



(12) 发明专利申请

(10) 申请公布号 CN 118819846 A

(43) 申请公布日 2024. 10. 22

(21) 申请号 202410903489.7

(22) 申请日 2024.07.05

(71) 申请人 昆仑芯(北京)科技有限公司

地址 100101 北京市海淀区上地十街10号1
幢4层CW区

(72) 发明人 张孝慈 陈庆澍 王勇 欧阳剑

(74) 专利代理机构 北京市汉坤律师事务所

11602

专利代理师 姜浩然 吴丽丽

(51) Int. Cl.

G06F 9/50 (2006.01)

G06F 9/48 (2006.01)

G06F 16/27 (2019.01)

G06N 3/0455 (2023.01)

权利要求书2页 说明书11页 附图5页

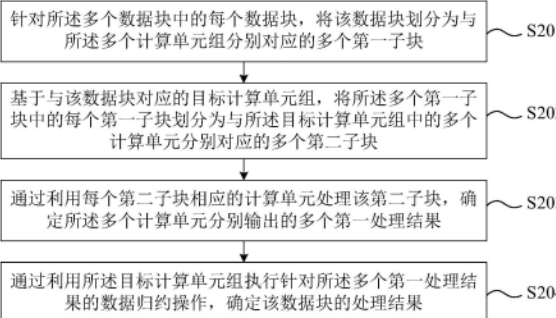
(54) 发明名称

应用于分布式系统的数据处理方法及装置、
设备和介质

(57) 摘要

本公开提供了一种应用于分布式系统的数据处理方法及装置、设备和介质,涉及计算机技术领域,尤其涉及芯片技术和分布式数据处理技术领域。实现方案为:针对待处理数据包括的多个数据块中的每个数据块,将该数据块划分为与
所述多个计算单元组分别对应的多个第一子块;基于与该数据块对应的目标计算单元组,将所述多个第一子块中的每个第一子块划分为与
所述目标计算单元组中的多个计算单元分别对应的多个第二子块;通过利用每个第二子块相应的计算单元处理该第二子块,确定所述多个计算单元
分别输出的多个第一处理结果;以及通过利用所述目标计算单元组执行针对所述多个第一处理结果的数据归约操作,确定该数据块的处理结果。

200



1. 一种应用于分布式系统的数据处理方法,其中,所述分布式系统中的多个计算单元能够构成多个计算单元组,待处理数据包括多个数据块,所述方法包括:

针对所述多个数据块中的每个数据块,

将该数据块划分为与所述多个计算单元组分别对应的多个第一子块;

基于与该数据块对应的目标计算单元组,将所述多个第一子块中的每个第一子块划分为与所述目标计算单元组中的多个计算单元分别对应的多个第二子块;

通过利用每个第二子块相应的计算单元处理该第二子块,确定所述多个计算单元分别输出的多个第一处理结果;以及

通过利用所述目标计算单元组执行针对所述多个第一处理结果的数据归约操作,确定该数据块的处理结果。

2. 如权利要求1所述的方法,其中,所述目标计算单元组中的多个计算单元构成环形拓扑,所述通过利用所述目标计算单元组执行针对所述多个第一处理结果的数据归约操作,确定该数据块的处理结果包括:

基于所述目标计算单元组中的多个计算单元在所述环形拓扑中的排列顺序,依次累加每个计算单元的第一处理结果;以及

基于累加结果确定该数据块的处理结果。

3. 如权利要求2所述的方法,其中,所述基于所述目标计算单元组中的多个计算单元在所述环形拓扑中的排列顺序,依次累加每个计算单元相应的第一处理结果包括:

利用所述多个计算单元中的每个计算单元执行如下步骤:

从所述环形拓扑中位于该计算单元之前的前序计算单元接收待累加数据;

通过基于所述待累加数据和该计算单元的第一处理结果执行累加操作,确定中间数据;以及

将所述中间数据发送至所述环形拓扑中位于该计算单元之后的后继计算单元,以使得所述后继计算单元能够基于所述中间数据和所述后继计算单元的第一处理结果执行累加操作。

4. 如权利要求1-3中任一项所述的方法,其中,所述待处理数据是利用特征编码网络得到的,并且其中,所述待处理数据包括的多个数据块是利用如下的划分操作得到的:基于所述特征编码网络的输出维数,将所述待处理数据划分为大小相同的多个数据块。

5. 如权利要求4所述的方法,其中,基于所述特征编码网络的输出维数,将所述待处理数据划分为大小相同的多个数据块包括:基于所述特征编码网络的输出维数和所述待处理数据的数据格式,将所述待处理数据划分为大小相同的多个数据块。

6. 一种应用于分布式系统的数据处理装置,其中,所述分布式系统中的多个计算单元能够构成多个计算单元组,待处理数据包括多个数据块,所述装置包括:

第一划分单元,被配置为针对所述多个数据块中的每个数据块,将该数据块划分为与所述多个计算单元组分别对应的多个第一子块;

第二划分单元,被配置为基于与该数据块对应的目标计算单元组,将所述多个第一子块中的每个第一子块划分为与所述目标计算单元组中的多个计算单元分别对应的多个第二子块;

第一处理单元,被配置为通过利用每个第二子块相应的计算单元处理该第二子块,确

定所述多个计算单元分别输出的多个第一处理结果;以及

第二处理单元,被配置为通过利用所述目标计算单元组执行针对所述多个第一处理结果的数据归约操作,确定该数据块的处理结果。

7.如权利要求6所述的装置,其中,所述目标计算单元组中的多个计算单元构成环形拓扑,所述第二处理单元包括:

累加子单元,被配置为基于所述目标计算单元组中的多个计算单元在所述环形拓扑中的排列顺序,依次累加每个计算单元的第一处理结果;以及

确定子单元,被配置为基于累加结果确定该数据块的处理结果。

8.如权利要求7所述的装置,其中,所述累加子单元被配置为利用所述多个计算单元中的每个计算单元执行如下步骤:

从所述环形拓扑中位于该计算单元之前的前序计算单元接收待累加数据;

通过基于所述待累加数据和该计算单元的第一处理结果执行累加操作,确定中间数据;以及

将所述中间数据发送至所述环形拓扑中位于该计算单元之后的后继计算单元,以使得所述后继计算单元能够基于所述中间数据和所述后继计算单元的第一处理结果执行累加操作。

9.如权利要求6-8中任一项所述的装置,其中,所述待处理数据是利用特征编码网络得到的,并且其中,所述待处理数据包括的多个数据块是利用如下的划分操作得到的:基于所述特征编码网络的输出维数,将所述待处理数据划分为大小相同的多个数据块。

