实施例中，所述调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务包括：

获取所述多个第一计算单元的第三数量和所述训练任务队列中的训练任务的第四数量；

判断所述第三数量是否小于所述第四数量；

当所述第三数量大于或等于所述第四数量时，用所述多个第一计算单元和所述第二分布式系统的多个第二计算单元联合处理所述训练任务队列中的全部训练任务；

当所述第三数量小于所述第四数量时，用所述多个第一计算单元和所述第二分布式系统的多个第二计算单元联合，分批处理所述训练任务队列中的训练任务。

实施例一的联邦学习模型训练方法通过分布式系统的多个计算单元对待处理匹配任务和训练任务进行异步处理，避免待处理任务在单位时间点对计算资源的需求压力，同时提高容错性，提升了训练联邦学习模型的效率。

在另一实施例中，所述联邦学习模型训练方法还包括：

从所述第一分布式系统获取多个空闲状态的第一计算单元；

将所述多个空闲状态的第一计算单元动态分配给所述分配任务队列和所述训练任务队列，使所述多个空闲状态的第一计算单元处理所述分配任务队列中的分配任务和/或所述训练任务队列中的训练任务。

所述将所述多个空闲状态的第一计算单元动态分配给所述分配任务队列和所述训练任务队列包括：

获取所述分配任务队列中的分配任务数量和所述训练任务队列中的训练任务数量；

按所述分配任务数量与所述训练任务数量的比例，将所述多个空闲状态的第一计算单元等比例分配给所述分配任务队列和所述训练任务队列。

根据任务队列和计算资源分配任务，从而有效利用计算资源。

在另一实施例中，所述联邦学习模型训练方法还包括：

对所述多个第一计算单元进行心跳检测，检测出停止心跳的第一计算单元；发送对所述停止心跳的第一计算单元的提醒信息给用户。

实施例二

图 2 是本申请实施例二提供的联邦学习模型训练装置的结构图。所述联邦学习模型训练装置 20 应用于第一分布式系统中，所述第一分布式系统由多个计算机设备构成。所述联邦学习模型训练装置 20 用于训练联邦学习模型。

如图 2 所示，所述联邦学习模型训练装置 20 可以包括获取模块 201、加密模块 202、分箱模块 203、第一加入模块 204、匹配模块 205、第二加入模块 206、处理模块 207。

获取模块 201，用于从第二分布式系统获取预设分箱规则、预设加密方法。

所述第一分布式系统可以是数据请求方，所述第二分布式系统可以是数据提供方，数据请求方需要联合数据提供方进行联邦学习，并不需要获取数据提供方的数据。例如，数据请求方为保险公司，数据提供方为金融公司；保险公司需要在不获取金融公司的数据的情况下，与金融公司进行联邦学习，以获得对客户进行可靠性评分的评分模型。

所述第一分布式系统可以使用所述预设分箱规则对样本数据进行分箱。所述预设分箱规则包括：根据数值范围或除数取余结果将样本数据分入不同的样本箱。所述第二分布式系统也是用所述预设分箱规则对样本数据进行分箱，使两个系统中相同的样本数据被分入相对应的样本箱。

所述预设加密方法可以是不可回溯的均匀加密算法。所述预设加密方法可以是哈希加密方法、对称加密方法、非对称加密方法。所述预设加密方法可以用于对样本数据的识别信息或所述第一分布式系统与所述第二分布式系统之间的通信数据进行加密。所述预设加密方法可以使所述第一分布式系统和所述第二分布式系统不获取对方的明文数据。所述第一分布式系统和所述第二分布式系统使用相同的预设加密方法，当所述第一分布式相同中的样本数据的标识信息与所述第二分布式系统中的样本数据的标识信息相同时，加密后的两个标识信息也相同。

加密模块 202，用于根据所述预设加密方法对所述第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息。

所述第一样本数据可以包括多个用户的基本信息和一些特征数据等，所述第一标识信息可以是加密后的用户的身份证号码或手机号。

例如，保险公司中的第一样本数据包括 3 条样本数据，分别为{用户 1 的标识信息，年龄，保险特征数据}、{用户 2 的标识信息，年龄，保险特征数据}、{用户 3 的标识信息，年龄，保险特征数据}。可以根据哈希算法（预设加密方法）对每个用户的识别信息进行加密。

分箱模块 203，用于根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱。

在一具体实施例中，所述联邦学习模型训练装置还包括预处理模块，用于在所述根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱之前，对所述第一样本数据进行数据清洗、格式转换、特征工程等预处理。

还可以对所述第一样本数据进行数据混淆、数据编码等预处理。

在一具体实施例中，所述根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱包括：

获取所述预设分箱规则中的预设除数；

对每个所述第一标识信息与所述预设除数进行相除并获取余数；

根据所述第一标识信息对应的余数对所述第一样本数据进行分箱，得到多个第一样本箱；

判断每个第一样本箱中的第一样本数量是否大于预设数量阈值；

对第一样本数量大于所述预设数量阈值的第一样本箱进行均匀分割，得到多个子箱。

例如，用 30 个第一标识信息分别除以 6（预设除数），得到各个第一标识信息的余数。将第一标识信息的余数为 1 的第一样本数据分入第 1 个第一样本箱；将第一标识信息的余数为 2 的第一样本数据分入第 2 个第一样本箱；将第一标识信息的余数为 3 的第一样本数据分入第 3 个第一样本箱；将第一标识信息的余数为 4 的第一样本数据分入第 4 个第一样本箱；将第一标识信息的余数为 5 的第一样本数据分入第 5 个第一样本箱；将第一标识信息的余数为 0 的第一样本数据分入第 6 个第一样本箱。如果判断第 2 个第一样本箱中的第一样本数量大于预设数量阈值，将第 2 个第一样本箱均匀分割为 2 个子箱（或其它数量）。

