

(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
2019 年 3 月 7 日 (07.03.2019)



(10) 国际公布号
WO 2019/042199 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01) **G06N 99/00** (2010.01)
- (21) 国际申请号: PCT/CN2018/101693
- (22) 国际申请日: 2018 年 8 月 22 日 (22.08.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201710762586.9 2017年8月30日 (30.08.2017) CN
- (71) 申请人: 第四范式 (北京) 技术有限公司(**THE FOURTH PARADIGM (BEIJING) TECH CO LTD**) [CN/CN]; 中国北京市海淀区上地东路 35 号颐泉汇大厦写字楼 A 座 610 室, Beijing 100085 (CN)。
- (72) 发明人: 戴文渊(**DAI, Wenyuan**); 中国北京市海淀区上地东路 35 号颐泉汇大厦写字楼 A 座 610 室, Beijing 100085 (CN)。 杨强(**YANG, Qiang**); 中国北京市海淀区上地东路 35 号颐泉汇大厦写字楼 A 座 610 室, Beijing 100085 (CN)。 陈雨强(**CHEN,**

Yuqiang); 中国北京市海淀区上地东路 35 号颐泉汇大厦写字楼 A 座 610 室, Beijing 100085 (CN)。 刘一鸣(**LIU, Yiming**); 中国北京市海淀区上地东路 35 号颐泉汇大厦写字楼 A 座 610 室, Beijing 100085 (CN)。 石光川(**SHI, Guangchuan**); 中国北京市海淀区上地东路 35 号颐泉汇大厦写字楼 A 座 610 室, Beijing 100085 (CN)。

(74) 代理人: 北京展翼知识产权代理事务所(特殊普通合伙)(**ZYX INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区中关村东路 18 号财智国际大厦 B-1503, Beijing 100083 (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,

(54) **Title:** DISTRIBUTED SYSTEM FOR EXECUTING MACHINE LEARNING, AND METHOD THEREFOR

(54) 发明名称: 用于执行机器学习的分布式系统及其方法

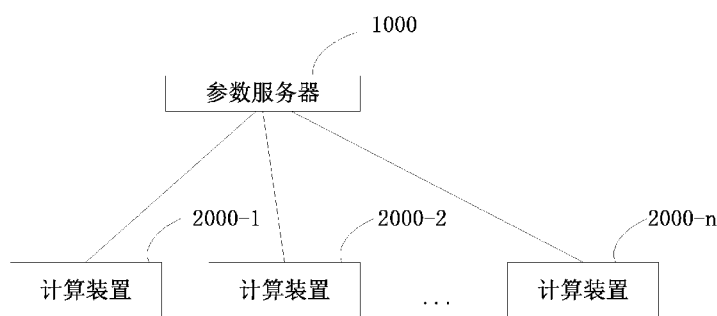


图 1

1000 PARAMETER SERVER
2000-1, 2000-2, 2000-N COMPUTING DEVICE

(57) **Abstract:** Provided are a distributed system for executing machine learning and a method therefor. The distributed system comprises: a parameter server, for maintaining parameters of a plurality of machine learning models, the parameters being in the form of a key-value pair, the parameter server saving a plurality of key-value pairs having the same key in the form of a single key corresponding to a plurality of values; and a plurality of computing devices, configured to execute, in parallel, algorithms concerning the plurality of machine learning models.

(57) **摘要:** 提供一种用于执行机器学习的分布式系统及其方法。所述分布式系统包括: 参数服务器, 用于维护多个机器学习模型的参数, 其中, 所述参数具有键值对的形式, 所述参数服务器按照单个键对应于多个值的形式来保存具有相同键的多个键值对; 以及多个计算装置, 被配置为并行地执行关于所述多个机器学习模型的算法。



WO 2019/042199 A1

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, 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))。

用于执行机器学习的分布式系统及其方法

5 技术领域

本公开的示例性实施例总体说来涉及人工智能领域，更具体地讲，涉及一种用于执行机器学习的分布式系统以及利用所述分布式系统来执行机器学习的方法。

10 背景技术

随着数据规模的迅速增长，机器学习被广泛应用于各种领域以挖掘数据的价值。然而，为了执行机器学习，一般物理机的内存已经远远不足，为此，实践中往往需要利用分布式机器学习系统来完成机器学习模型的训练或相应的预估。在现有的分布式机器学习系统中，通常由多个计算装置并行地执行关于同一机器学习模型的训练或预估，由参数服务器来存储该机器学习模型的参数，并支持各个计算装置对该机器学习模型的参数的读写。

参数服务器具有支持海量维度、高并发度、负载均衡等优点，但也存在内存占用大、运算量大、网络带宽占用大等缺点。通常一个参数服务器仅用于存储一个机器学习模型的参数，如果想要同时进行关于多个机器学习模型的训练和/或预估，则需要参数服务器存储多个机器学习模型的参数，那么存储开销也会呈线性增长。

发明内容

本公开的示例性实施例在于提供一种用于执行机器学习的分布式系统及其方法，以解决在同时执行多个机器学习模型的运算时参数服务器的存储开销过大的问题。

根据本公开的示例性实施例，提供一种用于执行机器学习的分布式系统，包括：参数服务器，用于维护多个机器学习模型的参数，其中，所述参数具有键值对的形式，所述参数服务器按照单个键对应于多个值的形式来保存具有相同键的多个键值对；以及多个计算装置，被配置为并行地执行关于所述多个机器学习模型的算法，其中，所述算法为关于机器学习模型的训练算法

或关于机器学习模型的预估算法；

其中，所述参数服务器将计算装置为了执行关于机器学习模型的训练算法而需要读取的参数提供给所述计算装置，以使所述计算装置利用读取的参数来执行所述训练算法，并且，所述参数服务器根据所述计算装置执行所述训练算法的结果来更新所述机器学习模型的参数；

或者，所述参数服务器将计算装置为了执行关于机器学习模型的预估算法而需要读取的参数提供给所述计算装置，以使所述计算装置利用读取的参数来执行所述预估算法；

或者，所述参数服务器将计算装置为了执行关于机器学习模型的训练算法而需要读取的参数提供给所述计算装置，以使所述计算装置利用读取的参数来执行所述训练算法，并且，所述参数服务器根据所述计算装置执行所述训练算法的结果来更新所述机器学习模型的参数；以及所述参数服务器将计算装置为了执行关于机器学习模型的预估算法而需要读取的参数提供给所述计算装置，以使所述计算装置利用读取的参数来执行所述预估算法。

根据本公开的另一示例性实施例，提供一种利用分布式系统来执行机器学习的方法，其中，所述分布式系统中的参数服务器用于维护多个机器学习模型的参数，其中，所述参数具有键值对的形式，所述参数服务器按照单个键对应于多个值的形式来保存具有相同键的多个键值对；所述分布式系统中的多个计算装置被配置为并行地执行关于所述多个机器学习模型的算法，其中，所述算法为关于机器学习模型的训练算法或关于机器学习模型的预估算法，所述方法包括由所述参数服务器执行的如下步骤：

所述参数服务器将计算装置为了执行关于机器学习模型的训练算法而需要读取的参数提供给所述计算装置，以使得所述计算装置利用读取的参数来执行所述训练算法；所述参数服务器根据所述计算装置执行所述训练算法的结果来更新所述机器学习模型的参数；

或者，所述参数服务器将计算装置为了执行关于机器学习模型的预估算法而需要读取的参数提供给所述计算装置，以使得所述计算装置利用读取的参数来执行所述预估算法；

或者，所述参数服务器将计算装置为了执行关于机器学习模型的训练算法而需要读取的参数提供给所述计算装置，以使得所述计算装置利用读取的参数来执行所述训练算法；所述参数服务器根据所述计算装置执行所述训练

