(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 118051779 B
(45)授权公告日 2024.07.16

(21)申请号 202410438532.7
(22)申请日 2024.04.12
(65)同一申请的已公布的文献号
申请公布号 CN 118051779 A
(43)申请公布日 2024.05.17
(73)专利权人 清华大学
地址 100084 北京市海淀区清华园1号
(72)发明人 汪玉 黄子潇 宁雪妃
(74)专利代理机构 北京清亦华知识产权代理事务所(普通合伙) 11201
专利代理师 黄德海
(51)Int.Cl.
G06F 18/214(2023.01)

(56)对比文件
CN 117407713 A,2024.01.16
CN 116629352 A,2023.08.22

审查员 覃冬梅

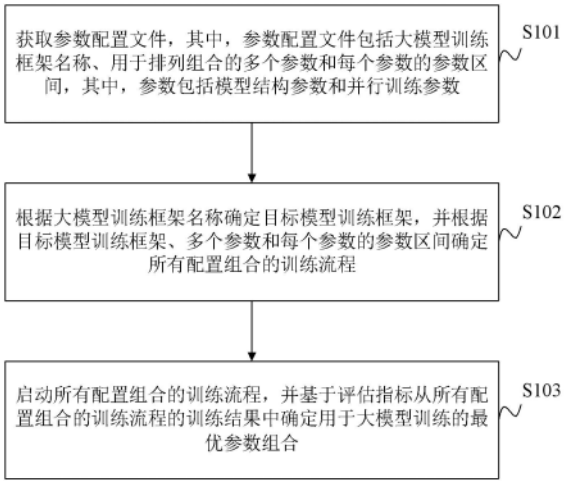
权利要求书2页 说明书10页 附图3页

(54)发明名称

用于大模型训练的参数自动搜索方法、装置及电子设备

(57)摘要

本发明涉及深度学习技术领域,特别涉及一种用于大模型训练的参数自动搜索方法、装置及电子设备,包括:获取参数配置文件,其包括大模型训练框架名称、多个参数和每个参数的参数区间;根据大模型训练框架名称确定目标模型训练框架,根据目标模型训练框架、多个参数和每个参数的参数区间确定所有配置组合的训练流程;启动所有配置组合的训练流程,并基于评估指标从所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合。由此,通过目标模型训练框架对参数配置组合进行枚举训练,即可得到最优参数配置组合,解决了当前确定最优参数配置的过程繁琐耗时,导致模型开发周期较长的问题,提高用户确定最优参数配置的效率,降低开发成本。



CN 118051779 B

1.一种用于大模型训练的参数自动搜索方法,其特征在于,包括以下步骤:

获取参数配置文件,其中,所述参数配置文件包括大模型训练框架名称、用于排列组合的多个参数和每个参数的参数区间,其中,所述参数包括模型结构参数和并行训练参数;

根据所述大模型训练框架名称确定目标模型训练框架,并根据所述目标模型训练框架、所述多个参数和所述每个参数的参数区间确定所有配置组合的训练流程;

启动所述所有配置组合的训练流程,并基于评估指标从所述所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合;

其中,所述根据所述目标模型训练框架、所述多个参数和所述每个参数的参数区间确定所有配置组合的训练流程,包括:从所述参数配置文件中获取每次训练的迭代次数;根据所述多个参数和所述每个参数的参数区间确定所有参数的配置组合;基于所述迭代次数和所述所有参数的配置组合确定所述所有配置组合的训练流程;

其中,所述基于评估指标从所述所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合,包括:从所述参数配置文件中获取需保留的最优参数组合的数量;基于所述训练结果获取每个配置组合的评估指标值;基于所述最优参数组合的数量和所述每个配置组合的评估指标值,确定所述最优参数组合;

其中,所述评估指标值包括:利用监控工具直接获取的评估指标值,和/或,由所述目标模型训练框架进行计算得到的无法直接获取的评估指标值,和/或,通过所述目标模型训练框架调用用户自定义的指标计算方式得到的评估指标值。