或者，所述根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分

箱包括：

获取所述预设分箱规则中的多个预设取值范围，每个预设取值范围对应一个第一样本箱；

截取每个所述第一标识信息的预设位置的数值，得到每个所述第一标识信息的截取数值；

确定所述第一标识信息的截取数值所属的目标预设取值范围；

根据所述第一标识信息的截取数值所属的目标预设取值范围，将所述第一样本数据分入目标预设取值范围对应的第一样本箱；

判断每个第一样本箱中的第一样本数量是否大于预设数量阈值；

对第一样本数量大于所述预设数量阈值的第一样本箱进行均匀分割，得到多个子箱。

例如，获取预设分箱规则中的 5 个预设取值范围（如 0-1，2-3，4-5，6-7，8-9），5 个预设取值范围分别对应 5 个第一样本箱。截取各个第一标识信息的最后一位数值（如十进制 0-9），得到每个第一标识信息的截取数值。确定每个第一标识信息的截取数值所属的目标预设取值范围，如截取数值 0、1 属于取值范围 0-1（对应第 1 个第一样本箱），截取数值 2、3 属于取值范围 2-3（对应第 2 个第一样本箱），截取数值 4、5 属于取值范围 4-5（对应第 3 个第一样本箱），截取数值 6、7 属于取值范围 6-7（对应第 4 个第一样本箱），截取数值 8、9 属于取值范围 8-9（对应第 5 个第一样本箱）。将第一标识信息的截取数值为 0 的第一样本数据分入第 1 个第一样本箱；将第一标识信息的截取数值为 1 的第一样本数据分入第 1 个第一样本箱；将第一标识信息的截取数值为 2 的第一样本数据分入第 2 个第一样本箱；以此类推。如果判断第 3 个第一样本箱中的第一样本数量大于预设数量阈值，将第 3 个第一样本箱均匀分割为 2 个子箱（或其它数量）。

通过分箱对需要计算的样本数据进行合理的分配。可以理解为把样本数据均衡分块成多个任务，交给不同的计算单元进行处理。缩小了遍历的计算量和访问压力。

第一加入模块 204，用于将所述多个第一样本箱加入匹配任务队列。

可以将每个第一样本箱打包为待处理的匹配任务，将多个匹配任务加入匹配任务队列。所述第一分布式系统中的处于空闲状态的计算单元可以依次读取并处理所述匹配任务队列中的匹配任务，以达到异步处理匹配任务的目的，提升处理效率。

匹配模块 205，用于控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱。

可以从所述第一分布式系统获取执行任务需要的多个第一计算单元。对所述多个第一计算单元进行配置。为保证配置的一致性和易操作性，可以使用 docker 向预设代码仓库拉取镜像。

在一具体实施例中，所述控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱包括：

每次控制一个第一计算单元读取所述匹配任务队列中的一个匹配任务，获取所述匹配任务中的第一样本箱；

控制所述第一计算单元从所述第二分布式系统中获取与所述第一样本箱关联的第二标识信息；

控制所述第一计算单元从所述第一样本箱中匹配出与所述第二标识信息一致的第一标识信息对应的样本数据，得到多个目标样本数据；

通过所述第一计算单元，删除所述第一样本箱中除所述多个目标样本数据外的样本数据，得到目标第一样本箱。

在另一实施例中，所述第一样本箱包括多个子箱，所述控制所述第一计算单元从所述第一样本箱中匹配出与所述第二标识信息一致的第一标识信息对应的样本数据包括：

控制所述第一计算单元从所述第一样本箱的多个子箱中匹配出与所述第二标识信息一致的第一标识信息对应的样本数据，得到多个目标样本数据。将第一样本箱分为多

个子箱，可以减少所述第一计算单元在短时间内对高运行资源的需求。避免卡顿或宕机，从而提升计算效率。

在另一实施例中，所述控制所述第一计算单元从所述第二分布式系统中获取与所述第一样本箱关联的第二标识信息包括：

为所述第一计算单元及所述第二分布式系统中的与所述第一计算单元关联的第二计算单元建立通信通道；

控制所述第一计算单元通过所述通信通道从所述关联的第二计算单元中获取与所述第一样本箱关联的第二样本箱，从所述关联的第二样本箱中获取第二样本数据的第二标识信息。

所述多个第一样本箱与所述第二分布式系统中的多个第二样本箱一一关联，处理两个关联样本箱的两个计算单元一一关联。例如第一样本箱 A 与第二样本箱 B 关联，第一计算单元 a 处理第一样本箱 A，第二计算单元 b 处理第二样本箱 B，第一计算单元 a 与第二计算单元 b 关联。

所述第一分布式系统中的一个第一计算单元与所述第二分布式系统中的一个第二计算单元一一关联，两个计算单元分别处理的两个样本箱也一一关联。

所述第二分布式系统根据所述预设加密方法对所述第二分布式系统的第二样本数据的标识信息进行加密，得到所述第二标识信息。所述第二分布式系统根据所述预设分箱规则对所述第二样本数据进行分箱，得到多个第二样本箱。

在一具体实施例中，所述控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱包括：

获取所述多个第一计算单元的第一数量和所述匹配任务队列中的匹配任务的第二数量；

判断所述第一数量是否小于所述第二数量；

当所述第一数量大于或等于所述第二数量时，用所述多个第一计算单元根据所述第一标识信息处理所述匹配任务队列中的全部匹配任务；

当所述第一数量小于所述第二数量时，用所述多个第一计算单元根据所述第一标识信息分批处理所述匹配任务队列中的匹配任务。

第二加入模块 206，用于将所述多个目标第一样本箱加入训练任务队列。

可以将各个目标第一样本箱打包为待处理的训练任务，将各个训练任务加入训练任务队列。所述第一分布式系统中的处于空闲状态的计算单元可以依次读取并处理所述训练任务队列中的训练任务，以达到异步处理训练任务的目的，提升处理效率。

在一具体实施例中，所述将所述多个目标第一样本箱加入训练任务队列包括：

获取每个目标第一样本箱中的第二样本数量；

当所述第二样本数量小于预设第一样本数量阈值时，合并第二样本数量小于所述预设第一样本数量阈值的目标第一样本箱，将所述合并后的第一样本箱加入所述训练任务队列；

当所述第二样本数量大于所述预设第二样本数量阈值时，拆分第二样本数量大于所述预设第二样本数量阈值的目标第一样本箱，将所述拆分后的第一样本箱加入所述训练任务队列；

其中，所述预设第二样本数量阈值大于所述预设第一样本数量阈值，所述合并后的目标第一样本箱的第二样本数量小于或等于所述预设第二样本数量阈值，且大于或等于所述预设第一样本数量阈值，所述拆分后的第一样本箱的第二样本数量小于或等于预设第二样本数量阈值，且大于或等于所述预设第一样本数量阈值。