算法的结果来更新所述机器学习模型的参数；以及所述参数服务器将计算装置为了执行关于机器学习模型的预估算法而需要读取的参数提供给所述计算装置，以使得所述计算装置利用读取的参数来执行所述预估算法。

5 根据本公开的另一示例性实施例，还提供一种参数服务器，其中，该参数服务器包括至少一个处理器和至少一个计算机可读存储介质；所述计算机可读存储介质上存储有可执行指令；当所述处理器执行所述可执行指令时实现如上所述的方法。

10 根据本公开的另一示例性实施例，还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程序，所述计算机程序在被参数服务器执行时，实现如上所述的方法。

在根据本公开示例性实施例的用于执行机器学习的分布式系统及其方法中，通过参数服务器来维护多个机器学习模型的参数，并且，参数服务器按照单个键对应于多个值的形式来保存具有相同键的多个键值对，从而能够有效减少在同时进行关于多个机器学习模型的训练和/或预估时参数服务器的存储开销。进一步地，还能够有效减少参数服务器的运算量和网络传输开销。

15 将在接下来的描述中部分阐述本公开总体构思另外的方面和/或优点，还有一部分通过描述将是清楚的，或者可以经过本公开总体构思的实施而得知。

附图说明

20 通过下面结合示例性地示出实施例的附图进行的描述，本公开示例性实施例的上述和其他目的和特点将会变得更加清楚，其中：

图 1 示出根据本公开示例性实施例的用于执行机器学习的分布式系统的框图；

图 2 示出根据本公开示例性实施例的保存机器学习模型的参数的示例；

25 图 3 示出根据本公开的另一示例性实施例的用于执行机器学习的分布式系统的框图；

图 4 示出根据本公开的另一示例性实施例的用于执行机器学习的分布式系统的框图；

30 图 5 示出根据本公开示例性实施例的客户端与关于机器学习模型的算法交互的示例；

图 6 示出根据本公开示例性实施例的利用分布式系统来执行机器学习的

方法的流程图；

图 7 示出根据本公开的另一示例性实施例的利用分布式系统来执行机器学习的方法的流程图；

图 8 示出根据本公开示例性实施例的客户端向服务器端请求计算装置为了执行关于机器学习模型的算法而需要读取的参数的方法的流程图；

图 9 示出根据本公开示例性实施例的客户端从服务器端获取计算装置需要读取的参数的方法的流程图。

具体实施方式

现将详细参照本公开的实施例，所述实施例的示例在附图中示出，其中，相同的标号始终指的是相同的部件。以下将通过参照附图来说明所述实施例，以便解释本公开。在此需要说明的是，在本公开中出现的“并且/或者”、“和/或”均表示包含三种并列的情况。例如“包括 A 和/或 B”表示如下三种并列的情况：（1）包括 A；（2）包括 B；（3）包括 A 和 B。又例如“执行步骤一并且/或者步骤二”表示如下三种并列的情况：（1）执行步骤一；（2）执行步骤二；（3）执行步骤一和步骤二。

本公开的示例性实施例所涉及的用于执行机器学习的分布式系统可由参数服务器和多个计算装置组成，其中，参数服务器通过与各个计算装置进行交互来维护多个机器学习模型的参数，多个计算装置并行地进行关于所述多个机器学习模型的训练和/或预估。应注意，这里所说的计算装置和/或参数服务器均由其所执行的处理或所实现的功能来限定，既可指示物理实体，也可指示虚拟实体，例如，计算装置可指示实际的计算机器，也可指示部署在该计算机器上的逻辑实体，同样，参数服务器既可指示实际的计算机器，也可作为一个或多个逻辑实体而与计算装置部署在同样或不同的计算机器上。

图 1 示出根据本公开示例性实施例的用于执行机器学习的分布式系统的框图。参照图 1，根据本公开示例性实施例的用于执行机器学习的分布式系统可包括：参数服务器 1000 和多个计算装置 2000（例如，2000-1、2000-2、...、2000-n（其中，n 为大于 1 的整数））。

具体说来，参数服务器 1000 用于维护多个机器学习模型的参数，其中，所述参数具有键值对（key-value）的形式。这里，机器学习模型可看作关于机器学习样本特征的函数，可通过针对全部机器学习样本不断地迭代训练，

逐渐收敛到该函数的参数最优解，相应地，一个 key-value 中的 key 可对应于一个特征的标识，value 可对应于该函数的参数之中与该特征相关的参数的值。

多个计算装置 2000 被配置为并行地执行关于所述多个机器学习模型的算法，其中，所述算法为关于机器学习模型的训练算法或关于机器学习模型的预估算法。具体说来，参数服务器 1000 将计算装置 2000 为了执行关于机器学习模型的训练算法而需要读取的参数提供给计算装置 2000，以使计算装置 2000 利用读取的参数来执行所述关于机器学习模型的训练算法，并且，参数服务器 1000 根据计算装置 2000 执行所述关于机器学习模型的训练算法的结果来更新所述机器学习模型的参数；并且/或者，参数服务器 1000 将计算装置 2000 为了执行关于机器学习模型的预估算法而需要读取的参数提供给计算装置 2000，以使计算装置 2000 利用读取的参数来执行所述关于机器学习模型的预估算法。

应注意，各个计算装置 2000 可执行至少一个关于机器学习模型的算法。作为示例，不同计算装置 2000 可针对各自的数据记录（这里，不同计算装置 2000 所针对的数据记录各不相同）执行相同的关于机器学习模型的算法（也即，关于同一机器学习模型的训练算法或关于同一机器学习模型的预估算法）；不同计算装置 2000 也可执行不同的关于机器学习模型的算法。

作为示例，参数服务器 1000 可按照单个键对应于多个值的形式来保存具有相同键的多个键值对。具体说来，每一个机器学习模型会对应一套键值对，在这套键值对中，不同键值对的键各不相同，每个键对应各自的值，参数服务器 1000 在保存多个机器学习模型的参数时，将具有相同键的对应于不同机器学习模型的键值对合并保存为单个键对应于多个值的形式，从而避免了在同时存储多个机器学习模型的参数时存储开销呈线性增长。如图 2 所示，机器学习模型 A 对应一套键值对，其中至少包括键 k1、k2、k3、...、km，分别对应值 v11、v12、v13、...、v1m；机器学习模型 B 对应另一套键值对，其中至少包括键 k1、k2、k3、...、km，分别对应值 v21、v22、v23、...、v2m，其中，m 为大于 1 的整数，可以看出，两套键值对之中的至少一部分键值对具有完全相同的键，因此，根据本公开的示例性实施例，参数服务器 1000 可按照单个键对应于多个值的形式来保存具有相同键的所述至少一部分键值对，例如，保存为键 k1 对应于值 v11 和 v21 两者的形式。

作为示例，参数服务器 1000 可部署在单个计算机上；或者，参数服务器 1000 可同时部署在多个计算机上。

图 3 示出根据本公开的另一示例性实施例的用于执行机器学习的分布式系统的框图。如图 3 所示，参数服务器 1000 可具有分布式结构，其中，在所述分布式结构下，参数服务器 1000 可包括至少一个服务器端 1001（图 3 中仅示出包括一个服务器端 1001 的情况）和多个客户端 1002（例如，1002-1、1002-2、...、1002-n），其中，客户端 1002 与计算装置 2000 一一对应，并且，对应的客户端 1002 和计算装置 2000 集成为一体。具体说来，对应的客户端 1002-1 和计算装置 2000-1 可集成在同一个虚拟机或物理机上，对应的客户端 1002-2 和计算装置 2000-2 可集成在同一个虚拟机或物理机上，以此类推，对应的客户端 1002-n 和计算装置 2000-n 可集成在同一个虚拟机或物理机上。