2.根据权利要求1所述的用于大模型训练的参数自动搜索方法,其特征在于,在获取所述参数配置文件之后,还包括:

识别在所述参数配置文件中的所述多个参数中未给出参数区间的目标参数;

获取所述目标参数的默认参数区间,并将所述默认参数区间作为所述目标参数的参数区间。

3.根据权利要求1所述的用于大模型训练的参数自动搜索方法,其特征在于,在根据所述大模型训练框架名称确定所述目标模型训练框架之后,还包括:

利用所述目标模型训练框架,校验所述多个参数中是否存在不满足预设兼容条件的不兼容参数;

若所述多个参数中存在不满足预设兼容条件的不兼容参数,则针对所述不兼容参数进行报错提醒。

4.根据权利要求3所述的用于大模型训练的参数自动搜索方法,其特征在于,在针对所述不兼容参数进行报错提醒之后,还包括:

接收用户针对所述不兼容参数反馈的参数修改指令;

基于所述参数修改指令修改所述不兼容参数。

5.根据权利要求1-4任一项所述的用于大模型训练的参数自动搜索方法,其特征在于,在启动所述所有配置组合的训练流程时,还包括:

记录训练启动失败的配置组合。

6.一种用于大模型训练的参数自动搜索装置,其特征在于,包括:

获取模块,用于获取参数配置文件,其中,所述参数配置文件包括大模型训练框架名称、用于排列组合的多个参数和每个参数的参数区间,其中,所述参数包括模型结构参数和

并行训练参数;

第一确定模块,用于根据所述大模型训练框架名称确定目标模型训练框架,并根据所述目标模型训练框架、所述多个参数和所述每个参数的参数区间确定所有配置组合的训练流程;

第二确定模块,用于启动所述所有配置组合的训练流程,并基于评估指标从所述所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合;

其中,所述第一确定模块,具体用于:从所述参数配置文件中获取每次训练的迭代次数;根据所述多个参数和所述每个参数的参数区间确定所有参数的配置组合;基于所述迭代次数和所述所有参数的配置组合确定所述所有配置组合的训练流程;

其中,所述第二确定模块,具体用于:从所述参数配置文件中获取需保留的最优参数组合的数量;基于所述训练结果获取每个配置组合的评估指标值;基于所述最优参数组合的数量和所述每个配置组合的评估指标值,确定所述最优参数组合;

其中,所述评估指标值包括:利用监控工具直接获取的评估指标值,和/或,由所述目标模型训练框架进行计算得到的无法直接获取的评估指标值,和/或,通过所述目标模型训练框架调用用户自定义的指标计算方式得到的评估指标值。

7.一种电子设备,其特征在于,包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述程序,以实现如权利要求1-5任一项所述的用于大模型训练的参数自动搜索方法。

8.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,该程序被处理器执行,以用于实现如权利要求1-5任一项所述的用于大模型训练的参数自动搜索方法。

用于大模型训练的参数自动搜索方法、装置及电子设备

技术领域

[0001] 本发明涉及深度学习技术领域,特别涉及一种用于大模型训练的参数自动搜索方法、装置及电子设备。

背景技术

[0002] 大模型是指模型参数量较大(通常在10亿及以上)的神经网络模型,在图像、文本、音频等多个领域都有典型的应用。在大模型训练框架出现之前,用户开发模型需要自己实现模型结构,当模型参数较大时,用户还需要自己实现模型的并行策略以加速模型的训练速度和减小显存占用。然而,开发模型并行策略对开发门槛要求较高,低效的分布式训练策略可能会大幅降低模型的训练效率甚至导致训练结果异常,并且手动逐个实现模型的并行策略容易导致代码可维护性和可扩展性较差。因此,提供支持高效分布式训练的大模型开源框架,减小用户在开发模型并行策略上的开发成本,是大模型训练中的重要研究方向。

