

(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
2025 年 6 月 26 日 (26.06.2025)



(10) 国际公布号
WO 2025/130623 A1

(51) 国际专利分类号:
H04L 47/10 (2022.01)

(21) 国际申请号: PCT/CN2024/136899

(22) 国际申请日: 2024 年 12 月 4 日 (04.12.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202311789769.1 2023年12月22日 (22.12.2023) CN

(71) 申请人: 中国电信股份有限公司技术创新中心 (**CHINA TELECOM CORPORATION LIMITED TECHNOLOGY INNOVATION CENTER**) [CN/CN]; 中国北京市昌平区北七家镇未来科技城南区中国电信北京信息科技创新园 11 层 1118 室、1116 室 102209 (CN)。中国电信股份有限公司 (**CHINA TELECOM CORPORATION LIMITED**) [CN/CN]; 中国北京市西城区金融大街 31 号 100033 (CN)。

(72) 发明人: 汪洋 (**WANG, Yang**); 中国北京市西城区金融大街 31 号 100033 (CN)。

(74) 代理人: 北京律智知识产权代理有限公司 (**BEIJING INTELLEGAL INTELLECTUAL PROPERTY AGENT LTD.**); 中国北京市朝阳区慧忠路 5 号 B1605、B1606、B1607 100101 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF,

(54) **Title:** NETWORK CONGESTION CONTROL METHOD FOR DISTRIBUTED MODEL TRAINING AND RELATED DEVICE

(54) 发明名称: 分布式模型训练的网络拥塞控制方法及相关设备

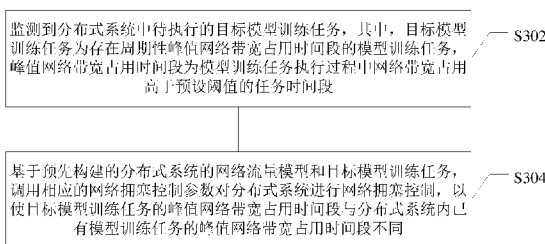


图 3

S302 Detect a target model training task to be executed in a distributed system, the target model training task being a model training task having a periodic peak network bandwidth occupation time period, and the peak network bandwidth occupation time period being a task time period when the network bandwidth occupation is higher than a preset threshold in the execution process of the model training task

S304 On the basis of a pre-constructed network traffic model of the distributed system and the target model training task, invoke a corresponding network congestion control parameter to perform network congestion control on the distributed system, such that the peak network bandwidth occupation time period of the target model training task is different from a peak network bandwidth occupation time period of an existing model training task in the distributed system

(57) **Abstract:** The present disclosure relates to the technical field of artificial intelligence. Provided are a network congestion control method and apparatus for distributed model training, a device, a medium and a computer program product. The method comprises: detecting a target model training task to be executed in a distributed system, the target model training task being a model training task having a periodic peak network bandwidth occupation time period, and the peak network bandwidth occupation time period being a task time period when the network bandwidth occupation is higher than a preset threshold in the execution process of the model training task; and, on the basis of a pre-constructed network traffic model of the distributed system and the target model training task, invoking a corresponding network congestion control parameter to perform network congestion control on the distributed system, such that the peak network bandwidth occupation time period of the target model training task is different from a peak network bandwidth occupation time period of an existing model training task in the distributed system. The present disclosure can reduce the occurrence probability of network congestion and improve the network utilization rate.



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：本公开提供了一种分布式模型训练的网络拥塞控制方法、装置、设备、介质及计算机程序产品，涉及人工智能技术领域。该方法包括：监测到分布式系统中待执行的目标模型训练任务，其中，目标模型训练任务为存在周期性峰值网络带宽占用时间段的模型训练任务，峰值网络带宽占用时间段为模型训练任务执行过程中网络带宽占用高于预设阈值的任务时间段；基于预先构建的分布式系统的网络流量模型和目标模型训练任务，调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制，以使目标模型训练任务的峰值网络带宽占用时间段与分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。本公开能够降低网络拥塞情况发生的概率，提高网络利用率。

分布式模型训练的网络拥塞控制方法及相关设备

相关申请的交叉引用

本公开要求于 2023 年 12 月 22 日提交的申请号为 202311789769.1、名称为“分布式模型训练的网络拥塞控制方法、装置、设备和介质”的中国专利申请的优先权，该中国专利

5 利申请的全部内容通过引用全部并入本文。

技术领域

本公开涉及人工智能技术领域，尤其涉及一种分布式模型训练的网络拥塞控制方法、装置、设备、介质及计算机程序产品。

10

背景技术

由于超大模型的机器学习应用往往需要大量的数据计算，因此，实际应用中为实现超大模型的机器学习应用通常会采用分布式模型进行训练作为工程解决方案。分布式机器学习将模型训练过程分散到多个计算节点上，每个节点都处理部分数据，以此来加速模型训练过程，实现大规模数据集的处理以及训练更复杂的模型。

15

在基于远程直接内存访问 RDMA 的分布式系统中，对于网络的拥塞控制通常采用公平策略，虽然 RDMA 技术提升了分布式系统的网络通信性能，但是 RDMA 的通用性适配并没有针对机器学习中的应用进行优化，可例如，存在两个模型训练任务的数据流量同时对网络产生压力时，根据 RDMA 拥塞控制的策略，会公平分配网络资源给不同的训练任务，而这种公平分配也使得两个模型训练任务在执行期间周期性的进入争抢网络资源和释放网络资源状态。特别的，当这两个模型训练任务是同类型任务时，该情况较为明显，系统中的网络拥塞将会使两个任务的执行时间都相对延迟。

20

相关技术中，实际的机器学习模型训练过程，模型训练任务并不是全时段的对网络有高流量的占用需求，这些任务对系统网络有可能产生周期性的高带宽压力。系统中出现多个模型训练任务时，相同的模型训练任务的数据同步阶段可能周期性的重叠，从而导致共享链路上周期性的网络拥塞，进而延长每个模型训练任务各自的整体训练时长。网络拥塞即在分组交换网络中传送分组的数目太多时，由于存储转发节点的资源有限而造成网络传输性能下降的情况，达成一种持续过载的网络状态，由此可能造成额外的网络开销，降低模型训练的效率。

25

需要说明的是，在上述背景技术部分公开的信息仅用于加强对本公开的背景的理解，因此可以包括不构成对本领域普通技术人员已知的现有技术的信息。

30

发明内容

本公开提供一种分布式模型训练的网络拥塞控制方法、装置、设备、介质及计算机程序产品，至少在一定程度上克服相关技术中模型训练时发生网络拥塞情况的问题。

本公开的其他特性和优点将通过下面的详细描述变得显然，或部分地通过本公开的实践而习得。

5 根据本公开的一个方面，提供了一种分布式模型训练的网络拥塞控制方法，包括：监测到分布式系统中待执行的目标模型训练任务，其中，所述目标模型训练任务为存在周期性峰值网络带宽占用时间段的模型训练任务，所述峰值网络带宽占用时间段为模型训练任务执行过程中网络带宽占用高于预设阈值的任务时间段；基于预先构建的所述分布式系统的网络流量模型和所述目标模型训练任务，调用相应的网络拥塞控制参数对所述分布式系
10 统进行网络拥塞控制，以使所述目标模型训练任务的峰值网络带宽占用时间段与所述分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。

在一些实施例中，所述监测到分布式系统中待执行的目标模型训练任务，包括：监测分布式系统是否接收到待执行的模型训练任务；若所述监测分布式系统接收到所述待执行的模型训练任务，则判断所述待执行的模型训练任务是否存在周期性峰值网络带宽占用时
15 间段；若所述待执行的模型训练任务存在周期性峰值网络带宽占用时间段，则将所述待执行的模型训练任务确定为目标模型训练任务。

在一些实施例中，判断所述待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段，包括：根据预设网络拥塞控制算法，确定待执行的模型训练任务在模型训练任务执行过程中峰值网络带宽占用时间段；根据所述待执行的模型训练任务的峰值网络带宽占
20 用时间段，判断所述待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段。