所述至少一个服务器端 1001 用于保存所述多个机器学习模型的参数。具体地，当存在单个服务器端 1001 时，该服务器端 1001 保存所述多个机器学习模型的参数；当存在多个服务器端 1001 时，每个服务器端 1001 分别保存所述多个机器学习模型的参数的一部分，并且，不同服务器端 1001 保存的参数无交集。作为示例，当存在多个服务器端 1001 时，可将键相同的键值对保存在同一参数服务器端 1001 中，以使服务器端 1001 能够按照单个键对应于多个值的形式来保存具有相同键的多个键值对。

每个客户端 1002 用于与一个或多个服务器端 1001 之间传送关于至少一个机器学习模型的算法所涉及的参数的参数操作指令，其中，与所述每个客户端 1002 对应的计算装置 2000 被配置为执行所述关于至少一个机器学习模型的算法。作为优选示例，在所述参数操作指令中，相同键可被压缩和/或合并，避免了相同键的重复传输，从而能够有效减少客户端 1002 与服务器端 1001 之间传送参数操作指令的网络开销。进一步地，当服务器端 1001 按照单个键对应于多个值的形式来保存具有相同键的多个键值对时，还能够避免在接收到参数操作指令之后重复查找相同键，从而有效减少服务器端 1001 在接收到参数操作指令之后查找所涉及的参数的运算量。

作为示例，参数操作指令中的相同键，可通过分桶方式或排序方式被合并。应该理解，参数操作指令中的相同键也可通过其他方式被合并，本公开对此不作限制。

作为示例，参数操作指令中的相同键可被直接压缩。作为另一示例，在

客户端 1002 与服务器端 1001 之间传送的参数操作指令也可以是压缩形式的参数操作指令，从而实现参数操作指令中的相同键被压缩。压缩算法能够隐式地合并相同数据，例如，可使用 Snappy 压缩算法、LZ4 压缩算法等各种适合的压缩算法。

- 5 此外，作为示例，参数操作指令中的相同键可先被合并，然后相同键被合并的参数操作指令可再被压缩，从而实现参数操作指令中的相同键被合并和压缩。

作为示例，每个客户端 1002 可从对应的计算装置 2000 接收所述关于至少一个机器学习模型的算法所涉及的参数的参数操作请求，针对保存所述参
10 数的一个或多个服务器端 1001 分别生成与所述参数操作请求相应的参数操作指令，并将生成的参数操作指令分别传送到所述一个或多个服务器端 1001。

进一步地，作为示例，每个客户端 1002 从所述一个或多个服务器端 1001 接收与所述参数的参数操作结果相应的参数操作指令，基于接收到的参数操作指令生成与所述参数操作请求之中的每一个对应的参数操作结果，并将生
15 成的参数操作结果传送到对应的计算装置 2000。

作为示例，所述参数操作请求可包括拉取（pull）操作请求和/或推送（push）操作请求。具体说来，pull 操作请求用于请求读取机器学习模型的参数，例如，pull 操作请求中可指定请求读取的机器学习模型及请求读取的参数
20 的 key。相应地，与 pull 操作请求对应的参数操作结果为请求读取的参数
的具体内容。push 操作请求用于请求更新机器学习模型的参数，例如，push 操作请求可指定请求更新的机器学习模型、请求更新的参数的 key、以及对与 key 对应的 value 的具体更新操作。相应地，与 push 操作请求对应的参数操作结果是参数服务器已完成更新操作的通知。

作为示例，参数服务器 1000 所包括的全部或部分服务器端 1001 之中的
25 每一个可和对应的一个计算装置 2000 集成为一体。图 4 示出根据本公开的另一示例性实施例的用于执行机器学习的分布式系统的框图。如图 4 所示，服务器端 1001、客户端 1002 与计算装置 2000 可一一对应，并且，对应的服务器端 1001、客户端 1002 和计算装置 2000 可集成为一体。具体说来，对应的服务器端 1001-1、客户端 1002-1 和计算装置 2000-1 可集成在同一个虚拟机
30 或物理机上，对应的服务器端 1002-1、客户端 1002-2 和计算装置 2000-2 可集成在同一个虚拟机或物理机上，以此类推，对应的服务器端 1001-n、客户

端 1002-n 和计算装置 2000-n 可集成在同一个虚拟机或物理机上。

应注意：在根据本公开的另一示例性实施例的用于执行机器学习的分布式系统中，每个计算装置 2000 为了执行关于机器学习模型的算法，可在本地使用与其集成为一体的对应的客户端 1001 所保存的参数，也可在需要使用其他客户端 1001 所维护的参数时，与所述其他客户端 1001 进行交互以得到执行关于机器学习模型的算法所需的参数。

作为示例，参数服务器 1000 可使用映射（Map）数据结构来存储键值对。作为示例，当参数服务器 1000 包括至少一个服务器端 1001 和多个客户端 1002 时，每个服务器端 1001 可使用映射（Map）数据结构来存储键值对。

作为示例，参数服务器 1000 可按照数据分片（shard）来存储键值对，从而提高参数服务器 1000 对参数进行读写操作的效率。作为示例，当参数服务器 1000 包括至少一个服务器端 1001 和多个客户端 1002 时，每个服务器端 1001 可按照数据分片来存储键值对。此外，作为示例，可针对每一个数据分片内的键值对，使用映射数据结构来进行存储。

作为示例，客户端 1002 与对应的计算装置 2000 之间针对每个算法，可采用单独的句柄（handler）来进行交互。参照图 5，计算装置 2000 所执行的关于机器学习模型的算法可不与对应的客户端 1002 直接进行交互，每个算法可通过一个单独的句柄与对应的客户端 1002 进行交互，例如，每个算法可通过对应的句柄向客户端 1002 发送 pull 操作请求、push 操作请求、wait 操作请求等参数操作请求，客户端 1002 也将与参数操作请求对应的参数操作结果发送到对应的句柄，算法可从句柄获知参数操作结果。由于关于机器学习模型的算法只需和句柄交互，因此，对于所述分布式系统的使用而言：同一计算装置 2000 执行的关于机器学习模型的算法彼此之间相对独立执行、并无影响，是纯透明的。

图 6 示出根据本公开示例性实施例的利用分布式系统来执行机器学习的方法的流程图。所述方法涉及的步骤可由之前描述的分布式系统中的计算装置和参数服务器（例如，参数服务器 1000；或者，服务器端 1001 和客户端 1002）来执行，例如，可由计算装置和参数服务器根据预先设定的配置来执行，其中，所述分布式系统中的参数服务器用于维护多个机器学习模型参数，其中，所述参数具有键值对的形式；所述分布式系统中的多个计算装置被配置为并行地执行关于所述多个机器学习模型的训练算法。

参照图 6，在步骤 S100 中，参数服务器将计算装置为了执行关于机器学习模型的训练算法而需要读取的参数提供给所述计算装置。

在步骤 S200 中，所述计算装置利用读取的参数来执行所述训练算法。

在步骤 S300 中，所述参数服务器根据所述计算装置执行所述训练算法的结果来更新所述机器学习模型的参数。

作为示例，所述参数服务器可按照单个键对应于多个值的形式来保存具有相同键的多个键值对。

本领域技术人员应该理解，在针对机器学习样本对机器学习模型进行训练的过程中，计算装置需要从本地或外部获取作为机器学习样本的数据记录。此外，在针对大量机器学习样本对机器学习模型进行训练的过程中，往往需要多次重复执行步骤 S100-S300。步骤 S100 到 S300 所涉及的操作已经参照图 1 到图 5 进行了详细描述，这里将不再赘述相关细节。