[0003] 相关技术中,支持高效分布式训练的大模型训练框架包括:基于深度学习框架PyTorch的大模型训练框架Megatron-LM;采用零冗余优化器内存优化技术(Zero Redundancy Optimizer,简称ZeRO)的开源大模型训练框架Megatron-DeepSpeed;无缝集成主流深度学习框架PyTorch的大模型训练框架。

[0004] 然而,利用上述大模型训练框架确定最优参数配置的过程依然繁琐且耗时,当用户更换集群拓扑或者机器型号时,还需要重新手动搜索模型训练的最优配置,导致模型开发周期较长,亟待解决。

发明内容

[0005] 本发明提供一种用于大模型训练的参数自动搜索方法、装置及电子设备,以解决当前确定最优参数配置的过程繁琐耗时,导致模型开发周期较长的问题,提高用户确定最优参数配置的效率,降低开发成本。

[0006] 为达到上述目的,本发明第一方面实施例提出一种用于大模型训练的参数自动搜索方法,包括以下步骤:

[0007] 获取参数配置文件,其中,所述参数配置文件包括大模型训练框架名称、用于排列组合的多个参数和每个参数的参数区间,其中,所述参数包括模型结构参数和并行训练参数;

[0008] 根据所述大模型训练框架名称确定目标模型训练框架,并根据所述目标模型训练框架、所述多个参数和所述每个参数的参数区间确定所有配置组合的训练流程;

[0009] 启动所述所有配置组合的训练流程,并基于评估指标从所述所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合。

[0010] 根据本发明的一个实施例,在获取所述参数配置文件之后,还包括:

[0011] 识别在所述参数配置文件中的所述多个参数中未给出参数区间的目标参数;

[0012] 获取所述目标参数的默认参数区间,并将所述默认参数区间作为所述目标参数的

参数区间。

[0013] 根据本发明的一个实施例,在根据所述大模型训练框架名称确定所述目标模型训练框架之后,还包括:

[0014] 利用所述目标模型训练框架,校验所述多个参数中是否存在不满足预设兼容条件的不兼容参数;

[0015] 若所述多个参数中存在不满足预设兼容条件的不兼容参数,则针对所述不兼容参数进行报错提醒。

[0016] 根据本发明的一个实施例,在针对所述不兼容参数进行报错提醒之后,还包括:

[0017] 接收用户针对所述不兼容参数反馈的参数修改指令;

[0018] 基于所述参数修改指令修改所述不兼容参数。

[0019] 根据本发明的一个实施例,所述根据所述目标模型训练框架、所述多个参数和所述每个参数的参数区间确定所有配置组合的训练流程,包括:

[0020] 从所述参数配置文件中获取每次训练的迭代次数;

[0021] 根据所述多个参数和所述每个参数的参数区间确定所有参数的配置组合;

[0022] 基于所述迭代次数和所述所有参数的配置组合确定所述所有配置组合的训练流程。

[0023] 根据本发明的一个实施例,所述基于评估指标从所述所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合,包括:

[0024] 从所述参数配置文件中获取需保留的最优参数组合的数量;

[0025] 基于所述训练结果获取每个配置组合的评估指标值;

[0026] 基于所述最优参数组合的数量和所述每个配置组合的评估指标值,确定所述最优参数组合。

[0027] 根据本发明的一个实施例,在启动所述所有配置组合的训练流程时,还包括:

[0028] 记录训练启动失败的配置组合。

[0029] 根据本发明实施例提出的用于大模型训练的参数自动搜索方法,通过获取参数配置文件,其包括大模型训练框架名称、多个参数和每个参数的参数区间,可以根据大模型训练框架名称确定目标模型训练框架,根据目标模型训练框架、多个参数和每个参数的参数区间确定所有配置组合的训练流程,启动所有配置组合的训练流程,并基于评估指标从所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合。由此,通过目标模型训练框架对参数配置组合进行枚举训练,即可得到最优参数配置组合,解决了当前确定最优参数配置的过程繁琐耗时,导致模型开发周期较长的问题,提高用户确定最优参数配置的效率,降低开发成本。

[0030] 为达到上述目的,本发明第二方面实施例提出一种用于大模型训练的参数自动搜索装置,包括:

[0031] 获取模块,用于获取参数配置文件,其中,所述参数配置文件包括大模型训练框架名称、用于排列组合的多个参数和每个参数的参数区间,其中,所述参数包括模型结构参数和并行训练参数;

[0032] 第一确定模块,用于根据所述大模型训练框架名称确定目标模型训练框架,并根据所述目标模型训练框架、所述多个参数和所述每个参数的参数区间确定所有配置组合的

训练流程；

[0033] 第二确定模块，用于启动所述所有配置组合的训练流程，并基于评估指标从所述所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合。

[0034] 根据本发明的一个实施例，在获取所述参数配置文件之后，所述获取模块，还用于：

[0035] 识别在所述参数配置文件中的所述多个参数中未给出参数区间的目标参数；

[0036] 获取所述目标参数的默认参数区间，并将所述默认参数区间作为所述目标参数的参数区间。

[0037] 根据本发明的一个实施例，在根据所述大模型训练框架名称确定所述目标模型训练框架之后，所述第一确定模块，还包括：

[0038] 校验单元，用于利用所述目标模型训练框架，校验所述多个参数中是否存在不满足预设兼容条件的不兼容参数；

[0039] 报错单元，用于在所述多个参数中存在不满足预设兼容条件的不兼容参数时，针对所述不兼容参数进行报错提醒。

[0040] 根据本发明的一个实施例，在针对所述不兼容参数进行报错提醒之后，所述报错单元，还用于：

[0041] 接收用户针对所述不兼容参数反馈的参数修改指令；

[0042] 基于所述参数修改指令修改所述不兼容参数。

[0043] 根据本发明的一个实施例，所述第一确定模块，具体用于：

[0044] 从所述参数配置文件中获取每次训练的迭代次数；

[0045] 根据所述多个参数和所述每个参数的参数区间确定所有参数的配置组合；

[0046] 基于所述迭代次数和所述所有参数的配置组合确定所述所有配置组合的训练流程。

[0047] 根据本发明的一个实施例，所述第二确定模块，具体用于：

[0048] 从所述参数配置文件中获取需保留的最优参数组合的数量；

[0049] 基于所述训练结果获取每个配置组合的评估指标值；

[0050] 基于所述最优参数组合的数量和所述每个配置组合的评估指标值，确定所述最优参数组合。

[0051] 根据本发明的一个实施例，在启动所述所有配置组合的训练流程时，所述第二确定模块，还用于：

[0052] 记录训练启动失败的配置组合。

[0053] 根据本发明实施例提出的用于大模型训练的参数自动搜索装置，通过获取参数配置文件，其包括大模型训练框架名称、多个参数和每个参数的参数区间，可以根据大模型训练框架名称确定目标模型训练框架，根据目标模型训练框架、多个参数和每个参数的参数区间确定所有配置组合的训练流程，启动所有配置组合的训练流程，并基于评估指标从所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合。由此，通过目标模型训练框架对参数配置组合进行枚举训练，即可得到最优参数配置组合，解决了当前确定最优参数配置的过程繁琐耗时，导致模型开发周期较长的问题，提高用户确定最优参数配置的效率，降低开发成本。

[0054] 为达到上述目的,本发明第三方面实施例提出一种电子设备,包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述程序,以实现如上述实施例所述的用于大模型训练的参数自动搜索方法。

[0055] 为达到上述目的,本发明第四方面实施例提出一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行,以用于实现如上述实施例所述的用于大模型训练的参数自动搜索方法。

[0056] 本发明附加的方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本发明的实践了解到。

附图说明

[0057] 本发明上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解,其中:

[0058] 图1为根据本发明实施例提供的一种用于大模型训练的参数自动搜索方法的流程图;

[0059] 图2为根据本发明的另一个实施例的用于大模型训练的参数自动搜索方法的流程图;

[0060] 图3为根据本发明实施例提供的用于大模型训练的参数自动搜索装置的方框示意图;