在一些实施例中，所述预设网络拥塞控制算法为数据中心量化拥塞通知 DCQCN 算法。

在一些实施例中，基于预先构建的所述分布式系统的网络流量模型和所述目标模型训练任务，调用相应的网络拥塞控制参数对所述分布式系统进行网络拥塞控制，包括：获取第一执行优先级和第二执行优先级，其中，所述第一执行优先级为目标模型训练任务的任
25 务执行优先级，所述第二执行优先级为已有模型训练任务的任务执行优先级；根据第一执行优先级和第二执行优先级，确定所述目标模型训练和所述已有模型训练任务的执行顺序，并调用相应的网络拥塞控制参数对所述分布式系统进行网络拥塞控制，以使所述任务执行优先级较高的模型训练任务优先执行，所述任务执行优先级较高的模型训练任务推迟执行。

在一些实施例中，所述预先构建的所述分布式系统的网络流量模型为根据分布式系统中一个或多个已有的周期性峰值网络带宽占用时间段的模型训练任务构建的网络流量模型。
30

在一些实施例中，所述调用相应的网络拥塞控制参数对所述分布式系统进行网络拥塞控制，以使目标模型训练任务的峰值网络带宽占用时间段与所述分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同，包括：所述调用相应的网络拥塞控制参数对所述
35 分布式系统进行网络拥塞控制，以使目标模型训练任务的峰值网络带宽占用时间段与所述

分布式系统内任意一个已有模型训练任务的峰值网络带宽占用时间段均不同。

根据本公开的另一个方面，还提供了一种分布式模型训练的网络拥塞控制装置，包括：训练任务监测模块，设置为监测到分布式系统中待执行的目标模型训练任务，其中，所述目标模型训练任务为存在周期性峰值网络带宽占用时间段的模型训练任务，所述峰值网络带宽占用时间段为模型训练任务执行过程中网络带宽占用高于预设阈值的任务时间段；网络拥塞控制模块，设置为基于预先构建的所述分布式系统的网络流量模型和所述目标模型训练任务，调用相应的网络拥塞控制参数对所述分布式系统进行网络拥塞控制，以使所述目标模型训练任务的峰值网络带宽占用时间段与所述分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。

根据本公开的另一个方面，还提供了一种电子设备，该电子设备包括：处理器；以及存储器，用于存储所述处理器的可执行指令；其中，所述处理器配置为经由执行所述可执行指令来执行上述任意一项所述的分布式模型训练的网络拥塞控制方法。

根据本公开的另一个方面，还提供了一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述任意一项所述的分布式模型训练的网络拥塞控制方法。

根据本公开的另一个方面，还提供了一种计算机程序产品，包括计算机程序，所述计算机程序被处理器执行时实现上述任意一项的分布式模型训练的网络拥塞控制方法。

本公开的实施例中提供的分布式模型训练的网络拥塞控制方法、装置、设备、介质及计算机程序产品，在监测到分布式系统中存在周期性峰值网络带宽占用时间段的待执行的目标模型训练任务后，结合预先构建的网络流量模型，调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制，以使目标模型训练任务与分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。本公开实施例能够通过将模型训练任务的网络拥塞参数根据用户需求进行自动调整，降低网络拥塞情况发生的概率，提高网络利用率。

应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，并不能限制本公开。

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，示出了符合本公开的实施例，并与说明书一起用于解释本公开的原理。显而易见地，下面描述中的附图仅仅是本公开的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1示出了本公开实施例中分布式模型训练的网络拥塞控制方法的示例性应用系统架构示意图；

图2示出了本公开实施例中一种网络感知系统的示意图；

图3示出本公开实施例中一种分布式模型训练的网络拥塞控制方法流程图；

图 4 示出本公开实施例中另一种分布式模型训练的网络拥塞控制方法流程图；

图 5 示出本公开实施例中一种分布式模型训练的网络拥塞控制装置示意图；

图 6 示出本公开实施例中一种电子设备的框图；

图 7 示出本公开实施例中一种计算机可读存储介质示意图。

5

具体实施方式

现在将参考附图更全面地描述示例实施方式。然而，示例实施方式能够以多种形式实施，且不应被理解为限于在此阐述的范例；相反，提供这些实施方式使得本公开将更加全面和完整，并将示例实施方式的构思全面地传达给本领域的技术人员。所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。

此外，附图仅为本公开的示意性图解，并非一定是按比例绘制。图中相同的附图标记表示相同或类似的部分，因而将省略对它们的重复描述。附图中所示的一些方框图是功能实体，不一定必须与物理或逻辑上独立的实体相对应。可以采用软件形式来实现这些功能实体，或在一个或多个硬件模块或集成电路中实现这些功能实体，或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

为便于理解，在介绍本公开实施例之前，首先对本公开实施例中涉及到的几个名词进行解释如下：

RDMA: Remote Direct Memory Access，远程直接内存访问；

PFC: Priority Flow Control，优先级流量控制，是一种流控制技术，它在以太网网络中用于管理数据流的优先级；

DCQCN: Data Center Quality of Service Congestion Notification，数据中心服务质量拥塞通知，是一种在数据中心网络中用于拥塞通知和质量服务的技术。

下面结合附图，对本公开实施例的具体实施方式进行详细说明。

图 1 示出了本公开实施例中分布式模型训练的网络拥塞控制方法的示例性应用系统架构示意图。如图 1 所示，相关技术中，该系统架构可以包括脊设备 101、叶设备 102、服务器 103 和客户端 104。其中，脊设备 101 可分别与第一叶设备 1021 和第二叶设备 1022 相连，第一叶设备可分别与第一服务器 1031、第二服务器 1032、第三服务器 1033 和第四服务器 1034 相连。此处以第一服务器 1031 举例，第一服务器 1031 与客户端 104 相连。

相关技术中，当前基于集群的分布式模型训练系统通过支持不同的框架（例如张量流 TensorFlow 等）进行端到端的模型训练，此时，模型训练任务承载在集群的多个节点上，分布式系统将模型训练任务切分成更小的批次，在各自节点进行计算，再通过同步机制更新统一的模型权重。在人脸识别的场景下，模型训练任务作业时可以选择多个服务器上的多个图形处理器 GPU 组合进行。在常用的分布式框架训练期间，模型每经历一个迭代或固定周期，多个服务器之间通过网络进行模型权重的同步，多

个服务器通常会遇到共享网络链路进行权重更新的情况。

在基于 RDMA 的分布式系统中，对于网络的拥塞控制使用公平策略，但在特殊场景（可例如大模型训练场景）的需求下，基于 PFC 的端口拥塞控制和基于流的 DCQCN 拥塞控制机制都有调节机制，能完成对特定任务（可例如大模型训练任务）

5 网络带宽进行设定。

因此，在本公开的一个实施例中，在相关技术的网络架构下，增加了网络感知系统 105。具体地，图 2 示出了本公开实施例中一种网络感知系统的示意图。该系统 105 包括：流量模型构建模块 1051、网络拥塞感知模块 1052 和参数记忆模块 1053。

10 在本公开的一个实施例中，流量模型构建模块 1051 拥有两个用户接口，一个接口面向当前的分布式系统，主要是进行模型训练任务的系统拓扑感知和网络感知，记录模型训练任务各节点的带宽敏感度；另一个接口面向模型训练任务，对模型训练任务的一部分超参数进行预设区域内的搜索适配，例如，数据并行的架构中，对模型训练任务的批量大小、工作节点等影像数据流量的超参数进行离散值配置。经过适配不同的超参数组合，构建系统中初始模型训练任务的流量模型表，用户可以参考流量模型表，来干预实际的模型训练计划，选择适合在当前系统拓扑下最合适的参数配置，
15 来完成初始流量模型的训练。