10.如权利要求9所述的装置,其中,所述待处理数据包括的多个数据块是利用如下的划分操作得到的:基于所述特征编码网络的输出维数和所述待处理数据的数据格式,将所述待处理数据划分为大小相同的多个数据块。

11.一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-5中任一项所述的方法。

12.一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使计算机执行根据权利要求1-5中任一项所述的方法。

13.一种计算机程序产品,包括计算机程序,其中,所述计算机程序在被处理器执行时实现根据权利要求1-5中任一项所述的方法。

应用于分布式系统的数据处理方法及装置、设备和介质

技术领域

[0001] 本公开涉及计算机技术领域,尤其涉及芯片技术和分布式数据处理技术领域,具体涉及一种应用于分布式系统的数据处理方法、装置、电子设备、计算机可读存储介质和计算机程序产品。

背景技术

[0002] 人工智能是研究使计算机来模拟人的某些思维过程和智能行为(如学习、推理、思考、规划等)的学科,既有硬件层面的技术也有软件层面的技术。人工智能硬件技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理等技术;人工智能软件技术主要包括计算机视觉技术、语音识别技术、自然语言处理技术以及机器学习/深度学习、大数据处理技术、知识图谱技术等几大方向。

[0003] 随着人工智能技术的发展,越来越多的应用基于人工智能技术取得了远超过传统算法的效果。深度学习是数据密集型算法和计算密集型算法,为了提高大规模深度学习模型的训练速度和推理速度,可以利用包含多个计算单元或是多个计算单元组的分布式系统来执行数据处理,满足算力要求。

[0004] 在此部分中描述的方法不一定是之前已经设想到或采用的方法。除非另有指明,否则不应假定此部分中描述的任何方法仅因其包括在此部分中就被认为是现有技术。类似地,除非另有指明,否则此部分中提及的问题不应认为在任何现有技术中已被公认。

发明内容

[0005] 本公开提供了一种应用于分布式系统的数据处理方法、装置、电子设备、计算机可读存储介质和计算机程序产品。

[0006] 根据本公开的一方面,提供了一种应用于分布式系统的数据处理方法,其中,所述分布式系统中的多个计算单元能够构成多个计算单元组,待处理数据包括多个数据块,所述方法包括:针对所述多个数据块中的每个数据块,将该数据块划分为与所述多个计算单元组分别对应的多个第一子块;基于与该数据块对应的目标计算单元组,将所述多个第一子块中的每个第一子块划分为与所述目标计算单元组中的多个计算单元分别对应的多个第二子块;通过利用每个第二子块相应的计算单元处理该第二子块,确定所述多个计算单元分别输出的多个第一处理结果;以及通过利用所述目标计算单元组执行针对所述多个第一处理结果的数据归约操作,确定该数据块的处理结果。

[0007] 根据本公开的另一方面,提供了一种应用于分布式系统的数据处理装置,其中,所述分布式系统中的多个计算单元能够构成多个计算单元组,待处理数据包括多个数据块,所述装置包括:第一划分单元,被配置为针对所述多个数据块中的每个数据块,将该数据块划分为与所述多个计算单元组分别对应的多个第一子块;第二划分单元,被配置为基于与该数据块对应的目标计算单元组,将所述多个第一子块中的每个第一子块划分为与所述目标计算单元组中的多个计算单元分别对应的多个第二子块;第一处理单元,被配置为通过

利用每个第二子块相应的计算单元处理该第二子块,确定所述多个计算单元分别输出的多个第一处理结果;以及第二处理单元,被配置为通过利用所述目标计算单元组执行针对所述多个第一处理结果的数据归约操作,确定该数据块的处理结果。

[0008] 根据本公开的另一方面,提供了一种电子设备,包括:至少一个处理器;以及与所述至少一个处理器通信连接的存储器;其中所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行上述数据处理方法。

[0009] 根据本公开的另一方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行上述数据处理方法。

[0010] 根据本公开的另一方面,提供了一种计算机程序产品,包括计算机程序,其中,计算机程序在被处理器执行时能够实现上述数据处理方法。

[0011] 根据本公开的一个或多个实施例,可以保证数据处理结果的稳定性,进而提升分布式数据处理的准确性。

[0012] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0013] 附图示例性地示出了实施例并且构成说明书的一部分,与说明书的文字描述一起用于讲解实施例的示例性实施方式。所示出的实施例仅出于例示的目的,并不限制权利要求的范围。在所有附图中,相同的附图标记指代类似但不一定相同的要素。

[0014] 图1示出了根据本公开示例性实施例的可以在其中实施本文描述的各种方法的示例性系统的示意图;

[0015] 图2示出了根据本公开示例性实施例的应用于分布式系统的数据处理方法的流程图;

[0016] 图3示出了根据本公开示例性实施例的分布式系统的结构示意图;

[0017] 图4示出了根据本公开示例性实施例的数据处理过程的示意图;

[0018] 图5示出了根据本公开示例性实施例的数据处理过程的示意图;

[0019] 图6示出了根据本公开示例性实施例的应用于分布式系统的数据处理装置的结构框图;

[0020] 图7示出了能够用于实现本公开实施例的示例性电子设备的结构框图。

具体实施方式

[0021] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0022] 在本公开中,除非另有说明,否则使用术语“第一”、“第二”等来描述各种要素不意图限定这些要素的位置关系、时序关系或重要性关系,这种术语只是用于将一个元件与另一元件区分开。在一些示例中,第一要素和第二要素可以指向该要素的同一实例,而在某些

情况下,基于上下文的描述,它们也可以指代不同实例。

[0023] 在本公开中对各种所述示例的描述中所使用的术语只是为了描述特定示例的目的,而并非旨在进行限制。除非上下文另外明确地表明,如果不特意限定要素的数量,则该要素可以是一个也可以是多个。此外,本公开中所使用的术语“和/或”涵盖所列出的项目中的任何一个以及全部可能的组合方式。

[0024] 分布式系统中可以利用多个计算单元组来执行数据运算和集合通信操作,以实现数据同步。现有技术中通常是基于当前负载状况来动态分发待处理数据,导致相同的数据可能在不同的轮次被分发到不同的计算单元组。可以理解地,当利用分布式系统进行数据处理时,需要通过各种类型的集合通信操作来实现系统中各个节点间的数据同步和数据交换。由于不同的计算单元组的数据集合逻辑存在差异,即可能影响计算结果的稳定性。

[0025] 例如,当分布式系统中的多个计算单元能够构成多个环形拓扑的计算单元组时,每个环形拓扑的计算单元组都可以应用环形全归约(ring all reduce)算法来针对计算单元组中各个计算单元的数据结果进行归约。由于ring all reduce算法流程的数据交换与累加步骤是基于单向环形拓扑、沿固定方向依次累加各个计算单元中的数据,而不同的数据累加顺序(即各个计算单元在环形拓扑中排列顺序)可能会导致累加结果存在差异,例如可能会受到加法计算中舍入操作的影响,进而影响数据处理结果的稳定性。