图 7 示出根据本公开的另一示例性实施例的利用分布式系统来执行机器学习的方法的流程图。所述方法涉及的步骤可由之前描述的分布式系统中的计算装置和参数服务器（例如，参数服务器 1000；或者，服务器端 1001 和客户端 1002）来执行，例如，可由计算装置和参数服务器根据预先设定的配置来执行，其中，所述分布式系统中的参数服务器用于维护多个机器学习模型的参数，其中，所述参数具有键值对的形式；所述分布式系统中的多个计算装置被配置为并行地执行关于所述多个机器学习模型的预估算法。

参照图 7，在步骤 S110 中，参数服务器将计算装置为了执行关于机器学习模型的预估算法而需要读取的参数提供给所述计算装置。

在步骤 S210 中，所述计算装置利用读取的参数来执行所述预估算法。

作为示例，所述参数服务器可按照单个键对应于多个值的形式来保存具有相同键的多个键值对。

本领域技术人员应该理解，在针对数据记录执行机器学习模型的预估（或测试）的过程中，需要从本地或外部获取用于预估的数据记录。此外，在针对大量数据记录执行机器学习模型的预估的过程中，往往需要多次重复执行步骤 S110 和 S210。步骤 S110 和 S210 所涉及的操作已经参照图 1 到图 5 进行了详细描述，这里将不再赘述相关细节。

此外，应该理解，图 6 和图 7 示出的根据本公开示例性实施例的利用分布式系统来执行机器学习的方法可以同时执行。

图 8 示出根据本公开示例性实施例的客户端向服务器端请求计算装置为了执行关于机器学习模型的算法而需要读取的参数的方法的流程图。所述方法可在步骤 S100 和/或步骤 S110 之前来执行。

这里，参数服务器可包括至少一个服务器端和多个客户端，其中，客户端与计算装置一一对应，并且，对应的客户端和计算装置集成为一体，其中，所述至少一个服务器端用于保存所述多个机器学习模型的参数；每个客户端用于与一个或多个服务器端之间传送关于至少一个机器学习模型的算法所涉及的参数的参数操作指令，其中，与所述每个客户端对应的计算装置被配置为执行所述关于至少一个机器学习模型的算法。

参照图 8，在步骤 S400 中，每个客户端从对应的计算装置接收所述关于至少一个机器学习模型的算法所涉及的参数的参数操作请求。

根据本公开的示例性实施例，每个计算装置在执行机器学习算法的过程中会向其对应的客户端请求获取和/或更新算法参数，这里，所述参数会分布地保存在一个或多个服务器端。为此，客户端在接收到任何参数操作请求之后，会将其拆分为与各个服务器端对应的参数操作请求部分，并将拆分后的各个部分保存在相应的队列中。例如，可针对每一个服务器端设置相应的队列。

在步骤 S500 中，针对保存所述参数的一个或多个服务器端分别生成与所述参数操作请求相应的参数操作指令。

作为示例，客户端每次生成参数操作指令所基于的参数操作请求可以是队列中缓存的各个部分，即，从该客户端上一次生成参数操作指令之后，到本次生成参数操作指令之前，从对应的计算装置接收到的针对相应服务器端的至少一个参数操作请求部分。作为示例，可每隔预定时间段执行一次生成参数操作指令的操作；也可每接收到预定数量的参数操作请求执行一次生成参数操作指令的操作。此外，作为示例，也可根据网络状况和/或服务器端的运行状态来确定每次执行生成参数操作指令的操作的时机。应该理解，也可根据其他预定策略来确定每次执行生成参数操作指令的操作的时机。

由于基于各个队列来分别生成与各个服务器端对应的参数操作指令，因此，考虑到每个队列中缓存有与至少一个机器学习算法相关的参数操作请求，相应生成的参数操作指令可基于相同或不同类型的参数操作请求，这些参数操作请求可针对相同或不同的机器学习算法。相应地，作为示例，在生成的

参数操作指令中，相同键可被压缩和/或合并。

在步骤 S600 中，将生成的参数操作指令分别传送到所述一个或多个服务器端。

图 9 示出根据本公开示例性实施例的客户端从服务器端获取计算装置需要读取的参数的方法的流程图。作为示例，该方法可对应于步骤 S100 和/或步骤 S110，在步骤 S600 之后被相应执行。具体说来，所述一个或多个服务器端在接收到所述每个客户端在步骤 S600 传送的参数操作指令之后，响应于接收到的参数操作指令对所述参数执行相应的操作，并返回与所述参数的参数操作结果相应的参数操作指令，其中，在返回的参数操作指令中，相同键可被压缩和/或合并。

参照图 9，在步骤 S111 中，所述每个客户端从所述一个或多个服务器端接收与所述参数的参数操作结果相应的参数操作指令。

在步骤 S112 中，基于接收到的参数操作指令生成与所述参数操作请求之中的每一个对应的参数操作结果。

在步骤 S113 中，将生成的参数操作结果传送到对应的计算装置。

例如，客户端可从对应的计算装置接收关于算法 A 的 pull 操作请求 1（所涉及的参数的键为 k1-k100）和关于算法 B 的 push 操作请求 2（所涉及的参数的键为 k20-k80），其中，键为 k1-k50 的参数保存在服务器端 1001-1，键为 k51-k100 的参数保存在服务器端 1001-2，根据本公开的示例性实施例，在按照预定策略将针对上述两个参数操作请求来生成参数操作指令的情况下，客户端 1002 可先根据 pull 操作请求 1 和 push 操作请求 2 所涉及的参数确定服务器端 1001-1 和服务器端 1001-2，然后针对服务器端 1001-1 生成与 pull 操作请求 1 和 push 操作请求 2 相应的关于键为 k1-k50 的参数的参数操作指令 C（这里，参数操作指令 C 可指示对算法 A 对应的机器学习模型的键为 k1-k50 的参数执行 pull 操作，对算法 B 对应的机器学习模型的键为 k20-k50 的参数执行 push 操作），并且，在该参数操作指令中，相同键 k20-k50 被合并和/或压缩；相应地，针对服务器端 1001-2 生成与 pull 操作请求 1 和 push 操作请求 2 相应的关于键为 k51-k100 的参数的参数操作指令 D（这里，参数操作指令 D 可指示对算法 A 对应的机器学习模型的键为 k51-k100 的参数执行 pull 操作，对算法 B 对应的机器学习模型的键为 k51-k80 的参数执行 push 操作），并且，在该参数操作指令中，相同键 k51-k80 被合并和/或压缩。服务器端