[0061] 图4为根据本发明实施例提供的电子设备的结构示意图。

具体实施方式

[0062] 下面详细描述本发明的实施例,实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,旨在用于解释本发明,而不能理解为对本发明的限制。

[0063] 下面参照附图描述根据本发明实施例提出的用于大模型训练的参数自动搜索方法、装置及电子设备。

[0064] 在介绍本发明实施例提出的用于大模型训练的参数自动搜索方法之前,先简单介绍相关技术中的可以实现高效并行策略的大模型训练框架。

[0065] 相关技术中,(1)大模型训练框架Megatron-LM基于深度学习框架PyTorch,实现了高效的并行策略,包括模型并行、数据并行以及流水线并行,同时还支持混合精度训练,可以提高计算性能并减小内存消耗;(2)主流的开源大模型训练框架Megatron-DeepSpeed采用了零冗余优化器内存优化技术,以降低模型训练的显存占用,同时在集群机器节点增加时依然有很好的可扩展性;(3)大模型训练框架通常无缝集成主流深度学习框架PyTorch,提供了高效的显存优化技术和模型并行策略接口,用户可以根据提供的接口调整合适的模型参数和并行策略。

[0066] 然而,目前主流的大模型训练框架往往都直接暴露接口给用户,用户在一次训练过程中可以手动调整模型的结构、参数量,以及模型训练时的并行策略,为了在特定的集群上获得最高的训练效率,用户往往需要在多个参数配置的排列组合之间多次尝试,并等待获取模型训练效率的数据进行记录,待所有排列组合的训练都尝试过后,再根据记录的数

据选择较为合适的训练参数配置,这个过程非常繁琐且耗时,并且当用户更换集群拓扑或者机器型号时,往往还需要重新手动搜索模型训练的最佳配置,导致模型开发周期较长。

[0067] 正是基于上述问题,本发明实施例提出一种用于大模型训练的参数自动搜索方法,通过目标模型训练框架对提供的参数配置组合进行枚举训练,并根据不同参数配置组合的运行情况实时获取显存占用情况、训练速度等指标,对不同参数配置组合的运行结果进行排序和过滤,帮助用户高效地搜索出给定硬件拓扑和模型结构下的最优参数配置及并行策略配置,减小用户在寻找训练配置上的开发成本。

[0068] 图1是本发明一个实施例的用于大模型训练的参数自动搜索方法的流程图。

[0069] 示例性的,如图1所示,该用于大模型训练的参数自动搜索方法包括以下步骤:

[0070] 在步骤S101中,获取参数配置文件,其中,参数配置文件包括大模型训练框架名称、用于排列组合的多个参数和每个参数的参数区间,其中,参数包括模型结构参数和并行训练参数。

[0071] 可以理解的是,参数配置文件可以由用户进行提供,其包括大模型训练框架名称、用于排列组合的多个参数和每个参数的参数区间,其中,参数包括模型结构参数和并行训练参数。其中,大模型训练框架指的是专门用于训练大规模深度学习模型的工具,其支持高效的大规模并行计算,可以处理大规模数据和模型,目前,主流的大模型训练框架包括:TensorFlow(由谷歌开发,支持分布式训练,具有强大的生态系统和广泛的社区支持)、PyTorch(由Facebook开发,具有简洁易用的API和灵活的动态图特性)和PaddlePaddle(飞桨,由PaddlePaddle开源社区开发,支持多种硬件平台和多种应用场景)等;为了在特定的模型配置下取得较好的搜索效果,用户可以提供用于排列组合的多个参数,常见的参数包括:(1)模型结构参数,如num-layers(大语言模型的层数)、hidden-size(隐藏层维度)、seq-length(输入序列长度)、micro-batch-size(单次训练所选取的样本数)和train-iters(训练迭代次数)等,(2)并行训练参数,包括但不限于:nproc-per-node(单机进程数,即调用的GPU(Graphics Processing Unit,图形处理器)数)、tensor-model-parallel-size(模型并行度,即将模型参数平均切分到所有GPU上的并行度)、pipeline-model-parallel-size(流水线并行度)、 sequence-parallel(序列并行度)等;此外,用户还需要提供每个参数的参数区间(取值范围或列表),如seq-length={1024, 2048, 4096}。