[0026] 基于此,本公开提供一种数据处理方法,当系统中的多个计算单元能够构成多个计算单元组时,以待处理数据中的数据块作为划分基础,将每个数据块划分为多个子块并分配给多个计算单元组,以保证数据块与计算单元组的划分-映射方式保持稳定,进而保证数据处理结果的稳定性。

[0027] 下面将结合附图详细描述本公开的实施例。

[0028] 图1示出了根据本公开的实施例可以将本文描述的各种方法和装置在其中实施的示例性系统100的示意图。参考图1,该系统100包括一个或多个客户端设备101、102、103、104、105和106、服务器120以及将一个或多个客户端设备耦接到服务器120的一个或多个通信网络110。客户端设备101、102、103、104、105和106可以被配置为执行一个或多个应用程序。

[0029] 在本公开的实施例中,服务器120可以运行使得能够执行数据处理方法的一个或多个服务或软件应用。

[0030] 在某些实施例中,服务器120还可以提供其他服务或软件应用,这些服务或软件应用可以包括非虚拟环境和虚拟环境。在某些实施例中,这些服务可以作为基于web的服务或云服务提供,例如在软件即服务(SaaS)模型下提供给客户端设备101、102、103、104、105和/或106的用户。

[0031] 在图1所示的配置中,服务器120可以包括实现由服务器120执行的功能的一个或多个组件。这些组件可以包括可由一个或多个处理器执行的软件组件、硬件组件或其组合。操作客户端设备101、102、103、104、105和/或106的用户可以依次利用一个或多个客户端应用程序来与服务器120进行交互以利用这些组件提供的服务。应当理解,各种不同的系统配置是可能的,其可以与系统100不同。因此,图1是用于实施本文所描述的各种方法的系统的一个示例,并且不旨在进行限制。

[0032] 用户可以使用客户端设备101、102、103、104、105和/或106来发送数据处理请求或

待处理数据。客户端设备可以提供使客户端设备的用户能够与客户端设备进行交互的接口。客户端设备还可以经由该接口向用户输出信息。尽管图1仅描绘了六种客户端设备,但是本领域技术人员将能够理解,本公开可以支持任何数量的客户端设备。

[0033] 客户端设备101、102、103、104、105和/或106可以包括各种类别的计算机设备,例如便携式手持设备、通用计算机(诸如个人计算机和膝上型计算机)、工作站计算机、可穿戴设备、智能屏设备、自助服务终端设备、服务机器人、游戏系统、瘦客户端、各种消息收发设备、传感器或其他感测设备等。这些计算机设备可以运行各种类别和版本的软件应用程序和操作系统,例如MICROSOFT Windows、APPLE iOS、类UNIX操作系统、Linux或类Linux操作系统(例如GOOGLE Chrome OS);或包括各种移动操作系统,例如MICROSOFT Windows Mobile OS、iOS、Windows Phone、Android。便携式手持设备可以包括蜂窝电话、智能电话、平板电脑、个人数字助理(PDA)等。可穿戴设备可以包括头戴式显示器(诸如智能眼镜)和其他设备。游戏系统可以包括各种手持式游戏设备、支持互联网的游戏设备等。客户端设备能够执行各种不同的应用程序,例如各种与Internet相关的应用程序、通信应用程序(例如电子邮件应用程序)、短消息服务(SMS)应用程序,并且可以使用各种通信协议。

[0034] 网络110可以是本领域技术人员熟知的任何类别的网络,其可以使用多种可用协议中的任何一种(包括但不限于TCP/IP、SNA、IPX等)来支持数据通信。仅作为示例,一个或多个网络110可以是局域网(LAN)、基于以太网的网络、令牌环、广域网(WAN)、因特网、虚拟网络、虚拟专用网络(VPN)、内部网、外部网、区块链网络、公共交换电话网(PSTN)、红外网络、无线网络(例如蓝牙、WIFI)和/或这些和/或其他网络的任意组合。

[0035] 服务器120可以包括一个或多个通用计算机、专用服务器计算机(例如PC(个人计算机)服务器、UNIX服务器、中端服务器)、刀片式服务器、大型计算机、服务器群集或任何其他适当的布置和/或组合。服务器120可以包括运行虚拟操作系统的一个或多个虚拟机,或者涉及虚拟化的其他计算架构(例如可以被虚拟化以维护服务器的虚拟存储设备的逻辑存储设备的一个或多个灵活池)。在各种实施例中,服务器120可以运行提供下文所描述的功能的一个或多个服务或软件应用。

[0036] 服务器120中的计算单元可以运行包括上述任何操作系统以及任何商业上可用的服务器操作系统的一个或多个操作系统。服务器120还可以运行各种附加服务器应用程序和/或中间层应用程序中的任何一个,包括HTTP服务器、FTP服务器、CGI服务器、JAVA服务器、数据库服务器等。

[0037] 在一些实施方式中,服务器120可以包括一个或多个应用程序,以分析和合并从客户端设备101、102、103、104、105和106的用户接收的数据馈送和/或事件更新。服务器120还可以包括一个或多个应用程序,以经由客户端设备101、102、103、104、105和106的一个或多个显示设备来显示数据馈送和/或实时事件。

[0038] 在一些实施方式中,服务器120可以为分布式系统的服务器,或者是结合了区块链的服务器。服务器120也可以是云服务器,或者是带人工智能技术的智能云计算服务器或智能云主机。云服务器是云计算服务体系中的一项主机产品,以解决传统物理主机与虚拟专用服务器(VPS, Virtual Private Server)服务中存在的管理难度大、业务扩展性弱的缺陷。

[0039] 系统100还可以包括一个或多个数据库130。在某些实施例中,这些数据库可以用

于存储数据和其他信息。例如,数据库130中的一个或多个可用于存储诸如音频文件和视频文件的信息。数据库130可以驻留在各种位置。例如,由服务器120使用的数据库可以在服务器120本地,或者可以远离服务器120且可以经由基于网络或专用的连接与服务器120通信。数据库130可以是不同的类别。在某些实施例中,由服务器120使用的数据库例如可以是关系数据库。这些数据库中的一个或多个可以响应于命令而存储、更新和检索到数据库以及来自数据库的数据。

[0040] 在某些实施例中,数据库130中的一个或多个还可以由应用程序使用来存储应用程序数据。由应用程序使用的数据库可以是不同类别的数据库,例如键值存储库,对象存储库或由文件系统支持的常规存储库。