1001-1 和服务器端 1001-2 分别响应于接收到的参数操作指令 C 和 D 对相应的参数执行相应的操作，并返回参数操作指令 C'（这里，参数操作指令 C' 可指示针对算法 A 对应的机器学习模型的键为 k1-k50 的参数的 pull 操作结果，针对算法 B 对应的机器学习模型的键为 k20-k50 的参数的 push 操作结果）和参数操作指令 D'（这里，参数操作指令 D' 可指示针对算法 A 对应的机器学习模型的键为 k51-k100 的参数的 pull 操作结果，针对算法 B 对应的机器学习模型的键为 k51-k80 的参数的 push 操作结果），其中，在参数操作指令 C' 和 D' 中，相同键被合并和/或压缩。客户端基于接收到的参数操作指令 C' 和 D' 分别生成与 pull 操作请求 1 和 push 操作请求 2 相应的参数操作结果 1'（即，针对算法 A 对应的机器学习模型的键为 k1-k100 的参数的 pull 操作结果）和 2'（即，针对算法 B 对应的机器学习模型的键为 k20-k80 的参数的 push 操作结果）。这里，应注意，参与合并的参数操作请求不仅可来自不同的算法，而且可来自相同的算法。此外，在每个参数操作请求内部，也可进行相同键的合并和/或压缩。

根据本公开示例性实施例的用于执行机器学习的分布式系统及其方法能够有效减少在同时进行关于多个机器学习模型的训练和/或预估时参数服务器的存储开销。进一步地，还能够有效减少参数服务器的运算量和网络传输开销。

根据本公开示例性实施例的用于执行机器学习的分布式系统及其方法尤其适用于模型参数调优、多用户共享集群执行机器学习任务等同时训练的多个机器学习模型参数之间存在至少部分重叠的 key 的情况。

应理解，根据本公开示例性实施例的分布式系统中的参数服务器、计算装置或组成它们的装置或单元等组成部分可被分别配置为执行特定功能的软件、硬件、固件或上述项的任意组合。例如，这些组成部分可对应于专用的集成电路，也可对应于纯粹的软件代码，还可对应于软件与硬件相结合的模块。当它们以软件、固件、中间件或微代码实现时，用于执行相应操作的程序代码或者代码段可以存储在诸如存储介质的计算机可读介质中，使得处理器可通过读取并运行相应的程序代码或者代码段来执行相应的操作。此外，这些组成部分所实现的一个或多个功能也可由物理实体设备（例如，计算机器等）中的组件来统一执行。

例如，根据本公开的一个示例性实施例，提供一种参数服务器，其中，

该参数服务器包括至少一个处理器和至少一个计算机可读存储介质；所述计算机可读存储介质上存储有可执行指令；当所述处理器执行所述可执行指令时实现本公开中所描述的方法。

5 根据本公开的另一示例性实施例，还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程序，所述计算机程序在被参数服务器执行时，实现如上述实施例中所述的任一方法。

10 应注意，根据本公开示例性实施例的分布式系统可完全依赖计算机程序的运行来实现相应的功能，即，各个组成部分在计算机程序的功能架构中与各步骤相应，使得整个系统通过专门的软件包（例如，lib 库）而被调用，以实现相应的功能。

虽然已表示和描述了本公开的一些示例性实施例，但本领域技术人员应该理解，在不脱离由权利要求及其等同物限定其范围的本公开的原理和精神的情况下，可以对这些实施例进行修改。

权 利 要 求 书

1、一种用于执行机器学习的分布式系统，包括：

5 参数服务器，用于维护多个机器学习模型的参数，其中，所述参数具有键值对的形式，所述参数服务器按照单个键对应于多个值的形式来保存具有相同键的多个键值对；以及

多个计算装置，被配置为并行地执行关于所述多个机器学习模型的算法，其中，所述算法为关于机器学习模型的训练算法或关于机器学习模型的预估算法；

10 其中，所述参数服务器将计算装置为了执行关于机器学习模型的训练算法而需要读取的参数提供给所述计算装置，以使所述计算装置利用读取的参数来执行所述训练算法，并且，所述参数服务器根据所述计算装置执行所述训练算法的结果来更新所述机器学习模型的参数；

15 或者，所述参数服务器将计算装置为了执行关于机器学习模型的预估算法而需要读取的参数提供给所述计算装置，以使所述计算装置利用读取的参数来执行所述预估算法；

20 或者，所述参数服务器将计算装置为了执行关于机器学习模型的训练算法而需要读取的参数提供给所述计算装置，以使所述计算装置利用读取的参数来执行所述训练算法，并且，所述参数服务器根据所述计算装置执行所述训练算法的结果来更新所述机器学习模型的参数；以及所述参数服务器将计算装置为了执行关于机器学习模型的预估算法而需要读取的参数提供给所述计算装置，以使所述计算装置利用读取的参数来执行所述预估算法。

25 2、根据权利要求 1 所述的分布式系统，其中，所述参数服务器包括至少一个服务器端和多个客户端，其中，客户端与计算装置一一对应，并且，对应的客户端和计算装置集成为一体，

其中，所述至少一个服务器端用于保存所述多个机器学习模型的参数；每个客户端用于与一个或多个服务器端之间传送关于至少一个机器学习模型的算法所涉及的参数的参数操作指令；与所述每个客户端对应的计算装置被配置为执行关于所述至少一个机器学习模型的算法，

30 其中，在所述参数操作指令中，相同键被压缩，或被合并，或被压缩和合并。

3、根据权利要求 1 所述的分布式系统，其中，

每个客户端用于在接收到对应的计算装置的操作请求之后，将其拆分为与各个服务器端对应的操作请求部分，并将拆分后的各个部分保存在相应的队列中；其中，针对每一个服务器端设置有相应的队列；

5 每个客户端用于基于相应队列中保存的至少一个操作请求部分生成操作指令，并发送给相应的服务器端。

4、根据权利要求 1 所述的分布式系统，其中，所述参数服务器使用映射数据结构来存储键值对。

10 5、根据权利要求 1 所述的分布式系统，其中，所述参数服务器按照数据分片来存储键值对。

6、根据权利要求 2 所述的分布式系统，其中，服务器端、客户端与计算装置一一对应，并且，对应的服务器端、客户端和计算装置集成为一体。

7、根据权利要求 2 所述的分布式系统，其中，相同键通过分桶方式或排序方式被合并。

15 8、根据权利要求 2 所述的分布式系统，其中，客户端与对应的计算装置之间针对每个算法，采用单独的句柄来进行交互。

9、一种利用分布式系统来执行机器学习的方法，其中，所述分布式系统中的参数服务器用于维护多个机器学习模型的参数，其中，所述参数具有键值对的形式，所述参数服务器按照单个键对应于多个值的形式来保存具有相同键的多个键值对；所述分布式系统中的多个计算装置被配置为并行地执行关于所述多个机器学习模型的算法，其中，所述算法为关于机器学习模型的训练算法或关于机器学习模型的预估算法，所述方法包括由所述参数服务器执行的如下步骤：

20

所述参数服务器将计算装置为了执行关于机器学习模型的训练算法而需要读取的参数提供给所述计算装置，以使得所述计算装置利用读取的参数来执行所述训练算法；所述参数服务器根据所述计算装置执行所述训练算法的结果来更新所述机器学习模型的参数；

25

或者，所述参数服务器将计算装置为了执行关于机器学习模型的预估算法而需要读取的参数提供给所述计算装置，以使得所述计算装置利用读取的参数来执行所述预估算法；

30

或者，所述参数服务器将计算装置为了执行关于机器学习模型的训练算

法而需要读取的参数提供给所述计算装置，以使得所述计算装置利用读取的参数来执行所述训练算法；所述参数服务器根据所述计算装置执行所述训练算法的结果来更新所述机器学习模型的参数；以及所述参数服务器将计算装置为了执行关于机器学习模型的预估算法而需要读取的参数提供给所述计算装置，以使得所述计算装置利用读取的参数来执行所述预估算法。