处理模块 207，用于调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

在一具体实施例中，所述调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务包括：

每次控制一个第一计算单元读取所述训练任务队列中的一个训练任务，获取所述训练任务中的目标第一样本箱，为所述第一计算单元及所述第二分布式系统中的与所述第一计算单元关联的第二计算单元建立通信通道；

控制所述第一计算单元通过所述通信通道与所述关联的第二计算单元基于所述目标第一样本箱联合训练联邦学习模型，得到所述目标第一样本箱对应的联邦学习模型。

在一具体实施例中，所述联邦学习模型为神经网络模型（人工智能模型），所述控制所述第一计算单元通过所述通信通道与所述关联的第二计算单元基于所述目标第一样本箱联合训练联邦学习模型包括：

控制所述第一计算单元以所述目标第一样本箱为输入，通过初始化的第一神经网络子模型计算第一中间输出；

控制所述第一计算单元通过所述通信通道将加密后的第一中间输出传输给所述关联的第二计算单元，使所述关联的第二计算单元以所述第一中间输出为输入，通过初始化的第二神经网络子模型计算第二中间输出，将所述第二中间输出通过所述通信通道传输给所述第一计算单元；

控制所述第一计算单元以所述第二中间输出为输入，通过所述第一神经网络子模型计算最终输出，根据所述最终输出和所述最终输出对应的标签通过反向传播算法优化所述第一神经网络子模型的参数；

控制所述第一计算单元通过反向传播算法计算中间误差，将所述中间误差通过所述通信通道传输给所述关联的第二计算单元，使所述关联的第二计算单元根据所述中间误差和所述第二中间输出通过反向传播算法优化所述第二神经网络子模型的参数；

组合所述第一神经网络子模型和所述第二神经网络子模型，得到所述联邦学习模型。

在另一实施例中，所述第一计算单元与所述关联的第二计算单元通过所述通信通道进行加密通信。用所述预设加密方法加密所述第一计算单元与所述关联的第二计算单元通过所述通信通道传输的数据。

在另一实施例中，所述联邦学习模型可以是树模型。

在一具体实施例中，所述调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务包括：

获取所述多个第一计算单元的第三数量和所述训练任务队列中的训练任务的第四数量；

判断所述第三数量是否小于所述第四数量；

当所述第三数量大于或等于所述第四数量时，用所述多个第一计算单元和所述第二分布式系统的多个第二计算单元联合处理所述训练任务队列中的全部训练任务；

当所述第三数量小于所述第四数量时，用所述多个第一计算单元和所述第二分布式系统的多个第二计算单元联合，分批处理所述训练任务队列中的训练任务。

实施例二的联邦学习模型训练装置 20 通过分布式系统的多个计算单元对待处理匹配任务和训练任务进行异步处理，避免待处理任务在单位时间点对计算资源的需求压力，同时提高容错性，提升了训练联邦学习模型的效率。

在另一实施例中，所述处理模块还用于从所述第一分布式系统获取多个空闲状态的第一计算单元；将所述多个空闲状态的第一计算单元动态分配给所述分配任务队列和所述训练任务队列，使所述多个空闲状态的第一计算单元处理所述分配任务队列中的分配任务和/或所述训练任务队列中的训练任务。

所述将所述多个空闲状态的第一计算单元动态分配给所述分配任务队列和所述训练任务队列包括：

获取所述分配任务队列中的分配任务数量和所述训练任务队列中的训练任务数量；

按所述分配任务数量与所述训练任务数量的比例，将所述多个空闲状态的第一计算单元等比例分配给所述分配任务队列和所述训练任务队列。

根据任务队列和计算资源分配任务，从而有效利用计算资源。

在另一实施例中，所述联邦学习模型训练装置 20 还包括发送模块，用于对所述多个第一计算单元进行心跳检测，检测出停止心跳的第一计算单元；发送对所述停止心跳的第一计算单元的提醒信息给用户。

实施例三

本实施例提供一种计算机可读存储介质，该计算机可读存储介质上存储有计算机可读指令，所述计算机可读存储介质可以是非易失性，也可以是易失性。该计算机可读指令被处理器执行时实现上述联邦学习模型训练方法实施例中的步骤，例如图 1 所示的步骤 101-107：

101，从第二分布式系统获取预设分箱规则、预设加密方法；

102，根据所述预设加密方法对第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

103，根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

104，将所述多个第一样本箱加入匹配任务队列；

105，控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

106，将所述多个目标第一样本箱加入训练任务队列；

107，调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

或者，该计算机可读指令被处理器执行时实现上述装置实施例中各模块的功能，例如图 2 中的模块 201-207：

获取模块 201，用于从第二分布式系统获取预设分箱规则、预设加密方法；

加密模块 202，用于根据所述预设加密方法对第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

分箱模块 203，用于根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

第一加入模块 204，用于将所述多个第一样本箱加入匹配任务队列；

匹配模块 205，用于控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

第二加入模块 206，用于将所述多个目标第一样本箱加入训练任务队列；

处理模块 207，用于调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

实施例四

图 3 为本申请实施例四提供的计算机设备的示意图。所述计算机设备 30 包括存储器 301、处理器 302 以及存储在所述存储器 301 中并可在所述处理器 302 上运行的计算机可读指令 303，例如联邦学习模型训练程序。所述处理器 302 执行所述计算机可读指令 303 时实现上述联邦学习模型训练方法实施例中的步骤，例如图 1 所示的步骤 101-107：

101，从第二分布式系统获取预设分箱规则、预设加密方法；

102，根据所述预设加密方法对第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

103，根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得

到多个第一样本箱；

104，将所述多个第一样本箱加入匹配任务队列；

105，控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

106，将所述多个目标第一样本箱加入训练任务队列；

107，调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

或者，该计算机可读指令被处理器执行时实现上述装置实施例中各模块的功能，例如图 2 中的模块 201-207：

获取模块 201，用于从第二分布式系统获取预设分箱规则、预设加密方法；

加密模块 202，用于根据所述预设加密方法对第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

分箱模块 203，用于根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

第一加入模块 204，用于将所述多个第一样本箱加入匹配任务队列；

匹配模块 205，用于控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

第二加入模块 206，用于将所述多个目标第一样本箱加入训练任务队列；

处理模块 207，用于调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

示例性的，所述计算机可读指令 303 可以被分割成一个或多个模块，所述一个或者多个模块被存储在所述存储器 301 中，并由所述处理器 302 执行，以完成本方法。所述一个或多个模块可以是能够完成特定功能的一系列计算机程序指令段，该指令段用于描述所述计算机可读指令 303 在所述计算机设备 30 中的执行过程。例如，所述计算机可读指令 303 可以被分割成图 2 中的获取模块 201、加密模块 202、分箱模块 203、第一加入模块 204、匹配模块 205、第二加入模块 206、处理模块 207，各模块具体功能参见实施例二。