具体地，流量模型表可以展示出不同参数下的模型训练任务的流量情况，具体的，可如表 1 所示：

表 1

模型参数组	网络吞吐量	任务时延
W1	TH1	LT1
W2	TH2	LT2

20 其中，W1 可以为系统中默认设置的多个超参数构成的模型参数组，TH1 和 LT1 可以为基于当前默认模型参数组计算得到的默认网络吞吐量和默认任务时延；W2 可以为初始模型训练任务的多个超参数构成的模型参数组，TH2 和 LT2 可以为基于当前初始模型训练任务的模型参数组计算得到的网络吞吐量和任务时延。比较两个模型参数组带来的效果，选择更优的数据来完成初始流量模型训练。

25 需要说明的是，上述预设区域是根据训练初始流量模型的超参数定制化选择搜索范围确定的，可根据实际情况进行调整，本公开实施例对此不做具体限定。

在本公开的一个实施例中，网络拥塞感知模块 1052 用来在模型训练任务执行阶段进行网络通信流量的感知，其中包括在启动时对系统中已有的模拟训练任务进行流量模型构建（即初始流量模型）的分析感知阶段，在模型训练任务执行期间对新接入的网络流量（即新模型训练任务）的抽样感知阶段，以及对模型训练任务的网络拥塞控制参数进行调整的固有感知阶段。
30

在本公开的一个实施例中，分析感知阶段指在系统中存在一项模型训练任务的场景下，对该模型训练任务的网络使用率进行感知，构建当前模型训练任务的网络流量模型，作为当前系统的初始流量模型，具体地，可通过一预设网络拥塞控制算法，感知周期性的数据流动，并忽略突发性的数据流动。上述预设网络拥塞控制算法可以指 DCQCN 算法，即通过分析算法中显式拥塞通知 ECN 标记数量来评估网络链路中的流量，需要说明的是，可以采用任意可感知到模型训练任务周期性的数据流动的网络拥塞控制算法，本公开对此不做具体限定。

在本公开的一个实施例中，获取初始流量模型后，开始进入抽样感知阶段。当存在新的模型训练任务在系统中运行时，通过一预设网络拥塞控制算法对网络流量的异常拥塞感知，确定当前模型训练任务是否存在周期性的数据流量（即周期性峰值网络带宽占用时间段）。若存在，则系统返回固有感知阶段，同时重新感知系统中是否还存在新模型训练任务接入。

在本公开的一个实施例中，固有感知阶段用于启动拥塞参数搜索功能，例如，当前系统中存在两个模型训练任务，通过调整一个某个节点的 RDMA 网络拥塞控制参数，将网络的占用率按一定比例调整为优先级第一项训练任务，使得系统的网络资源在发生拥塞时，更偏向于完成第一个训练任务。调整后，经过几轮训练迭代，网络拥塞感知模块判断系统是否主动推迟第二项训练任务进入占用网络高带宽的时间点（即是否找到适合两个模型训练任务的并发周期性）。当网络拥塞感知阶段识别到系统的拥塞控制达到当前情况的最佳状态时，该系统达到稳定，预计模型训练任务的完成时间最短，两项任务对共享链路上的网络占用率，在高占用需求阶段性能都会提升，类似将两个周期性的正弦波通过相位移动，规避波峰叠加的区域，降低整体网络带宽的压力，同时提高各自占用的网络带宽。

需要说明的是，模型训练任务的优先级可为用户预先设定的任务执行优先级，需要说明的是，本公开实施例对此不做具体限定。

在本公开的一个实施例中，参数记忆模块 1053 拥有两个接口，一个接口面向模型训练任务，另一个接口面向该模型训练任务执行期间的网络拥塞控制参数。该模块用于注册和存储模型训练任务的超参数、网络拥塞控制参数和当前系统中的网络流量模型，还用于提供和配置模型训练任务的超参数、网络拥塞控制参数的初始化值。该模块还可以开放给用户，方便用户在个性化的系统中设定参数信息，提高了系统的灵活性。

需要说明的是，上述初始化值在实际应用中通常由系统提供默认值，也可以由用户进行自主配置，本公开实施例对此不做具体限定。

在本公开的一个实施例中，该系统可为用户提供分析应用业务和系统的网络流量的功能，对模型训练任务进行的网络流量信息识别，并不特别的指定模型训练作业的系统拓扑结构，可以帮助用户在不同系统中进行网络的最优化配置。同时还结合业务

作业和网络性能参数，综合考虑整体的训练精度和训练效率，为用户在使用系统时提供训练作业和网络配置双重考虑下的推荐参数，提供两个维度的训练作业参数模型。

本领域技术人员可以知晓，图 1 中的脊设备、叶设备、服务器和客户端的数量仅仅是示意性的，根据实际需要，可以具有任意数目的脊设备、叶设备、服务器和客户端。本公开实施例对此不作限定。

在上述系统架构下，本公开实施例中提供了一种分布式模型训练的网络拥塞控制方法，该方法可以由任意具备计算处理能力的电子设备执行。

图 3 示出本公开实施例中一种分布式模型训练的网络拥塞控制方法流程图，如图 3 所示，该方法包括如下步骤：

S302，监测到分布式系统中待执行的目标模型训练任务，其中，目标模型训练任务为存在周期性峰值网络带宽占用时间段的模型训练任务，峰值网络带宽占用时间段为模型训练任务执行过程中网络带宽占用高于预设阈值的任务时间段。

在本公开的一个实施例中，一个或多个新接入分布式系统的模型训练任务中，存在周期性峰值网络带宽占用时间段的模型训练任务被确定为目标模型训练任务，其中，峰值网络带宽占用时间段可以指模型训练任务在任务执行期间，该模型训练任务的网络带宽占用高于预设阈值的任务时间段。需要说明的是，预设阈值可由用户根据实际情况进行设定，本公开实施例对此不做具体限定。

S304，基于预先构建的分布式系统的网络流量模型和目标模型训练任务，调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制，以使目标模型训练任务的峰值网络带宽占用时间段与分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。

在本公开的一个实施例中，预先构建的分布式系统的网络流量模型可以是根据分布式系统中一个或多个已有的周期性峰值网络带宽占用时间段的模型训练任务构建的网络流量模型；通过调用目标模型训练任务中相应的网络拥塞控制参数来调整该模型训练任务的峰值网络带宽占用时间段，使其与分布式系统内已有模型训练任务的峰值网络带宽占用时间段区别开。

由上述可知，本公开实施例在监测到分布式系统中存在周期性峰值网络带宽占用时间段的待执行的目标模型训练任务后，结合预先构建的网络流量模型，调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制，以使目标模型训练任务与分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。本公开实施例能够通过将模型训练任务的网络拥塞参数根据用户需求进行自动调整，降低网络拥塞情况发生的概率，提高网络利用率。

在本公开的一个实施例中，上述 S302，包括监测分布式系统是否接收到待执行的模型训练任务；若监测分布式系统接收到待执行的模型训练任务，则判断待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段；若待执行的模型训练任务存在

周期性峰值网络带宽占用时间段，则将待执行的模型训练任务确定为目标模型训练任务。

在本公开的一个实施例中，可通过 DCQCN 算法判断待执行的模型训练任务中是否存在周期性峰值网络带宽占用时间段，需要说明的是，本公开实施例还可采用其他任意可判断模型训练任务是否具有周期性峰值网络带宽占用时间段的网络拥塞控制算法，本公开实施例对此不做具体限定。

在本公开的一个实施例中，判断待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段，包括：根据预设网络拥塞控制算法，确定待执行的模型训练任务在模型训练任务执行过程中峰值网络带宽占用时间段；根据待执行的模型训练任务的峰值网络带宽占用时间段，判断待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段。

在本公开的一个实施例中，预设网络拥塞控制算法为数据中心量化拥塞通知 DCQCN 算法。

在本公开的一个实施例中，上述 S304，包括获取第一执行优先级和第二执行优先级，其中，第一执行优先级为目标模型训练任务的任务执行优先级，第二执行优先级为已有模型训练任务的任务执行优先级；根据第一执行优先级和第二执行优先级，确定目标模型训练和已有模型训练任务的执行顺序，并调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制，以使任务执行优先级较高的模型训练任务优先执行，任务执行优先级较高的模型训练任务推迟执行。