10、根据权利要求 9 所述的方法，其中，所述参数服务器包括至少一个服务器端和多个客户端，其中，客户端与计算装置一一对应，并且，对应的客户端和计算装置集成为一体，其中，与所述每个客户端对应的计算装置被配置为执行关于所述至少一个机器学习模型的算法，所述至少一个服务器端用于保存所述多个机器学习模型的参数；所述方法包括：

每个客户端从对应的计算装置接收所述关于至少一个机器学习模型的算法所涉及的参数的参数操作请求；

每个客户端与一个或多个服务器端之间传送关于至少一个机器学习模型的算法所涉及的参数的参数操作指令；其中，在所述参数操作指令中，相同键被压缩，或被合并，或被压缩和合并；

以及每个客户端将参数操作结果传送到对应的计算装置。

11、根据权利要求 10 所述的方法，其中，

客户端在接收到对应计算装置的参数操作请求之后，将其拆分为与各个服务器端对应的参数操作请求部分，并将拆分后的各个部分保存在相应的队列中；其中，针对每一个服务器端设置相应的队列；

客户端基于相应队列中保存的至少一个参数操作请求部分生成参数操作指令，并发送给相应的服务器端。

12、根据权利要求 10 所述的方法，其中，客户端与对应的计算装置之间针对每个算法，采用单独的句柄来进行交互。

13、根据权利要求 9 所述的方法，其中，所述参数服务器使用映射数据结构来存储键值对。

14、根据权利要求 9 所述的方法，其中，所述参数服务器按照数据分片来存储键值对。

15、根据权利要求 10 所述的方法，其中，服务器端、客户端与计算装置一一对应，并且，对应的服务器端、客户端和计算装置集成为一体。

16、根据权利要求 10 所述的方法，其中，相同键通过分桶方式或排序方

式被合并。

- 17、一种参数服务器，其中，该参数服务器包括至少一个处理器和至少一个计算机可读存储介质；所述计算机可读存储介质上存储有可执行指令；当所述处理器执行所述可执行指令时实现如权利要求 9-16 中任一项所述的方法。
- 5

18、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程序，所述计算机程序在被参数服务器执行时，实现如权利要求 9-16 中任一项所述的方法。