所述计算机设备 30 可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。本领域技术人员可以理解，所述示意图 3 仅仅是计算机设备 30 的示例，并不构成对计算机设备 30 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，例如所述计算机设备 30 还可以包括输入输出设备、网络接入设备、总线等。

所称处理器 302 可以是中央处理单元(Central Processing Unit，CPU)，还可以是其他通用处理器、数字信号处理器(Digital Signal Processor，DSP)、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、分立硬件组件等。通用处理器可以是微处理器或者该处理器 302 也可以是任何常规的处理器等，所述处理器 302 是所述计算机设备 30 的控制中心，利用各种接口和线路连接整个计算机设备 30 的各个部分。

所述存储器 301 可用于存储所述计算机可读指令 303，所述处理器 302 通过运行或执行存储在所述存储器 301 内的计算机可读指令或模块，以及调用存储在存储器 301 内的数据，实现所述计算机设备 30 的各种功能。所述存储器 301 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序（比如声音播放功能、图像播放功能等）等；存储数据区可存储根据计算机设备 30 的使用所创建的数据等。此外，存储器 301 可以包括硬盘、内存、插接式硬盘，智能存储卡（Smart Media Card， SMC），安全数字（Secure Digital， SD）卡，闪存卡（Flash Card）、至少一个

磁盘存储器件、闪存器件、只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）或其他非易失性/易失性存储器件。

所述计算机设备 30 集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读存储介质中。所述计算机可读存储介质可以是非易失性，也可以是易失性。基于这样的理解，本申请实现上述实施例方法中的全部或部分流程，也可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一计算机可读存储介质中，该计算机可读指令在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机可读指令可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读存储介质可以包括：能够携带所述计算机可读指令的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、只读存储器（ROM）、随机存取存储器（RAM）。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式。

所述作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块显示的部件可以是或者也可以不是物理模块，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，在本申请各个实施例中的各功能模块可以集成在一个处理模块中，也可以是各个模块单独物理存在，也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用硬件加软件功能模块的形式实现。

上述以软件功能模块的形式实现的集成的模块，可以存储在一个计算机可读存储介质中。上述软件功能模块存储在一个存储介质中，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）或处理器（processor）执行本申请各个实施例所述联邦学习模型训练方法的部分步骤。

对于本领域技术人员而言，显然本申请不限于上述示范性实施例的细节，而且在不背离本申请的精神或基本特征的情况下，能够以其他的具体形式实现本申请。因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本申请的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本申请内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。此外，显然“包括”一词不排除其他模块或步骤，单数不排除复数。系统权利要求中陈述的多个模块或装置也可以由一个模块或装置通过软件或者硬件来实现。第一，第二等词语用来表示名称，而并不表示任何特定的顺序。

最后应说明的是，以上实施例仅用以说明本申请的技术方案而非限制，尽管参照较佳实施例对本申请进行了详细说明，本领域的普通技术人员应当理解，可以对本申请的技术方案进行修改或等同替换，而不脱离本申请技术方案的精神和范围。

权 利 要 求 书

1. 一种联邦学习模型训练方法，所述联邦学习模型训练方法应用于第一分布式系统中，其中，所述联邦学习模型训练方法包括：

从第二分布式系统获取预设分箱规则、预设加密方法；

根据所述预设加密方法对所述第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

将所述多个第一样本箱加入匹配任务队列；

控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

将所述多个目标第一样本箱加入训练任务队列；

调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

2. 如权利要求 1 所述的联邦学习模型训练方法，其中，所述根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱包括：

获取所述预设分箱规则中的预设除数；

对每个所述第一标识信息与所述预设除数进行相除并获取余数；

根据所述第一标识信息对应的余数对所述第一样本数据进行分箱，得到多个第一样本箱；

判断每个第一样本箱中的第一样本数量是否大于预设数量阈值；

对第一样本数量大于所述预设数量阈值的第一样本箱进行均匀分割，得到多个子箱。

3. 如权利要求 1 所述的联邦学习模型训练方法，其中，所述控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱包括：

每次控制一个第一计算单元读取所述匹配任务队列中的一个匹配任务，获取所述匹配任务中的第一样本箱；

控制所述第一计算单元从所述第二分布式系统中获取与所述第一样本箱关联的第二标识信息；

控制所述第一计算单元从所述第一样本箱中匹配出与所述第二标识信息一致的第一标识信息对应的样本数据，得到多个目标样本数据；

通过所述第一计算单元，删除所述第一样本箱中除所述多个目标样本数据外的样本数据，得到目标第一样本箱。

4. 如权利要求 3 所述的联邦学习模型训练方法，其中，所述控制所述第一计算单元从所述第二分布式系统中获取与所述第一样本箱关联的第二标识信息包括：

为所述第一计算单元及所述第二分布式系统中的与所述第一计算单元关联的第二计算单元建立通信通道；

控制所述第一计算单元通过所述通信通道从所述关联的第二计算单元中获取与所述第一样本箱关联的第二样本箱，从所述关联的第二样本箱中获取第二样本数据的第二标识信息。

5. 如权利要求 1 所述的联邦学习模型训练方法，其中，所述将所述多个目标第一样本箱加入训练任务队列包括：

获取每个目标第一样本箱中的第二样本数量；

当所述第二样本数量小于预设第一样本数量阈值时，合并第二样本数量小于所述预

设第一样本数量阈值的目标第一样本箱，将所述合并后的第一样本箱加入所述训练任务队列；

当所述第二样本数量大于所述预设第二样本数量阈值时，拆分第二样本数量大于所述预设第二样本数量阈值的目标第一样本箱，将所述拆分后的第一样本箱加入所述训练任务队列；

其中，所述预设第二样本数量阈值大于所述预设第一样本数量阈值，所述合并后的目标第一样本箱的第二样本数量小于或等于所述预设第二样本数量阈值，且大于或等于所述预设第一样本数量阈值，所述拆分后的第一样本箱的第二样本数量小于或等于预设第二样本数量阈值，且大于或等于所述预设第一样本数量阈值。