[0041] 图1的系统100可以以各种方式配置和操作,以使得能够应用根据本公开所描述的各种方法和装置。

[0042] 图2示出了根据本公开示例性实施例的数据处理方法200的流程图,其中,所述分布式系统中的多个计算单元能够构成多个计算单元组,待处理数据包括多个数据块。如图2所示,方法200包括:

[0043] 步骤S201、针对所述多个数据块中的每个数据块,将该数据块划分为与所述多个计算单元组分别对应的多个第一子块;

[0044] 步骤S202、基于与该数据块对应的目标计算单元组,将所述多个第一子块中的每个第一子块划分为与所述目标计算单元组中的多个计算单元分别对应的多个第二子块;

[0045] 步骤S203、通过利用每个第二子块相应的计算单元处理该第二子块,确定所述多个计算单元分别输出的多个第一处理结果;以及

[0046] 步骤S204、通过利用所述目标计算单元组执行针对所述多个第一处理结果的数据归约操作,确定该数据块的处理结果。

[0047] 通过应用上述方法200,能够在系统中存在多个计算单元组的情况下,以待处理数据中数据块作为划分基础,将每个数据块划分为多个第一子块并分配给多个计算单元组,以保证数据块与计算单元组的划分-映射方式保持稳定,进而保证数据处理结果的稳定性。

[0048] 在一些示例中,待处理数据中的数据块信息可以是显式的,例如,待处理数据可以由自然断开的多个数据块构成的数据块序列。在一些示例中,也可以是在步骤S201之前先将待处理数据划分为多个数据块,例如,可以是按照预设的块大小来将待处理数据划分为多个数据块。

[0049] 在一些示例中,计算单元组可以是分布式系统中能够完成一组集合通信操作的多个计算单元的集合。可以理解地,不同计算单元组包括的计算单元可以相同,也可以不同。例如,当前述的集合通信操作是基于单向环拓扑执行的环形全归约操作,当分布式系统中的多个计算单元能够构成N个环形拓扑时,则该分布式系统中计算单元组的数量为2N。

[0050] 根据一些实施例,所述目标计算单元组中的多个计算单元构成环形拓扑,步骤S204中通过利用所述目标计算单元组执行针对所述多个第一处理结果的数据归约操作,确定该数据块的处理结果包括:基于所述目标计算单元组中的多个计算单元在所述环形拓扑中的排列顺序,依次累加每个计算单元的第一处理结果;以及基于累加结果确定该数据块的处理结果。由此,能够在基于ring-all-reduce算法实现集合通信的过程中,通过应用方法200所述的数据分片和算力分配方式来保证各个计算单元的中间处理结果的累加顺序保

持不变,进而保证累加结果的稳定性。

[0051] 根据一些实施例,所述基于所述目标计算单元组中的多个计算单元在所述环形拓扑中的排列顺序,依次累加每个计算单元相应的第一处理结果包括:利用所述多个计算单元中的每个计算单元执行如下步骤:从所述环形拓扑中位于该计算单元之前的前序计算单元接收待累加数据;通过基于所述待累加数据和该计算单元的第一处理结果执行累加操作,确定中间数据;以及将所述中间数据发送至所述环形拓扑中位于该计算单元之后的后继计算单元,以使得所述后继计算单元能够基于所述中间数据和所述后继计算单元的第一处理结果执行累加操作。由此,能够基于环形迭代操作更高效地实现全部计算单元的数据归约操作。

[0052] 图3示出了根据本公开示例性实施例的分布式系统300的结构示意图。如图3所示,在这一示例中,系统300包括2个处理核心(处理核心311和处理核心312)、4个通信接口(通信接口321、通信接口322、通信接口323和通信接口324)和8个计算单元。图中的点划线双向箭头示出了处理核心间的第一通信链路,虚线双向箭头示出了处理核心与计算单元间的第二通信链路,实线双向箭头示出了计算单元间的第三通信链路。参加图3示出的8个计算单元间多组点对点的第三通信链路可知,8个节点能够构成如下4组能够用于执行数据累加操作的单向环:

[0053] 单向环A:计算单元331—计算单元332—计算单元336—计算单元335—计算单元337—计算单元338—计算单元334—计算单元333;

[0054] 单向环B:计算单元333—计算单元334—计算单元338—计算单元337—计算单元335—计算单元336—计算单元332—计算单元331;

[0055] 单向环C:计算单元331—计算单元335—计算单元338—计算单元336—计算单元337—计算单元333—计算单元332—计算单元334;

[0056] 单向环D:计算单元334—计算单元332—计算单元333—计算单元337—计算单元336—计算单元338—计算单元335—计算单元331。

[0057] 可以理解地,在多个计算单元间的互联结构已经确定的情况下,可以基于各种已有的探测工具或分析算法来确定系统中的多个单向环拓扑,进而作为显式的拓扑信息保存,从而能够在执行本公开所述的方法200时直接应用。

[0058] 在一些示例中,处理核心可以是CPU(Central Processing Unit,中央处理器),运算节点可以是GPU(Graphics Processing Unit,图形处理器),第一通信链路可以是QPI(Quick Path Interconnect,快速信道互联)链路,第二通信链路可以是PCI(Peripheral Component Interconnect,外设部件互联)链路,第三通信链路可以是基于NVLink的点对点通信链路。

[0059] 如前所述,常规技术中通常是基于多个环形拓扑各自的负载状况来动态分发数据交换与累加任务,例如可以是基于数据规模来将同一批次(批次规模参数batch_size为可变参数)中的待处理的数据平均分布到多个环形拓扑中。在这种情况下,当数据规模为8192的数据处理请求A中的某次累加操作和数据规模为8192的数据处理请求B中的某次累加操作被合并到同一批次中(即batch_size=2),则数据处理请求A的前4096个数据可能被分发至单向环A,即按照单向环A中的节点排列顺序执行节点数据累加,而后4096个数据可能被分发至单向环B,即按照单向环B中的节点排列顺序执行节点数据累加。而当数据处理请求A

中的累加操作被单独放在一个批次中(即batch_size=1),则数据处理请求A的8192个数据可能被平均分配至上述的4个单向环中,每个单向环负责处理2048个数据。可以看出,在这种实现方式中,数据处理请求A所包含的各个数据元素在累加操作中的累加顺序可能发生变化,进而导致计算结果的不稳定。

[0060] 在一个示例中,当输入的待处理数据的大小为8时,该待处理数据可以被切分并分配到多个计算单元组处理。当分布式系统中计算单元组的数量为2时,基于待处理数据进行切分的第一切分结果的大小即为4。