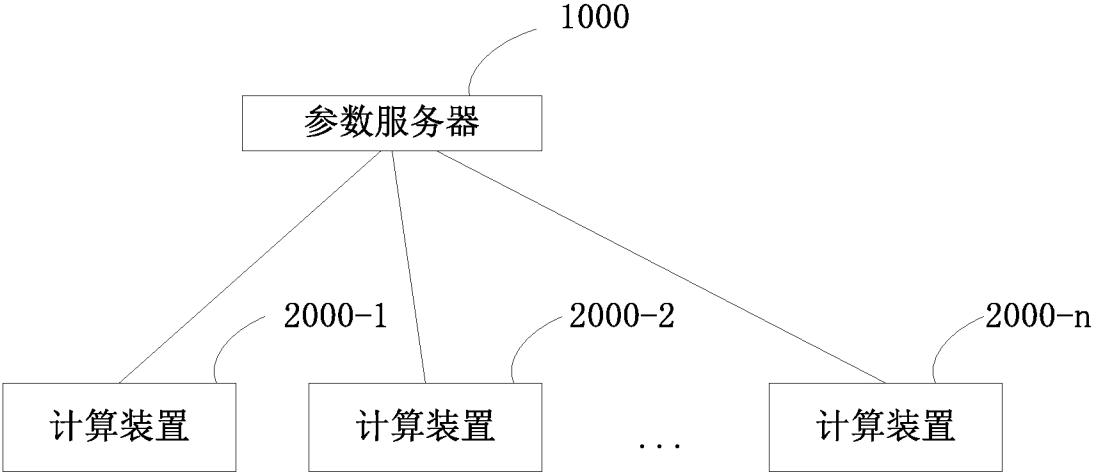


图 1

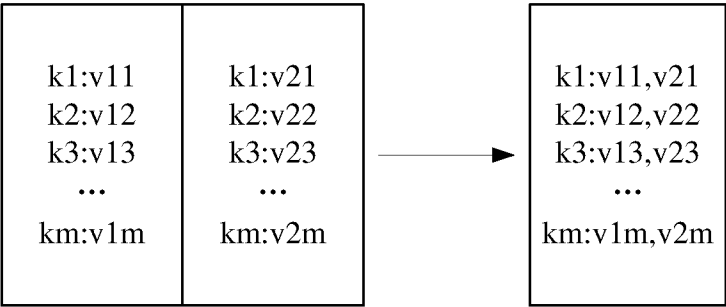


图 2

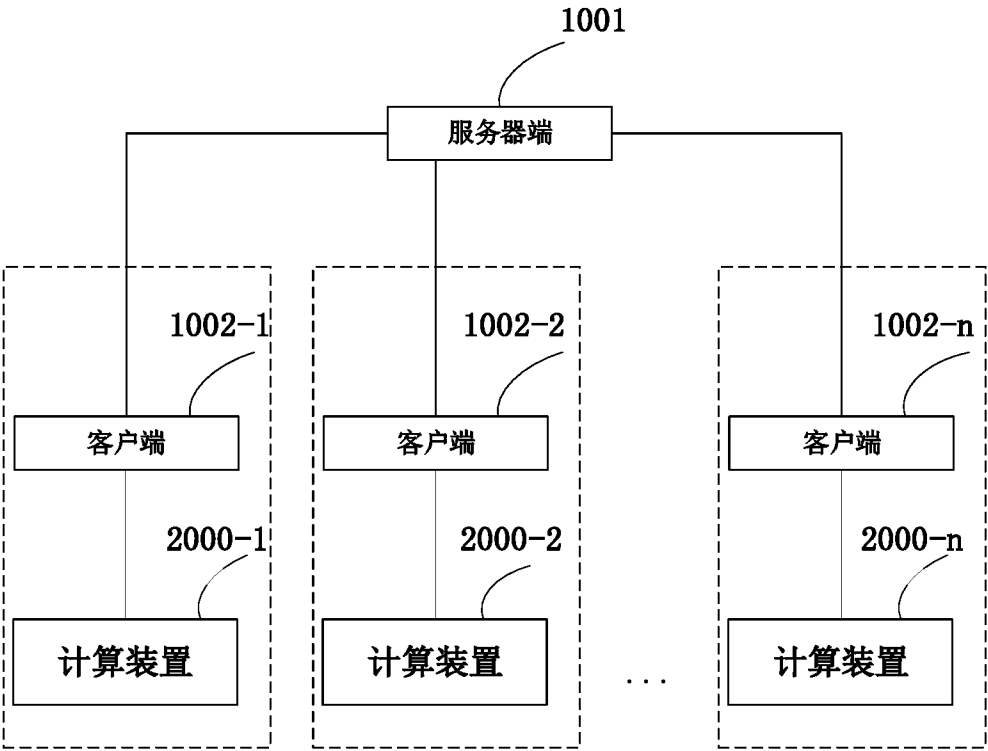


图 3

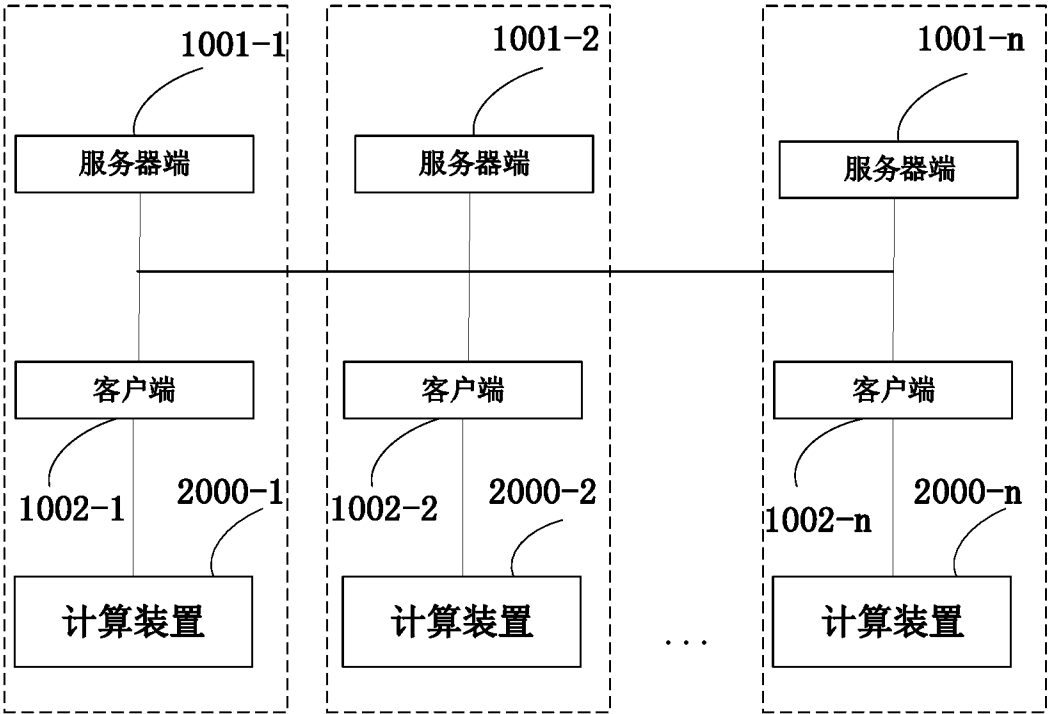


图 4

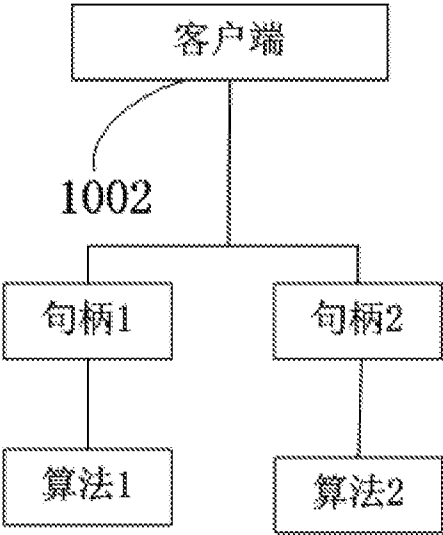


图 5

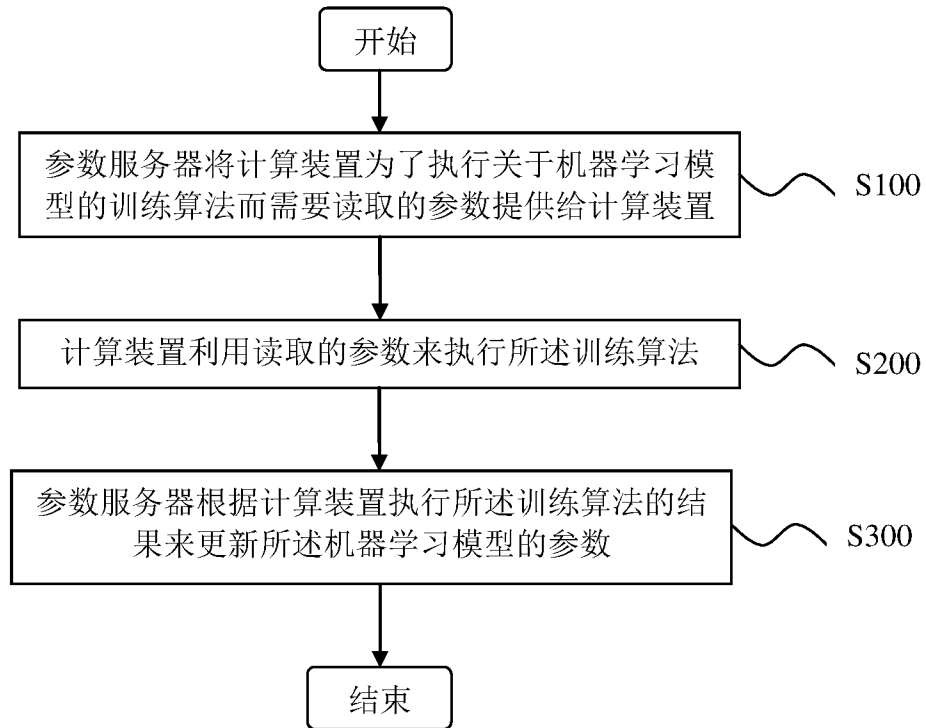


图 6

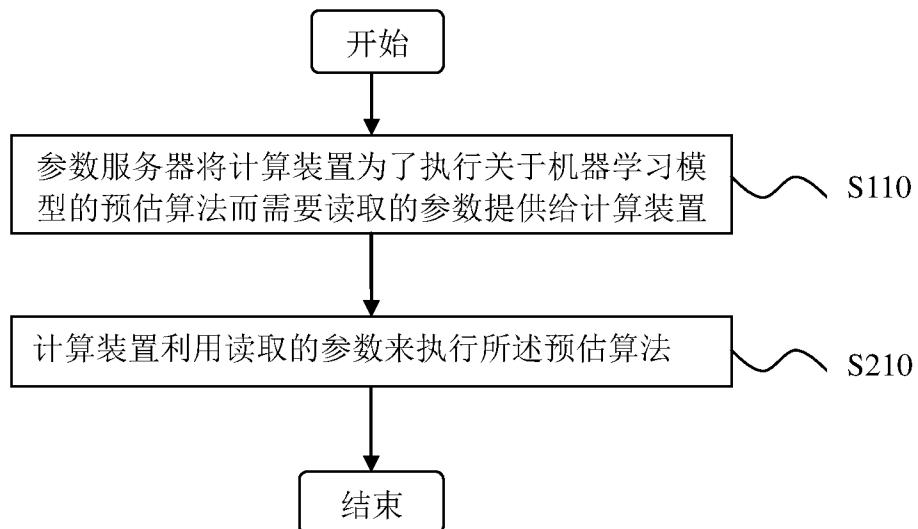


图 7

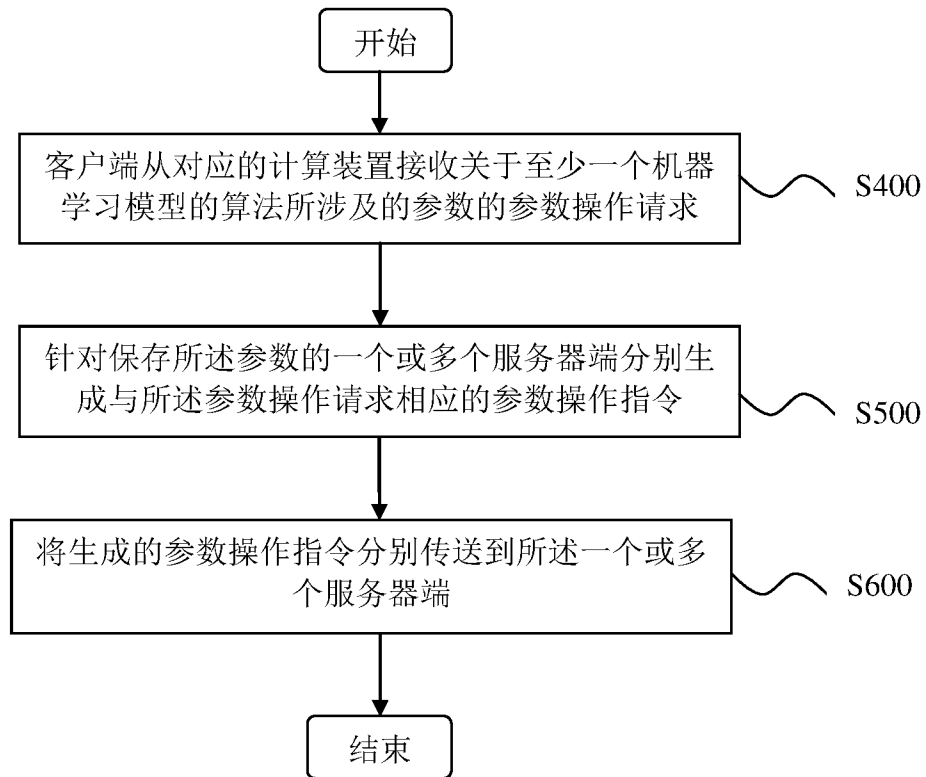


图 8

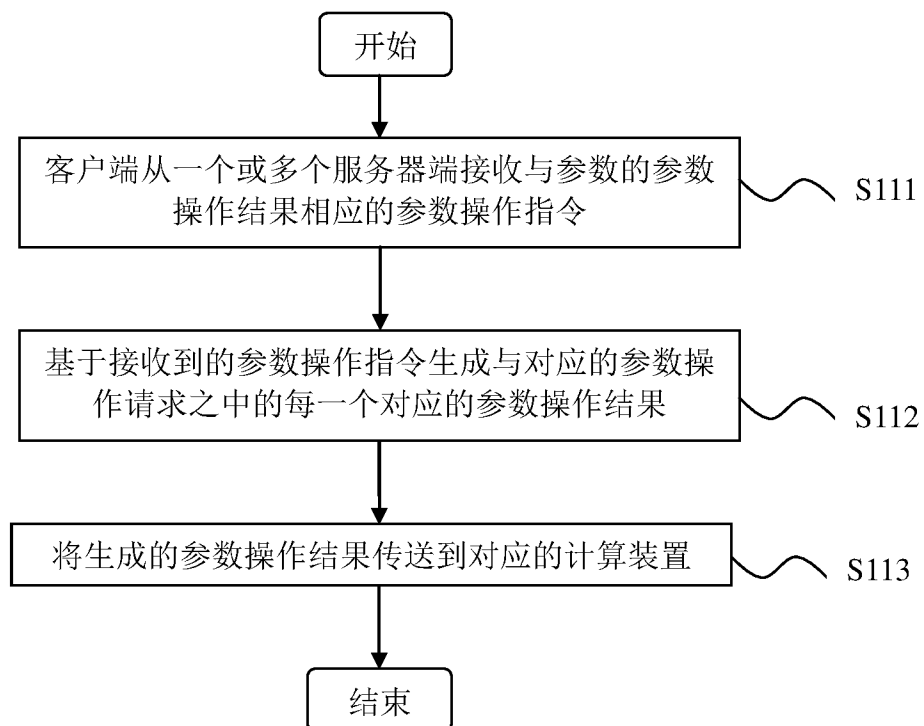


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/101693

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(2006.01)i; G06N 99/00(2010.01)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)

G06F, ; 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; IEEE: 分布式, 服务器, 客户端, 节点, 机器学习, 神经网络, 深度学习, 人工智能, 训练, 参数, 系数, 权值, 权重, 键, 值, distribution, server, client, node, machine learning, AI, neural network, NN, deep learning, CNN, DNN, train, parameter, weight, key, value

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 107729353 A (4PARADIGM (BEIJING) TECHNOLOGY CO., LTD.) 23 February 2018 (2018-02-23) claims 1-10, and description, paragraphs [0065]-[0072] and [0080]	1-18
A	CN 106156810 A (ALIBABA GROUP HOLDING LIMITED) 23 November 2016 (2016-11-23) description, paragraphs [0006]-[0036] and [0043]	1-18
A	CN 106779093 A (SHANGHAI ADVANCED RESEARCH INSTITUTE, CHINESE ACADEMY OF SCIENCES) 31 May 2017 (2017-05-31) entire document	1-18
A	US 2015234956 A1 (SAS INSTITUTE INC.) 20 August 2015 (2015-08-20) entire document	1-18
A	CN 105956166 A (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD. ET AL.) 21 September 2016 (2016-09-21) entire document	1-18
A	CN 102629219 A (PEKING UNIVERSITY) 08 August 2012 (2012-08-08) entire document	1-18

☐ 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

05 November 2018

Date of mailing of the international search report

22 November 2018

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
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/CN2018/101693

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107729353	A	23 February 2018	None			