6. 如权利要求 1 所述的联邦学习模型训练方法，其中，所述调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务包括：

每次控制一个第一计算单元读取所述训练任务队列中的一个训练任务，获取所述训练任务中的目标第一样本箱，为所述第一计算单元及所述第二分布式系统中的与所述第一计算单元关联的第二计算单元建立通信通道；

控制所述第一计算单元通过所述通信通道与所述关联的第二计算单元基于所述目标第一样本箱联合训练联邦学习模型，得到所述目标第一样本箱对应的联邦学习模型。

7. 如权利要求 6 所述的联邦学习模型训练方法，其中，所述第一计算单元与所述关联的第二计算单元通过所述通信通道进行加密通信。

8. 一种联邦学习模型训练装置，所述联邦学习模型训练装置运行于第一分布式系统中，其中，所述联邦学习模型训练装置包括：

获取模块，用于从第二分布式系统获取预设分箱规则、预设加密方法；

加密模块，用于根据所述预设加密方法对所述第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

分箱模块，用于根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

第一加入模块，用于将所述多个第一样本箱加入匹配任务队列；

匹配模块，用于控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

第二加入模块，用于将所述多个目标第一样本箱加入训练任务队列；

处理模块，用于调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

9. 一种计算机设备，多个所述计算机设备组成第一分布式系统，其中，每个计算机设备包括处理器，所述处理器用于执行存储器中存储的计算机可读指令以实现以下步骤：

从第二分布式系统获取预设分箱规则、预设加密方法；

根据所述预设加密方法对所述第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

将所述多个第一样本箱加入匹配任务队列；

控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

将所述多个目标第一样本箱加入训练任务队列；

调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型；

所述第一分布式系统是数据请求方，所述第二分布式系统是数据提供方，数据请求方需要联合数据提供方进行联邦学习。

10. 如权利要求 9 所述的计算机设备，其中，所述处理器执行所述存储器中存储的计算机可读指令以实现所述根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱时，包括：

获取所述预设分箱规则中的预设除数；

对每个所述第一标识信息与所述预设除数进行相除并获取余数；

根据所述第一标识信息对应的余数对所述第一样本数据进行分箱，得到多个第一样本箱；

判断每个第一样本箱中的第一样本数量是否大于预设数量阈值；

对第一样本数量大于所述预设数量阈值的第一样本箱进行均匀分割，得到多个子箱。

11. 如权利要求 9 所述的计算机设备，其中，所述处理器执行所述存储器中存储的计算机可读指令以实现所述控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱时，包括：

每次控制一个第一计算单元读取所述匹配任务队列中的一个匹配任务，获取所述匹配任务中的第一样本箱；

控制所述第一计算单元从所述第二分布式系统中获取与所述第一样本箱关联的第二标识信息；

控制所述第一计算单元从所述第一样本箱中匹配出与所述第二标识信息一致的第一标识信息对应的样本数据，得到多个目标样本数据；

通过所述第一计算单元，删除所述第一样本箱中除所述多个目标样本数据外的样本数据，得到目标第一样本箱。

12. 如权利要求 11 所述的计算机设备，其中，所述处理器执行所述存储器中存储的计算机可读指令以实现所述控制所述第一计算单元从所述第二分布式系统中获取与所述第一样本箱关联的第二标识信息时，包括：

为所述第一计算单元及所述第二分布式系统中的与所述第一计算单元关联的第二计算单元建立通信通道；

控制所述第一计算单元通过所述通信通道从所述关联的第二计算单元中获取与所述第一样本箱关联的第二样本箱，从所述关联的第二样本箱中获取第二样本数据的第二标识信息。

13. 如权利要求 9 所述的计算机设备，其中，所述处理器执行所述存储器中存储的计算机可读指令以实现所述将所述多个目标第一样本箱加入训练任务队列时，包括：

获取每个目标第一样本箱中的第二样本数量；

当所述第二样本数量小于预设第一样本数量阈值时，合并第二样本数量小于所述预设第一样本数量阈值的目标第一样本箱，将所述合并后的第一样本箱加入所述训练任务队列；

当所述第二样本数量大于所述预设第二样本数量阈值时，拆分第二样本数量大于所述预设第二样本数量阈值的目标第一样本箱，将所述拆分后的第一样本箱加入所述训练任务队列；

其中，所述预设第二样本数量阈值大于所述预设第一样本数量阈值，所述合并后的目标第一样本箱的第二样本数量小于或等于所述预设第二样本数量阈值，且大于或等于所述预设第一样本数量阈值，所述拆分后的第一样本箱的第二样本数量小于或等于预设第二样本数量阈值，且大于或等于所述预设第一样本数量阈值。

14. 如权利要求 9 所述的计算机设备，其中，所述处理器执行所述存储器中存储的计算机可读指令以实现所述调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务时，包括：

每次控制一个第一计算单元读取所述训练任务队列中的一个训练任务，获取所述训练任务中的目标第一样本箱，为所述第一计算单元及所述第二分布式系统中的与所述第一计算单元关联的第二计算单元建立通信通道；

控制所述第一计算单元通过所述通信通道与所述关联的第二计算单元基于所述目标第一样本箱联合训练联邦学习模型，得到所述目标第一样本箱对应的联邦学习模型。

15. 一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机可读指令，其中，所述计算机可读指令被处理器执行时实现以下步骤：

从第二分布式系统获取预设分箱规则、预设加密方法；

根据所述预设加密方法对第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息；

根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱，得到多个第一样本箱；

将所述多个第一样本箱加入匹配任务队列；

控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱；

将所述多个目标第一样本箱加入训练任务队列；

调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型。

16. 如权利要求 15 所述的存储介质，其中，所述计算机可读指令被所述处理器执行以实现所述根据所述预设分箱规则和所述第一标识信息对所述第一样本数据进行分箱时，包括：

获取所述预设分箱规则中的预设除数；

对每个所述第一标识信息与所述预设除数进行相除并获取余数；

根据所述第一标识信息对应的余数对所述第一样本数据进行分箱，得到多个第一样本箱；

判断每个第一样本箱中的第一样本数量是否大于预设数量阈值；

对第一样本数量大于所述预设数量阈值的第一样本箱进行均匀分割，得到多个子箱。

17. 如权利要求 15 所述的存储介质，其中，所述计算机可读指令被所述处理器执行以实现所述控制所述第一分布式系统中的多个第一计算单元根据所述第一标识信息从所述匹配任务队列中匹配出多个目标第一样本箱时，包括：