[0061] 在此基础上,每个计算单元组内部可以利用多步操作来处理该计算单元组的第一切分结果。图4示出了根据本公开示例性实施例的数据处理过程的示意图。如图4所示,计算单元0、计算单元1、计算单元2、计算单元3构成的2个计算单元组可以基于环形迭代步骤实现数据累加,每个计算单元在每一步骤中针对大小为1的第二切分结果执行传输和累加。图4即示出了第一步骤的传输过程的示意图。在这种情况下,计算单元组A得到的累加结果为 $[a_0+b_0+c_0+d_0, b_1+c_1+d_1+a_1, c_2+d_2+a_2+b_2, d_3+a_3+b_3+c_3]$,计算单元组B得到的累加结果为 $[d_4+a_4+b_4+c_4, c_5+d_5+a_5+b_5, b_6+c_6+d_6+a_6, a_7+b_7+c_7+d_7]$ 。

[0062] 在这一示例中,假设每个数据处理请求的长度为4,则输入的待处理数据中包含2个数据处理请求(例如,计算单元0的数据处理请求 $query_0=[a_0, a_1, a_2, a_3]$,数据处理请求 $query_1=[a_4, a_5, a_6, a_7]$),并且2个数据处理请求被分别分配到计算单元组A和计算单元组B处理。可以看出,计算单元组A和计算单元组B的数据累加顺序不同,从而可能导致不同数据处理请求的处理结果存在差异。

[0063] 在另一示例中,假设每个数据处理请求的长度为2,则输入的待处理数据中包含4个数据处理请求,并且每个计算单元组负责处理2个数据处理请求(例如,计算单元0中与计算单元组A对应的数据处理请求 $query_0=[a_0, a_1,]$,数据处理请求 $query_1=[a_2, a_3]$)。可以看出,不同数据处理请求在同一计算单元组的切分结果内的起始位置不同也会导致数据累加顺序不同,从而可能导致不同数据处理请求的处理结果存在差异。

[0064] 图5示出了根据本公开示例性实施例的数据处理过程的示意图。如图5所示,在这一示例中,通过应用上述的方法200所述的数据切分和算力分配方法,可以以待处理数据中的数据块为最小切分单元。例如,当数据块大小为8、计算单元组的数量为2时,即可得到 $[a_0, a_1, a_2, a_3]$ 和 $[a_4, a_5, a_6, a_7]$ 两个第一子块,并且将两个第一子块分别分配给计算单元组A和计算单元组B进行处理。在这种情况下,无论数据处理请求中包含多少数据块,每个数据元素对应的算力分配方式和数据处理逻辑都可以保持稳定。

[0065] 在一个示例中,图5的数据处理过程例如是可以基于如下步骤实现的:

[0066] 步骤S1、将每个数据块顺序切分成第一数量(即计算单元组的数量)个第一子块,并将每个第一子块分配给一个计算单元组处理;

[0067] 步骤S2、将每个第一子块内部顺序切分成第二数量(即计算单元组中计算单元的数量)个第二子块;

[0068] 步骤S3、在每个计算单元内部遍历所有的数据块,从该计算单元组相应的第一子块内部选取一个第二子块发送至该计算单元组中的后继计算单元,可以理解地,每个计算单元选取的第二子块的排列序数互不相同;

[0069] 通过循环执行上述的步骤S3,即可令每个计算单元组中的每个计算单元在每一轮

次中从该计算单元组中的前序计算单元接收数据,与该计算单元的相应数据执行累加计算,再将累加结果发送至该计算单元组中的后继计算单元。当环形迭代步骤执行完毕后,每个计算单元组的每个计算单元上均拥有该计算单元组中所有计算单元的数据累加结果。

[0070] 根据一些实施例,所述待处理数据是利用特征编码网络得到的,并且其中,所述待处理数据包括的多个数据块是利用如下的划分操作得到的:基于所述特征编码网络的输出维数,将所述待处理数据划分为大小相同的多个数据块。由此,可以应用上述的方法200实现特征数据处理,通过基于特征编码向量空间维数来确定待处理数据中的数据块,可以更简便高效地得到固定大小的数据块,平衡分布式系统中各个计算单元组的计算负载,提升处理效率。

[0071] 在一些示例中,待处理数据可以是对自然语言文本数据进行嵌入表示(embedding)得到的。通过对文本数据的编码数据进行运算,则可以实现各种类型的自然语言处理任务,例如可以将自然语言文本数据映射到嵌入空间来得到语义特征张量,进而基于此进行语义分析、实体抽取等操作。可以理解地,嵌入空间中的语义特征张量的数据长度必然是嵌入空间维数的整数倍,在这种情况下,通过基于嵌入空间维数来划分待处理数据的数据块,即可简便高效地将待处理数据划分为整数个数据块,进而通过针对数据块来执行数据切分和算力分配操作来保证每个数据块的分布式处理逻辑的稳定性,进而保证针对待处理数据的运算结果的稳定性。

[0072] 根据一些实施例,基于所述特征编码网络的输出维数,将所述待处理数据划分为大小相同的多个数据块包括:基于所述特征编码网络的输出维数和所述待处理数据的数据格式,将所述待处理数据划分为大小相同的多个数据块。由此,能够进一步结合数据格式来准确地确定每个数据块的数据量,平衡分布式系统中各个计算单元组的计算负载,提升处理效率。

[0073] 根据本公开的一方面,还提供一种应用于分布式系统的数据处理装置。图6示出了根据本公开示例性实施例的应用于分布式系统的数据处理装置600的结构框图,其中,所述分布式系统中的多个计算单元能够构成多个计算单元组,待处理数据包括多个数据块。如图6所示,装置600包括:

[0074] 第一划分单元601,被配置为针对所述多个数据块中的每个数据块,将该数据块划分为与所述多个计算单元组分别对应的多个第一子块;

[0075] 第二划分单元602,被配置为基于与该数据块对应的目标计算单元组,将所述多个第一子块中的每个第一子块划分为与所述目标计算单元组中的多个计算单元分别对应的多个第二子块;

[0076] 第一处理单元603,被配置为通过利用每个第二子块相应的计算单元处理该第二子块,确定所述多个计算单元分别输出的多个第一处理结果;以及

[0077] 第二处理单元604,被配置为通过利用所述目标计算单元组执行针对所述多个第一处理结果的数据归约操作,确定该数据块的处理结果。

[0078] 根据一些实施例,所述目标计算单元组中的多个计算单元构成环形拓扑,所述第二处理单元604包括:累加子单元,被配置为基于所述目标计算单元组中的多个计算单元在所述环形拓扑中的排列顺序,依次累加每个计算单元的第一处理结果;以及确定子单元,被配置为基于累加结果确定该数据块的处理结果。

[0079] 根据一些实施例,所述累加子单元被配置为利用所述多个计算单元中的每个计算单元执行如下步骤:从所述环形拓扑中位于该计算单元之前的前序计算单元接收待累加数据;通过基于所述待累加数据和该计算单元的第一处理结果执行累加操作,确定中间数据;以及将所述中间数据发送至所述环形拓扑中位于该计算单元之后的后继计算单元,以使得所述后继计算单元能够基于所述中间数据和所述后继计算单元的第一处理结果执行累加操作。