需要说明的是，第一执行优先级和第二执行优先级均可由用户预先进行设定，本公开实施例对此不做具体限定。

在本公开的一个实施例中，预先构建的分布式系统的网络流量模型为根据分布式系统中一个或多个已有的周期性峰值网络带宽占用时间段的模型训练任务构建的网络流量模型。

在本公开的一个实施例中，可通过 DCQCN 判断已有的模型训练任务是否存在周期性峰值网络带宽占用时间段，需要说明的是，本公开实施例还可采用其他任意可判断模型训练任务是否具有周期性峰值网络带宽占用时间段的网络拥塞控制算法，本公开实施例对此不做具体限定。

在本公开的一个实施例中，上述 S304，包括：调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制，以使目标模型训练任务的峰值网络带宽占用时间段与分布式系统内任意一个已有模型训练任务的峰值网络带宽占用时间段均不同。

在本公开的一个实施例中，还可在多用户的云计算业务场景中，基于不同业务，将可以结合最大化网络性能的业务虚拟化到相近的硬件实现上，给用户提供有效的网络配置和系统拓扑建议。

在本公开的一个实施例中，在云服务商提供相同资源的条件下，为用户提供更高

效的服务，同时也能帮助和指导用户对资源的利用有更多的可控性，提升用户认知体验。还可以帮助云服务商提升用户资源的配置效率，云服务商可以根据用户的需求，将可以互补的资源提供给多用户，节省资源成本。

图 4 示出本公开实施例中另一种分布式模型训练的网络拥塞控制方法流程图，如图 4 所示，该方法包括如下步骤：

S401，在系统启动时，在流量模型构建模块注册训练任务，对初始网络流量模型进行构建。

在本公开的一个实施例中，对分布式系统中接收到的初始模型训练任务进行判断，基于存在周期性峰值网络带宽占用时间段的模型训练任务训练初始网络流量模型，即对存在周期性峰值网络带宽占用时间段的模型训练任务，在面向用户的接口进行注册，而对仅存在突发性峰值网络带宽占用的模型训练任务不进行注册。

在本公开的一个实施例中，可通过 DCQCN 算法判断已有的模型训练任务是否存在周期性峰值网络带宽占用时间段，需要说明的是，本公开实施例还可采用其他任意可判断模型训练任务是否具有周期性峰值网络带宽占用时间段的网络拥塞控制算法，本公开实施例对此不做具体限定。

S402，在初始网络流量模型构建完成后，开始进入抽样感知阶段。

S403，判断是否存在新模型训练任务接入当前分布式系统。若是，则执行 S404；若否，则执行 S410。

在本公开的一个实施例中，可通过 DCQCN 判断当前分布式系统中网络流量执行周期是否出现变化，来判断系统中是否接入新模型训练任务，需要说明的是，本公开实施例还可采用其他任意可判断模型训练任务是否具有周期性峰值网络带宽占用时间段的网络拥塞控制算法，本公开实施例对此不做具体限定。

S404，进行网络拥塞感知，启动拥塞感知功能。

S405，判断接收到的新模型训练任务中是否存在周期性峰值网络带宽占用时间段。若是，则执行 S406；若否，则标记该任务，后续不做处理，并执行 S407。

在本公开的一个实施例中，可通过 DCQCN 判断待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段，需要说明的是，本公开实施例还可采用其他任意可判断模型训练任务是否具有周期性峰值网络带宽占用时间段的网络拥塞控制算法，本公开实施例对此不做具体限定。

S406，进行参数搜索及配置，并对分布式系统进行调优。

在本公开的一个实施例中，按照用户预先设置的任务执行优先级，调整模型训练任务的网络拥塞控制参数，以使模型训练任务的峰值网络带宽占用时间段与分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。保持当前网络拥塞控制参数不变，对模型训练任务中的超参数进行搜索并配置，使得分布式系统达到最优状态。

S407，对当前参数进行存储。

S408, 判断当前系统是否达到最优。若是, 则执行 S409; 若否, 则执行 S406。

在本公开的一个实施例中, 当多个模型训练任务在系统中同时执行时, 系统进入稳定周期性的运行, 预计模型训练任务的完成时间最短时, 则证明当前分布式系统已达到性能最优。

5 S409, 保持当前参数配置。

S410, 判断当前系统是否同时满足系统结束以及当前系统中注册的新任务小于 1。若是, 则执行 S411; 若否, 则执行 S402。

S411, 向用户输出模型训练报告。

10 当系统结束且系统中注册的模型训练任务小于 1 时, 模块完成本阶段的拥塞感知任务, 在模型训练期间对超参数的搜索、对网络拥塞参数的调整和最终训练得到的网络流量模型进行整理分析, 将上述数据作为模型训练报告发送给用户并提出合理建议。

15 由上述可知, 本公开实施例在监测到分布式系统中存在周期性峰值网络带宽占用时间段的待执行的目标模型训练任务后, 结合预先构建的网络流量模型, 调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制, 以使目标模型训练任务与分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。可以进行对模型训练任务的标记和网络监控, 在同时执行多个训练任务时, 有侧重性的调节网络优先级, 在不影响训练业务准确率的情况下, 对并发训练任务进行网络流量模型的识别, 通过定制化的拥塞控制, 即根据系统的推荐隐性使系统进入高利用率模式, 达成网络利用率的最大化。

20 在本公开的一个实施例中, 当存在新模型训练任务接入当前分布式系统时, 在开始进行网络拥塞情况感知的同时, 可同时触发系统中新模型训练任务的监测任务, 此时, 仍继续监测系统中是否存在新模型训练任务。

在本公开的一个实施例中, 可通过如下公式进行网络拥塞感知:

$$R_t = R_o \times (1 + g) \quad (1)$$

25 其中, R_t 表示拥塞感知时网络拥塞参数组, R_o 表示拥塞感知前的拥塞参数组, g 表示拥塞控制带宽的变化率, $g \in [-1, 1]$ 。

30 其中, g 值越远离 0, 变化率越高, 回到高速带宽的速度越快, 需要说明的是, g 值同时应考虑到整个系统变化的鲁棒性; 上述网络拥塞参数组为用于调整模型训练任务峰值网络带宽占用时间段的一系列网络拥塞参数的总称, 可例如网络带宽等, 在实际模型训练中涉及的网络拥塞参数均可加入到该网络拥塞参数组中, 本公开实施例对网络拥塞参数组中包含的参数类型不做具体限定。

在本公开的一个实施例中, 可通过如下公式体现参数搜索前后任务之间的时间差:

$$\Delta T = T \times (P_{train_n} + P_{net_n}) - T \times (P_{train_n_nonsense} + P_{net_n_nonsense}) \quad (2)$$

35 其中, ΔT 表示参数搜索配置后模型训练任务完成时间与原始模型训练任务完成时间的的时间差, $T \times (\cdot)$ 表示当前参数配置下的模型训练任务的完成时间, P_{train_n} 表示参

数搜索后模型训练任务的超参数综合值, P_{net_n} 表示参数搜索后模型训练任务的网络拥塞参数综合值, $P_{train_n_nonsense}$ 表示参数搜索前模型训练的超参数综合值, $P_{net_n_nonsense}$ 表示参数搜索前模型训练的网络拥塞参数综合值。

需要说明的是, 上述 ΔT 、 T 、 P_{train_n} 、 P_{net_n} 、 $P_{train_n_nonsense}$ 和 $P_{net_n_nonsense}$ 均为一个常数; 上述网络拥塞参数可以是当前数据采集时刻下由 R_t 、 R_o 和 g 构成的网络拥塞参数的综合值, 为一个常数; 超参数用于使当前模型训练后输出结果的效果最优, 可例如批量大小、工作节点等, 上述超参数可以是当前数据采集时刻下基于当前网络拥塞参数进行调节得到的一个或多个超参数构成的超参数组的综合值, 为一个常数。