每次控制一个第一计算单元读取所述匹配任务队列中的一个匹配任务，获取所述匹配任务中的第一样本箱；

控制所述第一计算单元从所述第二分布式系统中获取与所述第一样本箱关联的第二标识信息；

控制所述第一计算单元从所述第一样本箱中匹配出与所述第二标识信息一致的第一标识信息对应的样本数据，得到多个目标样本数据；

通过所述第一计算单元，删除所述第一样本箱中除所述多个目标样本数据外的样本数据，得到目标第一样本箱。

18. 如权利要求 17 所述的存储介质，其中，所述计算机可读指令被所述处理器执行以实现所述控制所述第一计算单元从所述第二分布式系统中获取与所述第一样本箱关联的第二标识信息时，包括：

为所述第一计算单元及所述第二分布式系统中的与所述第一计算单元关联的第二计算单元建立通信通道；

控制所述第一计算单元通过所述通信通道从所述关联的第二计算单元中获取与所述第一样本箱关联的第二样本箱，从所述关联的第二样本箱中获取第二样本数据的第二标识信息。

19. 如权利要求 15 所述的存储介质，其中，所述计算机可读指令被所述处理器执行以实现所述将所述多个目标第一样本箱加入训练任务队列时，包括：

获取每个目标第一样本箱中的第二样本数量；

当所述第二样本数量小于预设第一样本数量阈值时，合并第二样本数量小于所述预设第一样本数量阈值的目标第一样本箱，将所述合并后的第一样本箱加入所述训练任务队列；

当所述第二样本数量大于所述预设第二样本数量阈值时，拆分第二样本数量大于所述预设第二样本数量阈值的目标第一样本箱，将所述拆分后的第一样本箱加入所述训练任务队列；

其中，所述预设第二样本数量阈值大于所述预设第一样本数量阈值，所述合并后的目标第一样本箱的第二样本数量小于或等于所述预设第二样本数量阈值，且大于或等于所述预设第一样本数量阈值，所述拆分后的第一样本箱的第二样本数量小于或等于预设第二样本数量阈值，且大于或等于所述预设第一样本数量阈值。

20. 如权利要求 15 所述的存储介质，其中，所述计算机可读指令被所述处理器执行以实现所述调用所述多个第一计算单元和所述第二分布式系统中的多个第二计算单元联合处理所述训练任务队列中的训练任务时，包括：

每次控制一个第一计算单元读取所述训练任务队列中的一个训练任务，获取所述训练任务中的目标第一样本箱，为所述第一计算单元及所述第二分布式系统中的与所述第一计算单元关联的第二计算单元建立通信通道；