[0080] 根据一些实施例,所述待处理数据是利用特征编码网络得到的,并且其中,所述待处理数据包括的多个数据块是利用如下的划分操作得到的:基于所述特征编码网络的输出维数,将所述待处理数据划分为大小相同的多个数据块。

[0081] 根据一些实施例,所述待处理数据包括的多个数据块是利用如下的划分操作得到的:基于所述特征编码网络的输出维数和所述待处理数据的数据格式,将所述待处理数据划分为大小相同的多个数据块。

[0082] 根据本公开的另一方面,还提供一种电子设备,包括:至少一个处理器;以及与所述至少一个处理器通信连接的存储器;其中所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行上述的数据处理方法。

[0083] 根据本公开的另一方面,还提供一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行上述的数据处理方法。

[0084] 根据本公开的另一方面,还提供一种计算机程序产品,包括计算机程序,其中,所述计算机程序再被处理器执行时实现上述的数据处理方法。

[0085] 参考图7,现将描述可以作为本公开的服务器或客户端的电子设备700的结构框图,其是可以应用于本公开的各方面的硬件设备的示例。电子设备旨在表示各种形式的数字电子的计算机设备,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0086] 如图7所示,设备700包括计算单元701,其可以根据存储在只读存储器(ROM)702中的计算机程序或者从存储单元708加载到随机访问存储器(RAM)703中的计算机程序,来执行各种适当的动作和处理。在RAM 703中,还可存储设备700操作所需的各种程序和数据。计算单元701、ROM 702以及RAM 703通过总线704彼此相连。输入/输出(I/O)接口705也连接至总线704。

[0087] 设备700中的多个部件连接至I/O接口705,包括:输入单元706、输出单元707、存储单元708以及通信单元709。输入单元706可以是能向设备700输入信息的任何类别的设备,输入单元706可以接收输入的数字或字符信息,以及产生与电子设备的用户设置和/或功能控制有关的键信号输入,并且可以包括但不限于鼠标、键盘、触摸屏、轨迹板、轨迹球、操作杆、麦克风和/或遥控器。输出单元707可以是能呈现信息的任何类别的设备,并且可以包括但不限于显示器、扬声器、视频/音频输出终端、振动器和/或打印机。存储单元708可以包括但不限于磁盘、光盘。通信单元709允许设备700通过诸如因特网的计算机网络和/或各种电

信网络与其他设备交换信息/数据,并且可以包括但不限于调制解调器、网卡、红外通信设备、无线通信收发机和/或芯片组,例如蓝牙™设备、802.11设备、WiFi设备、WiMax设备、蜂窝通信设备和/或类似物。

[0088] 计算单元701可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元701的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元701执行上文所描述的各个方法和处理,例如数据处理方法。例如,在一些实施例中,数据处理方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元708。在一些实施例中,计算机程序的部分或者全部可以经由ROM 702和/或通信单元709而被载入和/或安装到设备700上。当计算机程序加载到RAM 703并由计算单元701执行时,可以执行上文描述的数据处理方法的一个或多个步骤。备选地,在其他实施例中,计算单元701可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行数据处理方法。

[0089] 本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、复杂可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0090] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0091] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0092] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用

任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0093] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)、互联网和区块链网络。

[0094] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。服务器可以是云服务器,也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0095] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本公开中记载的各步骤可以并行地执行、也可以顺序地或以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0096] 虽然已经参照附图描述了本公开的实施例或示例,但应理解,上述的方法、系统和设备仅仅是示例性的实施例或示例,本发明的范围并不由这些实施例或示例限制。实施例或示例中的各种要素可以被省略或者可由其等同要素替代。此外,可以通过不同于本公开中描述的次序来执行各步骤。进一步地,可以以各种方式组合实施例或示例中的各种要素。重要的是随着技术的演进,在此描述的很多要素可以由本公开之后出现的等同要素进行替换。