基于同一发明构思, 本公开实施例中还提供了一种分布式模型训练的网络拥塞控制装置, 如下面的实施例所述。由于该装置实施例解决问题的原理与上述方法实施例相似, 因此该装置实施例的实施可以参见上述方法实施例的实施, 重复之处不再赘述。

图 5 示出本公开实施例中一种分布式模型训练的网络拥塞控制装置示意图, 如图 5 所示, 该装置包括: 训练任务监测模块 501 和网络拥塞控制模块 502。

其中, 训练任务监测模块 501, 设置为监测到分布式系统中待执行的目标模型训练任务, 其中, 目标模型训练任务为存在周期性峰值网络带宽占用时间段的模型训练任务, 峰值网络带宽占用时间段为模型训练任务执行过程中网络带宽占用高于预设阈值的任务时间段; 网络拥塞控制模块 502, 设置为基于预先构建的分布式系统的网络流量模型和目标模型训练任务, 调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制, 以使目标模型训练任务的峰值网络带宽占用时间段与分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。

由上述可知, 本公开实施例在监测到分布式系统中存在周期性峰值网络带宽占用时间段的待执行的目标模型训练任务后, 结合预先构建的网络流量模型, 调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制, 以使目标模型训练任务与分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。本公开实施例能够通过将模型训练任务的网络拥塞参数根据用户需求进行自动调整, 降低网络拥塞情况发生的概率, 提高网络利用率。

在本公开的一个实施例中, 上述训练任务监测模块 501, 还设置为监测分布式系统是否接收到待执行的模型训练任务; 若监测分布式系统接收到待执行的模型训练任务, 则判断待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段; 若待执行的模型训练任务存在周期性峰值网络带宽占用时间段, 则将待执行的模型训练任务确定为目标模型训练任务。

在本公开的一个实施例中, 上述训练任务监测模块 501, 还设置为根据预设网络拥塞控制算法, 确定待执行的模型训练任务在模型训练任务执行过程中峰值网络带宽占用时间段; 根据待执行的模型训练任务的峰值网络带宽占用时间段, 判断待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段。

在本公开的一个实施例中，预设网络拥塞控制算法为数据中心量化拥塞通知 DCQCN 算法。

在本公开的一个实施例中，上述网络拥塞控制模块 502，还设置为获取第一执行优先级和第二执行优先级，其中，第一执行优先级为目标模型训练任务的任务执行优先级，第二执行优先级为已有模型训练任务的任务执行优先级；根据第一执行优先级和第二执行优先级，确定目标模型训练和已有模型训练任务的执行顺序，并调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制，以使任务执行优先级较高的模型训练任务优先执行，任务执行优先级较高的模型训练任务推迟执行。

在本公开的一个实施例中，预先构建的分布式系统的网络流量模型为根据分布式系统中一个或多个已有的周期性峰值网络带宽占用时间段的模型训练任务构建的网络流量模型。

在本公开的一个实施例中，上述网络拥塞控制模块 502，还设置为调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制，以使目标模型训练任务的峰值网络带宽占用时间段与分布式系统内任意一个已有模型训练任务的峰值网络带宽占用时间段均不同。

所属技术领域的技术人员能够理解，本公开的各个方面可以实现为系统、方法或程序产品。因此，本公开的各个方面可以具体实现为以下形式，即：完全的硬件实施方式、完全的软件实施方式（包括固件、微代码等），或硬件和软件方面结合的实施方式，这里可以统称为“电路”、“模块”或“系统”。

下面参照图 6 来描述根据本公开的这种实施方式的电子设备 600。图 6 显示的电子设备 600 仅仅是一个示例，不应对本公开实施例的功能和使用范围带来任何限制。

图 6 示出本公开实施例中一种电子设备的框图。下面参照图 6 来描述根据本公开的这种实施方式的电子设备 600。图 6 显示的电子设备 600 仅仅是一个示例，不应对本公开实施例的功能和使用范围带来任何限制。

如图 6 所示，电子设备 600 以通用计算设备的形式表现。电子设备 600 的组件可以包括但不限于：上述至少一个处理单元 610、上述至少一个存储单元 620、连接不同系统组件（包括存储单元 620 和处理单元 610）的总线 630。

其中，所述存储单元存储有程序代码，所述程序代码可以被所述处理单元 610 执行，使得所述处理单元 610 执行本说明书上述“示例性方法”部分中描述的根据本公开各种示例性实施方式的步骤。例如，所述处理单元 610 可以执行上述方法实施例的如下步骤：监测到分布式系统中待执行的目标模型训练任务，其中，目标模型训练任务为存在周期性峰值网络带宽占用时间段的模型训练任务，峰值网络带宽占用时间段为模型训练任务执行过程中网络带宽占用高于预设阈值的任务时间段；基于预先构建的分布式系统的网络流量模型和目标模型训练任务，调用相应的网络拥塞控制参数对分布式系统进行网络拥塞控制，以使目标模型训练任务的峰值网络带宽占用时间段与

分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。

存储单元 620 可以包括易失性存储单元形式的可读介质，例如随机存取存储单元 (RAM) 6201 和/或高速缓存存储单元 6202，还可以进一步包括只读存储单元 (ROM) 6203。

5 存储单元 620 还可以包括具有一组（至少一个）程序模块 6205 的程序/实用工具 6204，这样的程序模块 6205 包括但不限于：操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。

10 总线 630 可以为表示几类总线结构中的一种或多种，包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

电子设备 600 也可以与一个或多个外部设备 640（例如键盘、指向设备、蓝牙设备等）通信，还可与一个或者多个使得用户能与该电子设备 600 交互的设备通信，和/或与使得该电子设备 600 能与一个或多个其它计算设备进行通信的任何设备（例如路由
15 器、调制解调器等等）通信。这种通信可以通过输入/输出 (I/O) 接口 650 进行。并且，电子设备 600 还可以通过网络适配器 660 与一个或者多个网络（例如局域网 (LAN)，广域网 (WAN) 和/或公共网络，例如因特网）通信。如图所示，网络适配器 660 通过总线 630 与电子设备 600 的其它模块通信。应当明白，尽管图中未示出，可以结合电子设备 600 使用其它硬件和/或软件模块，包括但不限于：微代码、设备驱
20 动器、冗余处理单元、外部磁盘驱动阵列、RAID 系统、磁带驱动器以及数据备份存储系统等。

通过以上的实施方式的描述，本领域的技术人员易于理解，这里描述的示例实施方式可以通过软件实现，也可以通过软件结合必要的硬件的方式来实现。因此，根据本公开实施方式的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质（可以是 CD-ROM，U 盘，移动硬盘等）中或网络上，包括若干指令以使得一台计算设备（可以是个人计算机、服务器、终端装置、或者网络设备等）执行根据本公开实施方式的方法。

特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机程序产品，该计算机程序产品包括：计算机程序，所述计算机程序被处理器执行时实现上述分布式模型训练的网络拥塞控制方法。
30

在本公开的示例性实施例中，还提供了一种计算机可读存储介质，该计算机可读存储介质可以是可读信号介质或者可读存储介质。图 7 示出本公开实施例中一种计算机可读存储介质示意图，如图 7 所示，该计算机可读存储介质上存储有能够实现本公开上述方法的程序产品 700。在一些可能的实施方式中，本公开的各个方面还可以实现为一种程序产品的形式，其包括程序代码，当所述程序产品在终端设备上运行时，
35

所述程序代码用于使所述终端设备执行本说明书上述“示例性方法”部分中描述的根据本公开各种示例性实施方式的步骤。

本公开中的计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。

在本公开中，计算机可读存储介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了可读程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。可读信号介质还可以是可读存储介质以外的任何可读介质，该可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

在一些实施例中，计算机可读存储介质上包含的程序代码可以用任何适当的介质传输，包括但不限于无线、有线、光缆、RF 等等，或者上述的任意合适的组合。

在具体实施时，可以以一种或多种程序设计语言的任意组合来编写用于执行本公开操作的程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如 Java、C++ 等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中，远程计算设备可以通过任意种类的网络，包括局域网（LAN）或广域网（WAN），连接到用户计算设备，或者，可以连接到外部计算设备（例如利用因特网服务提供商来通过因特网连接）。