控制所述第一计算单元通过所述通信通道与所述关联的第二计算单元基于所述目标第一样本箱联合训练联邦学习模型，得到所述目标第一样本箱对应的联邦学习模型。

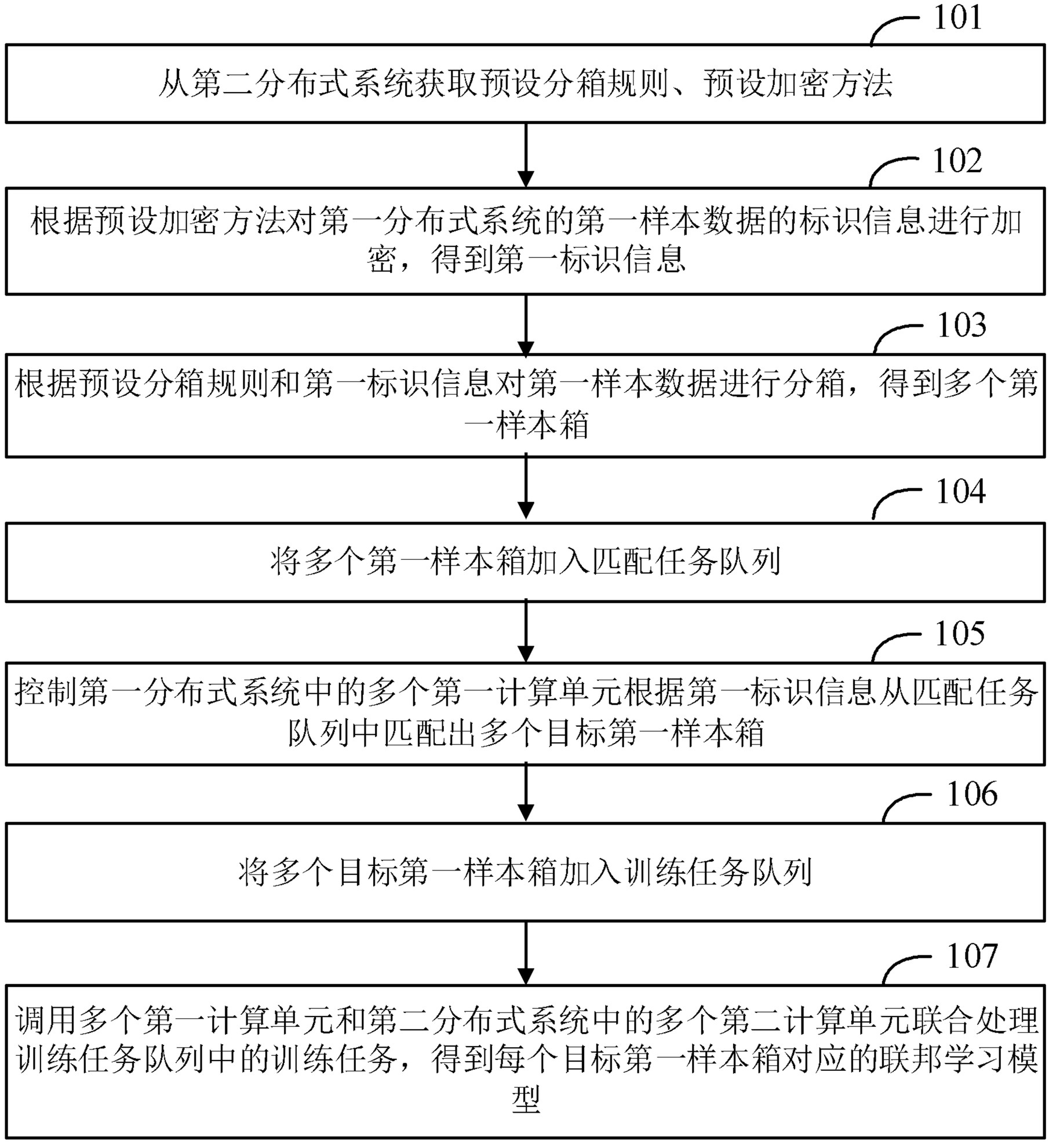

101
从第二分布式系统获取预设分箱规则、预设加密方法
102
根据预设加密方法对第一分布式系统的第一样本数据的标识信息进行加密，得到第一标识信息
103
根据预设分箱规则和第一标识信息对第一样本数据进行分箱，得到多个第一样本箱
104
将多个第一样本箱加入匹配任务队列
105
控制第一分布式系统中的多个第一计算单元根据第一标识信息从匹配任务队列中匹配出多个目标第一样本箱
106
将多个目标第一样本箱加入训练任务队列
107
调用多个第一计算单元和第二分布式系统中的多个第二计算单元联合处理训练任务队列中的训练任务，得到每个目标第一样本箱对应的联邦学习模型