CN	106156810	A	23 November 2016	None			
CN	106779093	A	31 May 2017	None			
US	2015234956	A1	20 August 2015	US	2018060470	A1	01 March 2018
				US	2015234955	A1	20 August 2015
				US	2017220713	A1	03 August 2017
				US	2016314226	A1	27 October 2016
CN	105956166	A	21 September 2016	None			
CN	102629219	A	08 August 2012	None			

国际检索报告

国际申请号

PCT/CN2018/101693

A. 主题的分类

G06F 17/30(2006.01)i; G06N 99/00(2010.01)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)

G06F, ; 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;IEEE: 分布式, 服务器, 客户端, 节点, 机器学习, 神经网络, 深度学习, 人工智能, 训练, 参数, 系数, 权值, 权重, 键, 值, distribution, server, client, node, machine learning, AI, neural network, NN, deep learning, CNN, DNN, train, parameter, weight, key, value

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 107729353 A (第四范式北京技术有限公司) 2018年 2月 23日 (2018 - 02 - 23) 权利要求1-10, 说明书第[0065]-[0072], [0080]段	1-18
A	CN 106156810 A (阿里巴巴集团控股有限公司) 2016年 11月 23日 (2016 - 11 - 23) 说明书第[0006]-[0036], [0043]段	1-18
A	CN 106779093 A (中国科学院上海高等研究院) 2017年 5月 31日 (2017 - 05 - 31) 全文	1-18
A	US 2015234956 A1 (SAS INSTITUTE INC.) 2015年 8月 20日 (2015 - 08 - 20) 全文	1-18
A	CN 105956166 A (北京京东尚科信息技术有限公司 等) 2016年 9月 21日 (2016 - 09 - 21) 全文	1-18
A	CN 102629219 A (北京大学) 2012年 8月 8日 (2012 - 08 - 08) 全文	1-18

☐ 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

2018年 11月 5日

国际检索报告邮寄日期

2018年 11月 22日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(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

欧晓丹

电话号码 86-(10)-53961416

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/101693

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	107729353	A	2018年 2月 23日	无			
CN	106156810	A	2016年 11月 23日	无			
CN	106779093	A	2017年 5月 31日	无			
US	2015234956	A1	2015年 8月 20日	US	2018060470	A1	2018年 3月 1日
				US	2015234955	A1	2015年 8月 20日
				US	2017220713	A1	2017年 8月 3日
				US	2016314226	A1	2016年 10月 27日
CN	105956166	A	2016年 9月 21日	无			
CN	102629219	A	2012年 8月 8日	无			

表 PCT/ISA/210 (同族专利附件) (2015年1月)