应当注意，尽管在上文详细描述中提及了用于动作执行的设备的若干模块或者单元，但是这种划分并非强制性的。实际上，根据本公开的实施方式，上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之，上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

此外，尽管在附图中以特定顺序描述了本公开中方法的各个步骤，但是，这并非要求或者暗示必须按照该特定顺序来执行这些步骤，或是必须执行全部所示的步骤才能实现期望的结果。附加的或备选的，可以省略某些步骤，将多个步骤合并为一个步骤执行，以及/或者将一个步骤分解为多个步骤执行等。

通过以上实施方式的描述，本领域的技术人员易于理解，这里描述的示例实施方式可以通过软件实现，也可以通过软件结合必要的硬件的方式来实现。因此，根据本公开实施方式的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质（可以是 CD-ROM，U 盘，移动硬盘等）中或网络上，包括若干指令以使得一台计算设备（可以是个人计算机、服务器、移动终端、或者网络设备等）

执行根据本公开实施方式的方法。

- 5 本领域技术人员在考虑说明书及实践这里公开的发明后，将容易想到本公开的其它实施方案。本公开旨在涵盖本公开的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的，本公开的真正范围和精神由所附的权利要求指出。

权利要求

1. 一种分布式模型训练的网络拥塞控制方法，包括：

监测到分布式系统中待执行的目标模型训练任务，其中，所述目标模型训练任务为存在周期性峰值网络带宽占用时间段的模型训练任务，所述峰值网络带宽占用时间段为模型训练任务执行过程中网络带宽占用高于预设阈值的任务时间段；

基于预先构建的所述分布式系统的网络流量模型和所述目标模型训练任务，调用相应的网络拥塞控制参数对所述分布式系统进行网络拥塞控制，以使所述目标模型训练任务的峰值网络带宽占用时间段与所述分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。

2. 根据权利要求 1 所述的分布式模型训练的网络拥塞控制方法，其中，所述监测到分布式系统中待执行的目标模型训练任务，包括：

监测分布式系统是否接收到待执行的模型训练任务；

若所述监测分布式系统接收到所述待执行的模型训练任务，则判断所述待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段；

若所述待执行的模型训练任务存在周期性峰值网络带宽占用时间段，则将所述待执行的模型训练任务确定为目标模型训练任务。

3. 根据权利要求 2 所述的分布式模型训练的网络拥塞控制方法，其中，判断所述待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段，包括：

根据预设网络拥塞控制算法，确定待执行的模型训练任务在模型训练任务执行过程中峰值网络带宽占用时间段；

根据所述待执行的模型训练任务的峰值网络带宽占用时间段，判断所述待执行的模型训练任务是否存在周期性峰值网络带宽占用时间段。

4. 根据权利要求 3 所述的分布式模型训练的网络拥塞控制方法，其中，所述预设网络拥塞控制算法为数据中心量化拥塞通知 DCQCN 算法。

5. 根据权利要求 1 所述的分布式模型训练的网络拥塞控制方法，其中，基于预先构建的所述分布式系统的网络流量模型和所述目标模型训练任务，调用相应的网络拥塞控制参数对所述分布式系统进行网络拥塞控制，包括：

获取第一执行优先级和第二执行优先级，其中，所述第一执行优先级为目标模型训练任务的执行优先级，所述第二执行优先级为已有模型训练任务的执行优先级；

根据第一执行优先级和第二执行优先级，确定所述目标模型训练和所述已有模型训练任务的执行顺序，并调用相应的网络拥塞控制参数对所述分布式系统进行网络拥塞控制，以使所述任务执行优先级较高的模型训练任务优先执行，所述任务执行优先级较高的模型训练任务推迟执行。

6. 根据权利要求 1 所述的分布式模型训练的网络拥塞控制方法, 其中, 所述预先构建的所述分布式系统的网络流量模型为根据分布式系统中一个或多个已有的周期性峰值网络带宽占用时间段的模型训练任务构建的网络流量模型。

5 7. 根据权利要求 6 所述的分布式模型训练的网络拥塞控制方法, 其中, 所述调用相应的网络拥塞控制参数对所述分布式系统进行网络拥塞控制, 以使目标模型训练任务的峰值网络带宽占用时间段与所述分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同, 包括:

10 调用相应的网络拥塞控制参数对所述分布式系统进行网络拥塞控制, 以使所述目标模型训练任务的峰值网络带宽占用时间段与所述分布式系统内任意一个已有模型训练任务的峰值网络带宽占用时间段均不同。

8. 一种分布式模型训练的网络拥塞控制装置, 包括:

15 训练任务监测模块, 设置为监测到分布式系统中待执行的目标模型训练任务, 其中, 所述目标模型训练任务为存在周期性峰值网络带宽占用时间段的模型训练任务, 所述峰值网络带宽占用时间段为模型训练任务执行过程中网络带宽占用高于预设阈值的任务时间段;

网络拥塞控制模块, 设置为基于预先构建的所述分布式系统的网络流量模型和所述目标模型训练任务, 调用相应的网络拥塞控制参数对所述分布式系统进行网络拥塞控制, 以使所述目标模型训练任务的峰值网络带宽占用时间段与所述分布式系统内已有模型训练任务的峰值网络带宽占用时间段不同。

20 9. 一种电子设备, 包括:

处理器; 以及

存储器, 用于存储所述处理器的可执行指令;

其中, 所述处理器配置为经由执行所述可执行指令来执行权利要求 1~7 中任意一项所述的分布式模型训练的网络拥塞控制方法。

25 10. 一种计算机可读存储介质, 其上存储有计算机程序, 所述计算机程序被处理器执行时实现权利要求 1~7 中任意一项所述的分布式模型训练的网络拥塞控制方法。