图 1

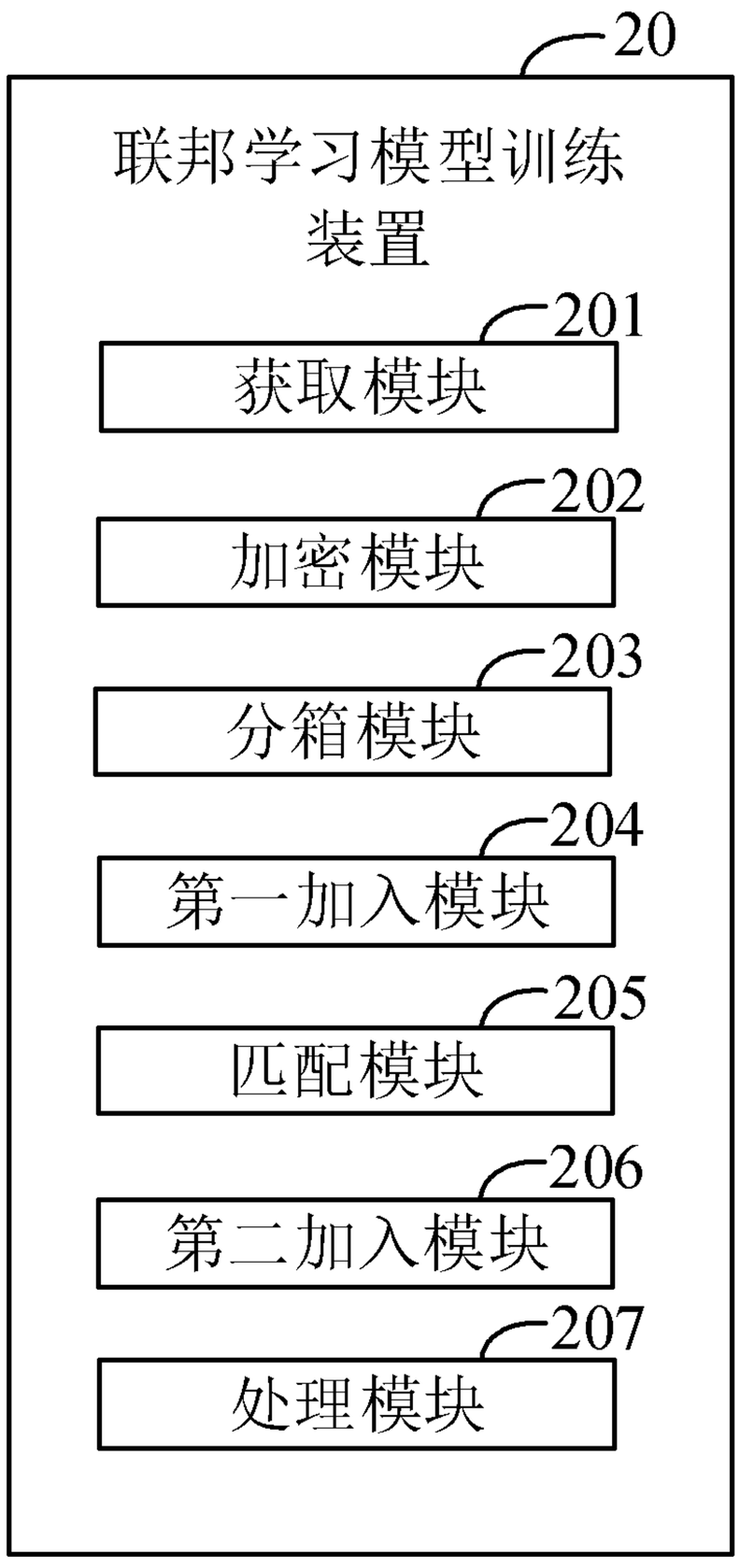

20
联邦学习模型训练装置
201
获取模块
202
加密模块
203
分箱模块
204
第一加入模块
205
匹配模块
206
第二加入模块
207
处理模块

图 2

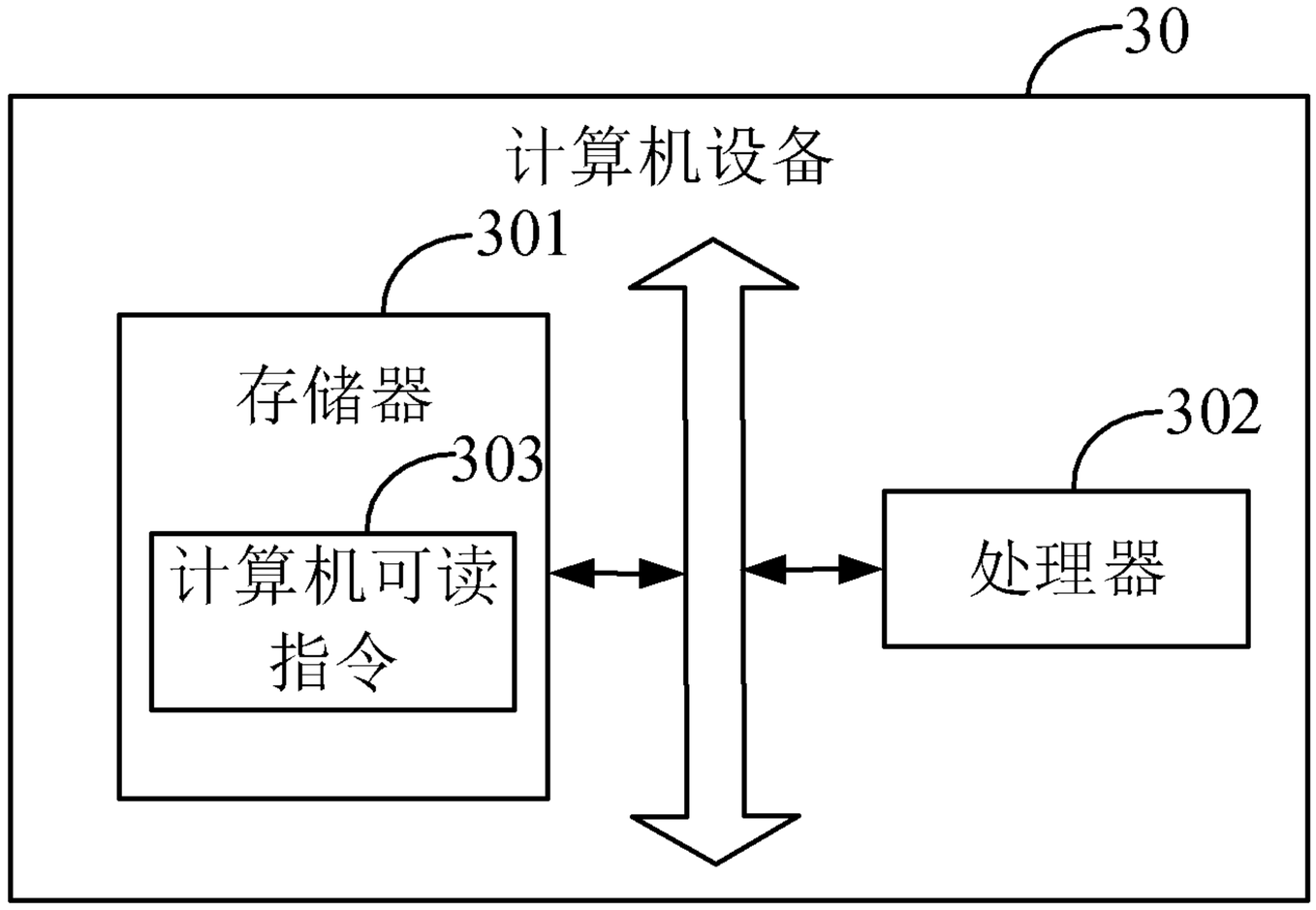


图 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2020/098884

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/04(2006.01)i; G06N 3/08(2006.01)i; 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

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)

EPODOC, WPI, CNPAT, CNKI, IEEE: 联邦, 学习, 分箱, 样本, 加密, 分布式, 第一, 第二, 匹配, 目标, 队列, federation, learning, binning, sample, encryption, distributed, first, second, matching, target, queue

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 110990857 A (ALIPAY (HANGZHOU) INFORMATION TECHNOLOGY CO., LTD.) 10 April 2020 (2020-04-10) abstract, description paragraphs 2-5, 71-101, 123-142 and 171-172, figures 1A-3 and 5	1-20
A	CN 111104731 A (BEIJING GEO POLYMERIZATION TECHNOLOGY CO., LTD.) 05 May 2020 (2020-05-05) entire document	1-20
A	CN 110825476 A (SHENZHEN QIANHAI WEBANK CO., LTD.) 21 February 2020 (2020-02-21) entire document	1-20
A	CN 110797124 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 14 February 2020 (2020-02-14) entire document	1-20
A	CN 110991552 A (ZHIFUBAO (HANGZHOU) INFORMATION TECHNOLOGY CO., LTD.) 10 April 2020 (2020-04-10) entire document	1-20
A	US 2014142966 A1 (AIONEX, INC.) 22 May 2014 (2014-05-22) 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	Date of mailing of the international search report
03 February 2021	**18 February 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/CN2020/098884

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110990857	A	10 April 2020	None			
CN	111104731	A	05 May 2020	None			
CN	110825476	A	21 February 2020	None			
CN	110797124	A	14 February 2020	None			
CN	110991552	A	10 April 2020	None			
US	2014142966	A1	22 May 2014	US	2009313046	A1	17 December 2009

国际检索报告

国际申请号
PCT/CN2020/098884

A. 主题的分类

G06N 3/04(2006.01)i; G06N 3/08(2006.01)i; 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

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））

EPODOC, WPI, CNPAT, CNKI, IEEE：联邦，学习，分箱，样本，加密，分布式，第一，第二，匹配，目标，队列，federation, learning, binning, sample, encryption, distributed, first, second, matching, target, queue

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 110990857 A (支付宝杭州信息技术有限公司) 2020年 4月 10日 (2020-04-10) 摘要, 说明书2-5、71-101、123-142和171-172段, 图1A-3和5	1-20
A	CN 111104731 A (北京集奥聚合科技有限公司) 2020年 5月 5日 (2020-05-05) 全文	1-20
A	CN 110825476 A (深圳前海微众银行股份有限公司) 2020年 2月 21日 (2020-02-21) 全文	1-20
A	CN 110797124 A (腾讯科技深圳有限公司) 2020年 2月 14日 (2020-02-14) 全文	1-20
A	CN 110991552 A (支付宝杭州信息技术有限公司) 2020年 4月 10日 (2020-04-10) 全文	1-20
A	US 2014142966 A1 (AIONEX, INC.) 2014年 5月 22日 (2014-05-22) 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	国际检索报告邮寄日期
2021年 2月 3日	2021年 2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 姜晓庆 电话号码 86-(10)-53961427

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2020/098884

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	110990857	A	2020年 4月 10日	无			
CN	111104731	A	2020年 5月 5日	无			
CN	110825476	A	2020年 2月 21日	无			
CN	110797124	A	2020年 2月 14日	无			
CN	110991552	A	2020年 4月 10日	无			
US	2014142966	A1	2014年 5月 22日	US	2009313046	A1	2009年 12月 17日