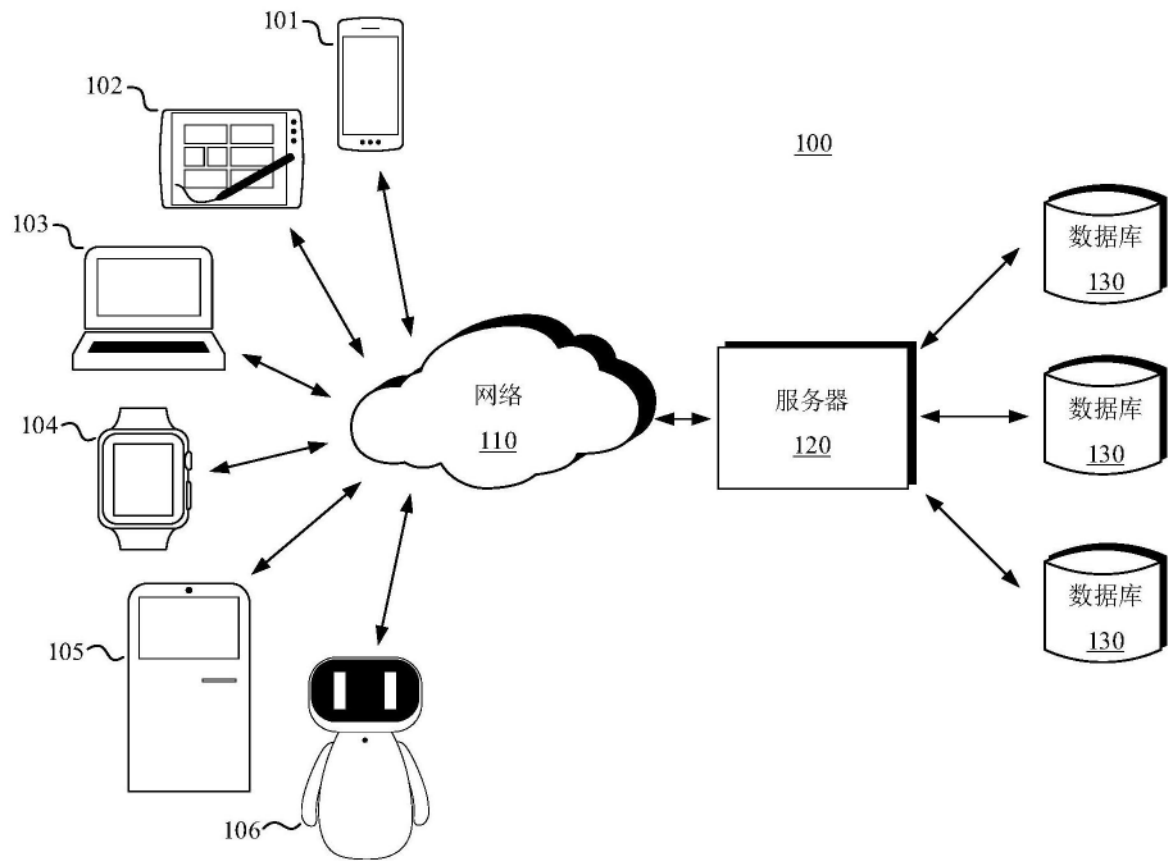


图1

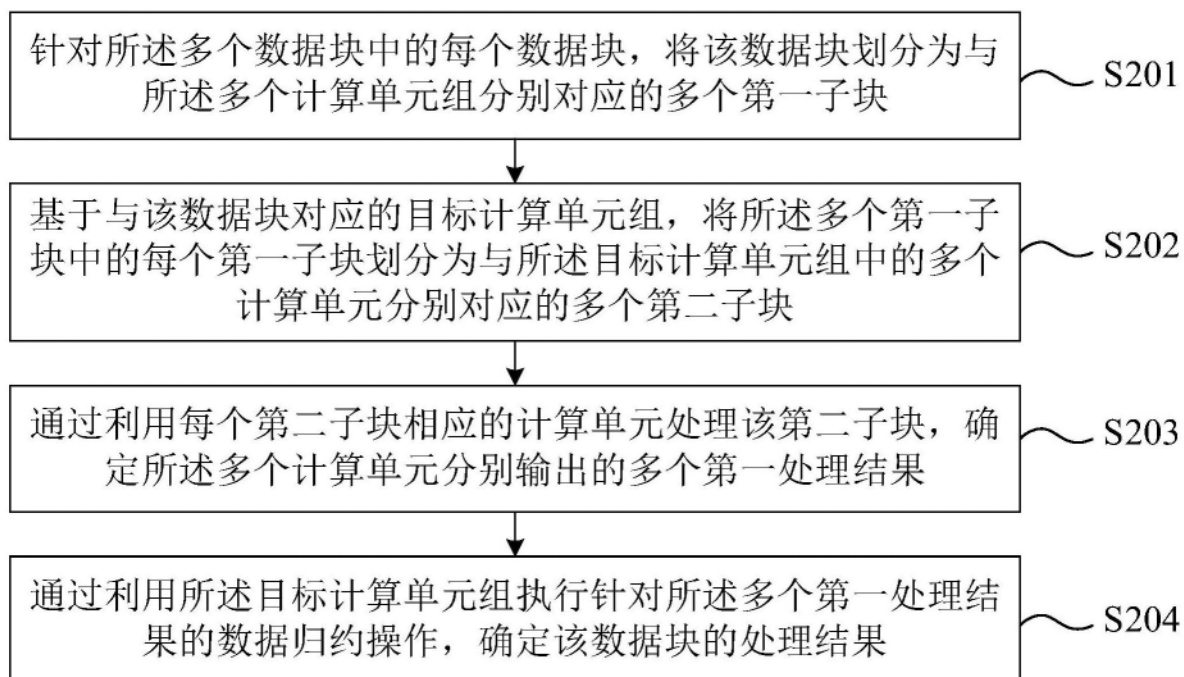
200

图2

系统300

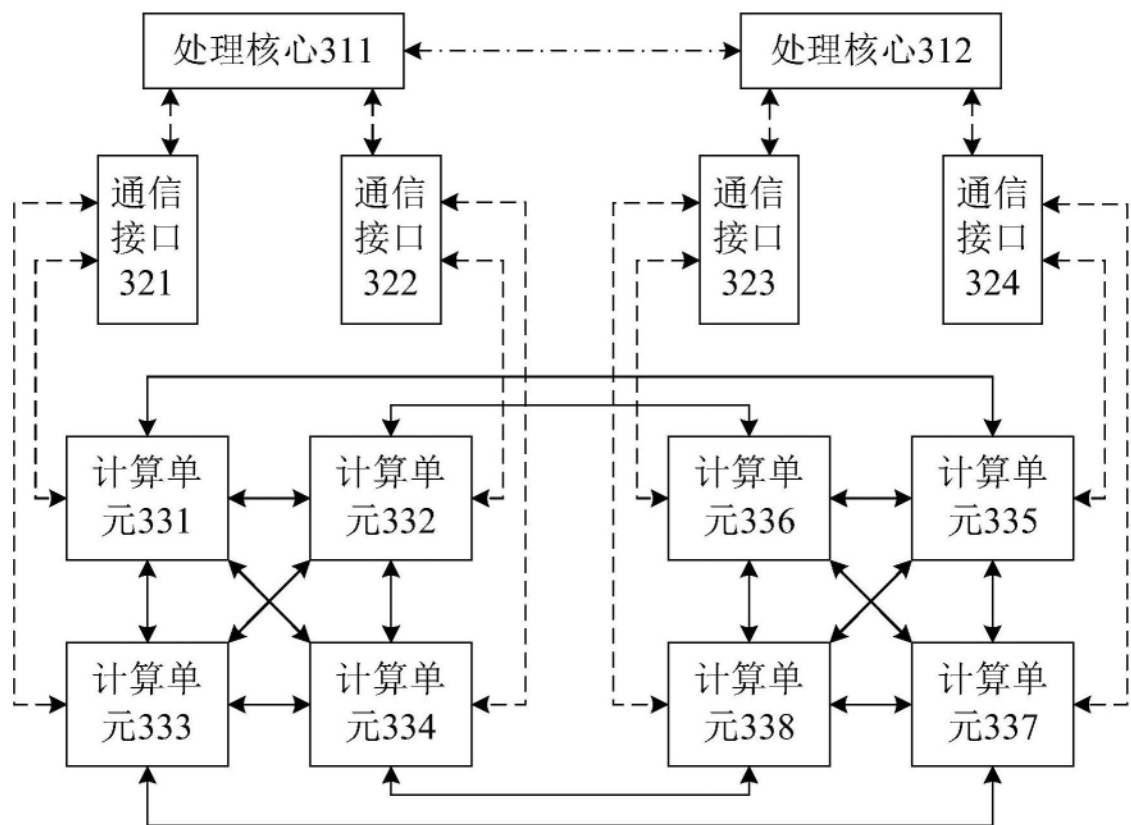
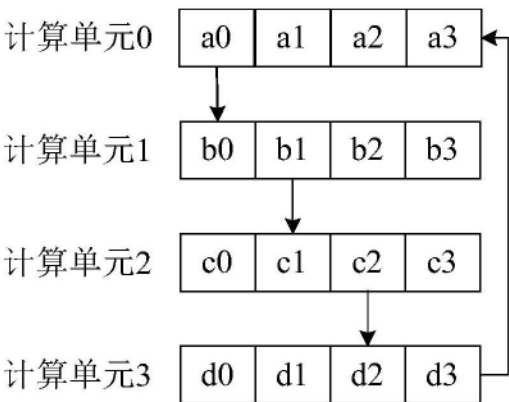