11. 一种计算机程序产品, 包括计算机程序, 所述计算机程序被处理器执行时实现权利要求 1~7 中任意一项所述的分布式模型训练的网络拥塞控制方法。

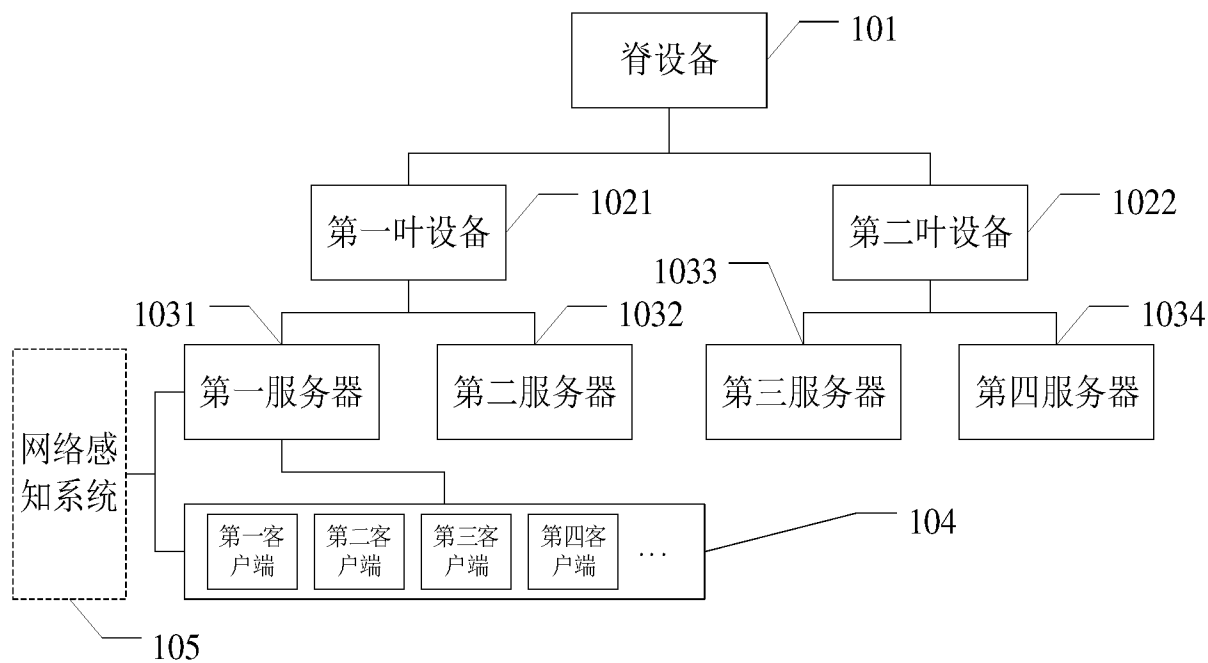


图 1

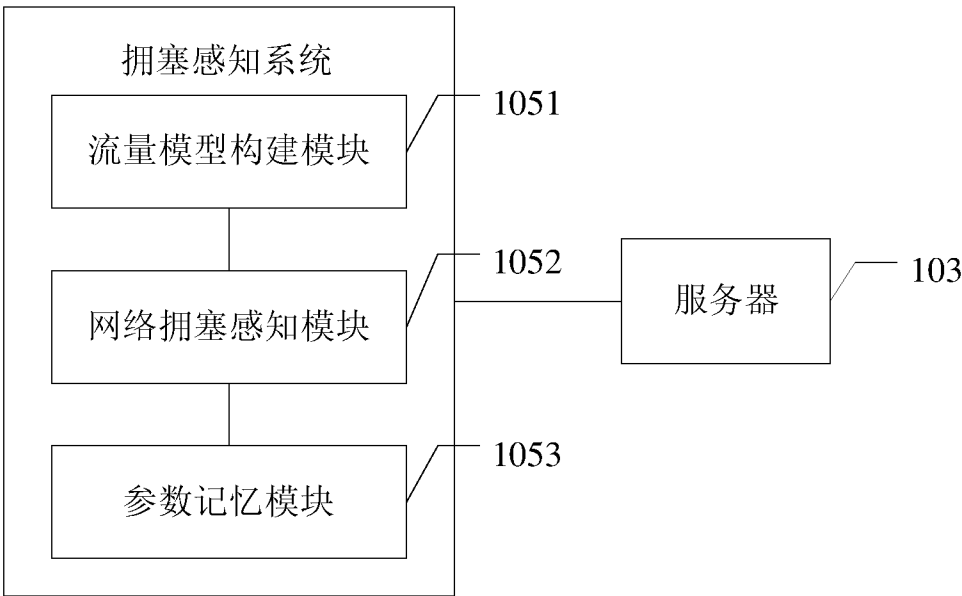


图 2

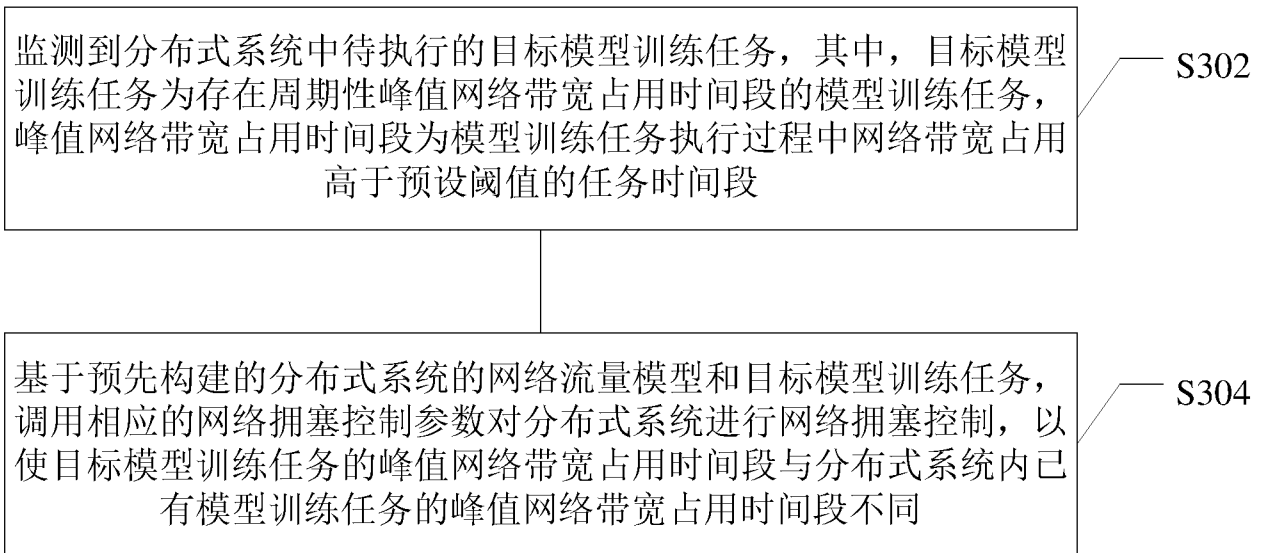


图 3

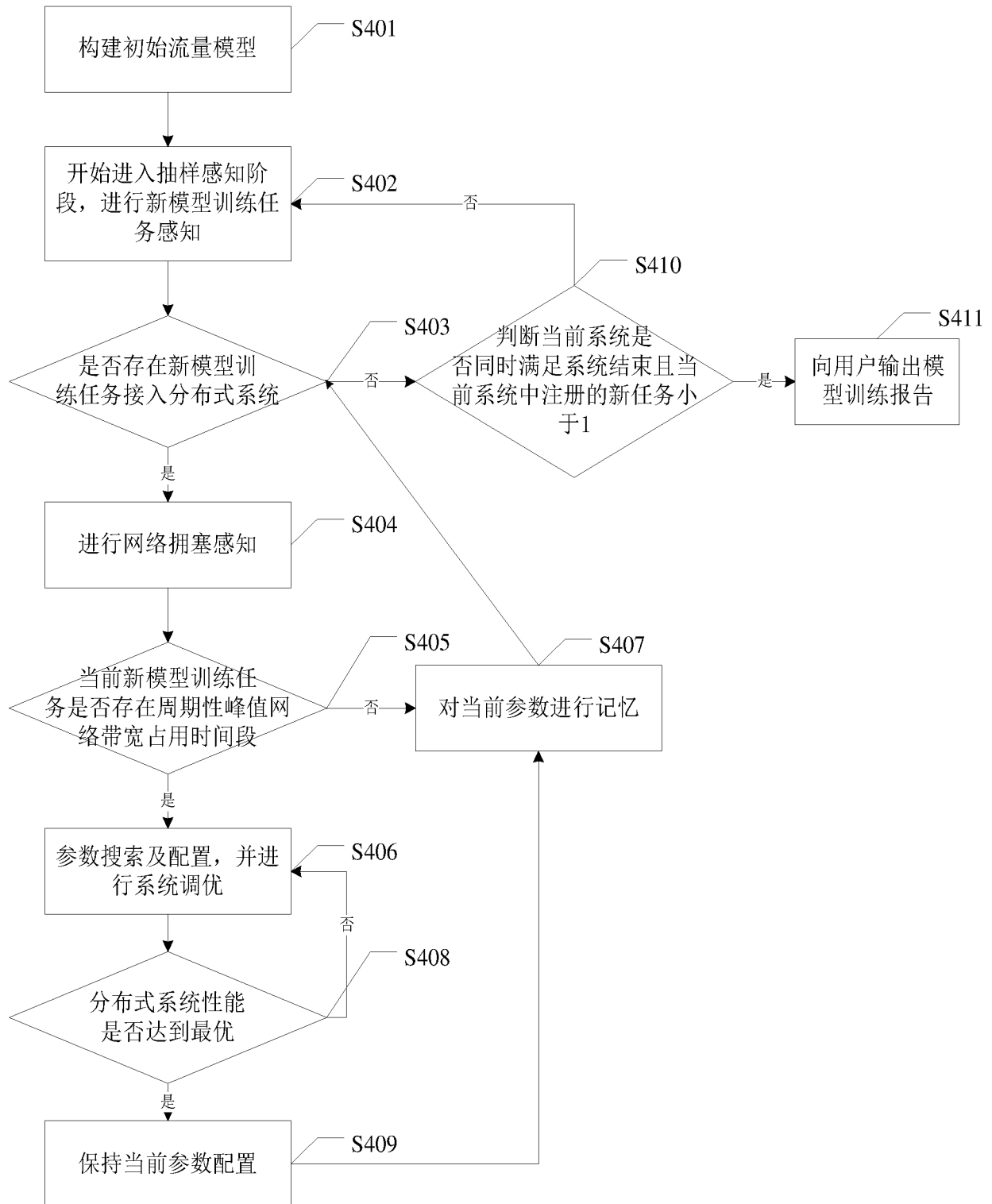


图 4

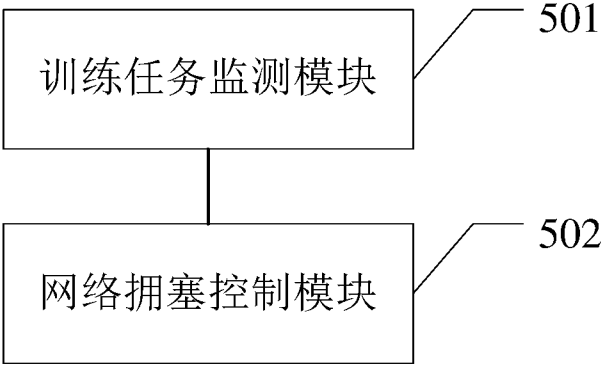


图 5

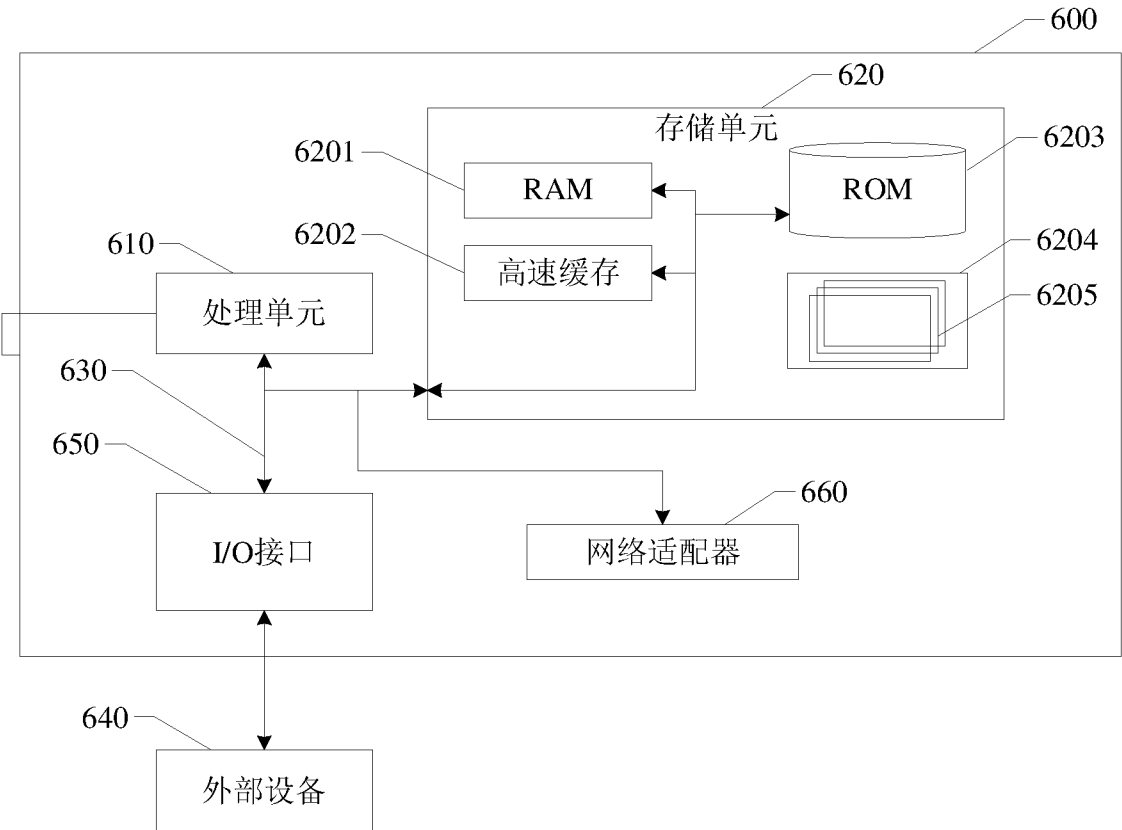


图 6

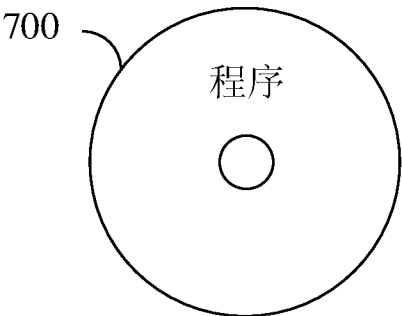


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/136899

A. CLASSIFICATION OF SUBJECT MATTER H04L 47/10(2022.01)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) IPC:H04L 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) CNTXT, CNKI, ENTXTC, DWPI, IEEE: 分布式, 训练任务, 拥塞控制, 周期性, 占用, 时间, 优先级, 顺序, distributed, training task, congestion control, periodic, occupation, time, priority, order		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 114089889 A (JDT HOLDINGS CO., LTD.) 25 February 2022 (2022-02-25) description, paragraphs [0064]-[0156]	1-11
PX	CN 117544565 A (TECHNOLOGY INNOVATION CENTER OF CHINA TELECOM CORP., LTD. et al.) 09 February 2024 (2024-02-09) claims 1-10	1-11
A	CN 111131080 A (UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA) 08 May 2020 (2020-05-08) entire document	1-11
A	CN 115220899 A (DOUYIN VISION CO., LTD. et al.) 21 October 2022 (2022-10-21) entire document	1-11
A	US 2022012642 A1 (INTERNATIONAL BUSINESS MACHINES CORP.) 13 January 2022 (2022-01-13) entire document	1-11
☐ 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 “D” document cited by the applicant in the international application “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 13 March 2025		Date of mailing of the international search report 18 March 2025
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/ CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2024/136899

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	114089889	A	25 February 2022	None			
CN	117544565	A	09 February 2024	None			
CN	111131080	A	08 May 2020	None			
CN	115220899	A	21 October 2022	WO	2024041400	A1	29 February 2024
US	2022012642	A1	13 January 2022	None			

A. 主题的分类

H04L 47/10(2022.01)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)

IPC:H04L

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))

CNTEXT;CNKI;ENTXTC;DWPI;IEEE:分布式, 训练任务, 拥塞控制, 周期性, 占用, 时间, 优先级, 顺序, distributed, training task, congestion control, periodic, occupation, time, priority, order

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
X	CN 114089889 A (京东科技控股股份有限公司) 2022年2月25日 (2022 - 02 - 25) 说明书第[0064]-[0156]段	1-11
PX	CN 117544565 A (中国电信股份有限公司技术创新中心 等) 2024年2月9日 (2024 - 02 - 09) 权利要求1-10	1-11
A	CN 111131080 A (电子科技大学) 2020年5月8日 (2020 - 05 - 08) 全文	1-11
A	CN 115220899 A (抖音视界有限公司 等) 2022年10月21日 (2022 - 10 - 21) 全文	1-11
A	US 2022012642 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2022年1月13日 (2022 - 01 - 13) 全文	1-11

☐ 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

“D” 申请人在国际申请中引证的文件

“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

2025年3月13日

国际检索报告邮寄日期

2025年3月18日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

林桂荣

电话号码 (+86) 010-53961573

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2024/136899

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	114089889	A	2022年2月25日	无	
CN	117544565	A	2024年2月9日	无	
CN	111131080	A	2020年5月8日	无	
CN	115220899	A	2022年10月21日	WO 2024041400 A1	2024年2月29日
US	2022012642	A1	2022年1月13日	无	