图3

计算单元组A [0-1-2-3-0]



计算单元组B [3-2-1-0-3]

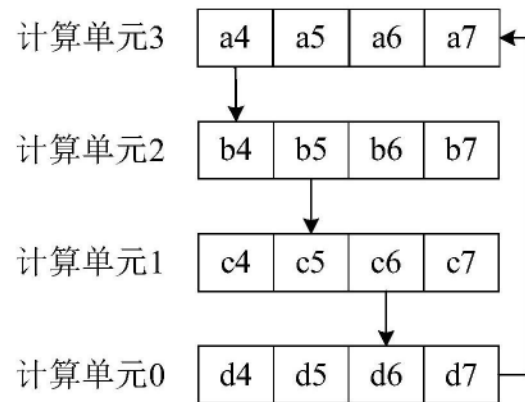


图4

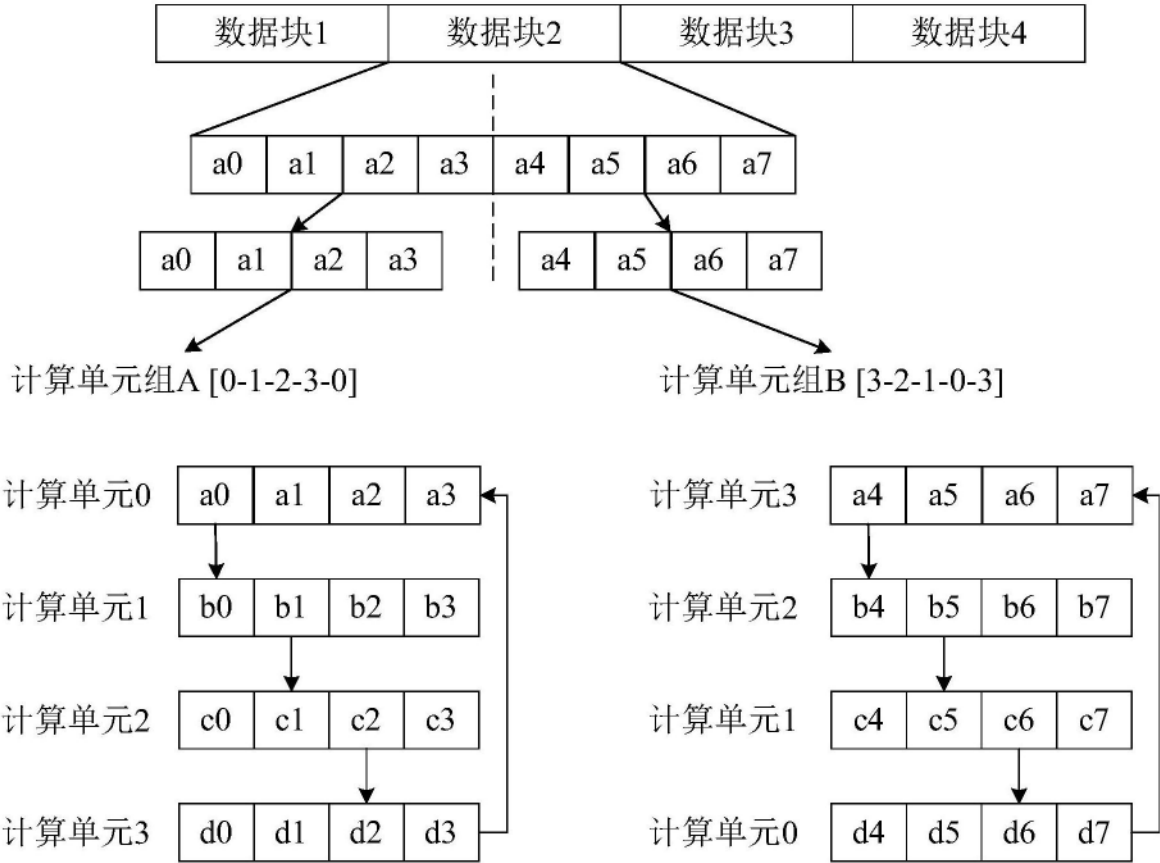


图5



图6

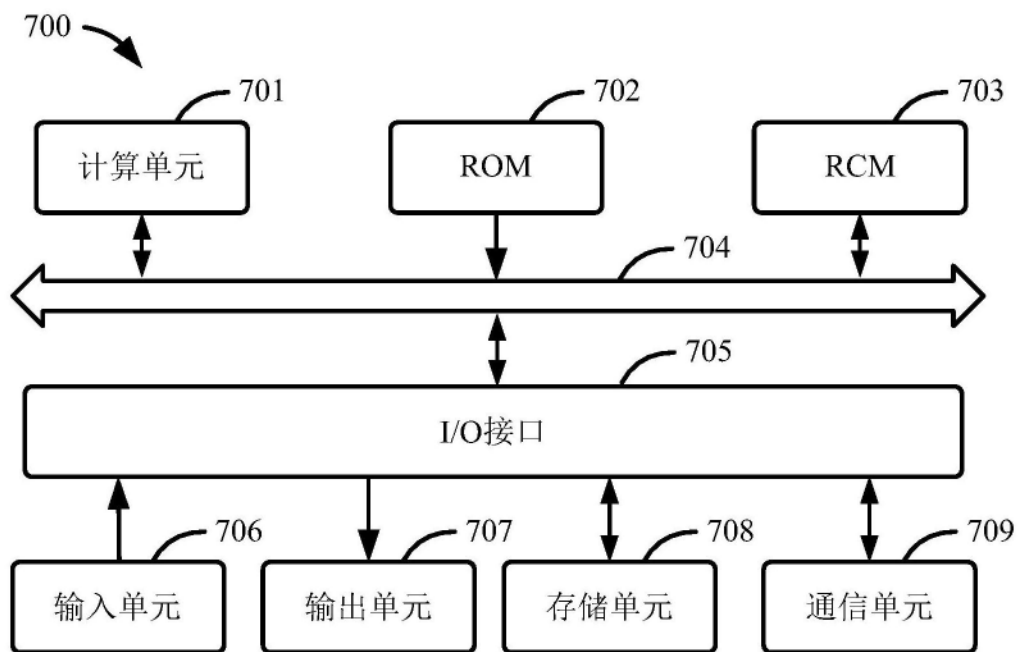


图7