[0072] 在步骤S102中,根据大模型训练框架名称确定目标模型训练框架,并根据目标模型训练框架、多个参数和每个参数的参数区间确定所有配置组合的训练流程。

[0073] 也就是说,在步骤S101中提供大模型训练框架名称之后,便可以根据大模型训练框架名称确定目标模型训练框架,基于目标模型训练框架、多个参数和每个参数的参数区间即可确定关于所有配置组合的训练流程,其中,目标训练框架可以自动枚举所有参数的配置组合,作为单次训练的配置。

[0074] 为便于进一步理解,下面详细说明如何根据目标模型训练框架、多个参数和每个参数的参数区间确定所有配置组合的训练流程。

[0075] 作为一种可能实现的方式,根据目标模型训练框架、多个参数和每个参数的参数区间确定所有配置组合的训练流程,包括:从参数配置文件中获取每次训练的迭代次数;根据多个参数和每个参数的参数区间确定所有参数的配置组合;基于迭代次数和所有参数的配置组合确定所有配置组合的训练流程。

[0076] 具体而言,每次训练的迭代次数取决于参数配置文件中的total-iters(总的训练迭代次数)参数,根据多个参数和每个参数的参数区间可以确定所有参数的配置组合。例如,假设参数seq-length的参数区间为:seq-length={1024, 2048, 4096},在枚举seq-length的配置组合时可以分别用1024、2048和4096这三个数值依次与其他参数进行排列组合,作为以参数seq-length为主的运行一次训练模型的配置组合,即包括(1024,num-layers)、(1024,hidden-size)、……、(4096,pipeline-model-parallel-size)、(4096,sequence-parallel)等,利用该参数(seq-length)完成一次训练后,可以同理枚举以下一个参数为主的配置组合,开启下一次训练,由此,便可以根据多个参数和每个参数的参数区间确定所有参数的配置组合,并在得到所有参数的配置组合之后,基于迭代次数和所有参数的配置组合确定所有配置组合的训练流程,直到枚举完所有参数的配置组合。

[0077] 需要说明的是,总的训练次数为所有参数的参数区间的笛卡尔积的基数(即,笛卡尔积中元素的数量),其中,笛卡尔积又称直积,如数学领域的两个集合X和Y,集合X中的每个元素与集合Y中的每个元素分别构成有序对,所有有序对组成的集合叫做集合X和Y的笛卡尔积,即假设集合X={a, b},集合Y={0, 1, 2},则两个集合的笛卡尔积为{(a, 0), (a, 1), (a, 2), (b, 0), (b, 1), (b, 2)},两个以上集合的笛卡尔积同理,为避免冗余,此处不做赘述。

[0078] 在步骤S103中,启动所有配置组合的训练流程,并基于评估指标从所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合。

[0079] 也就是说,基于已经提供的参数和所有参数的配置组合,可以将训练流程拆分为多个并行任务,每个任务对应一个特定的参数的配置组合,将以每个参数为主的配置组合分配到多个计算节点进行并行处理,每个计算节点使用相应的参数的配置组合进行大模型训练,在每个节点完成训练后,即可得到所有配置组合的训练流程的训练结果,使用适当的评估指标(如准确率、交叉验证损失等)对所有训练结果进行综合评估,便可以确定用于大模型训练的最优参数组合。

[0080] 下面详细介绍如何基于评估指标从所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合。

[0081] 作为一种可能实现的方式,基于评估指标从所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合,包括:从参数配置文件中获取需保留的最优参数组合的数量;基于训练结果获取每个配置组合的评估指标值;基于最优参数组合的数量和每个配置组合的评估指标值,确定最优参数组合。

[0082] 具体而言,结合图2所示,每次启动一个训练实例,在训练过程中记录的指标,可以通过目标模型训练框架的报告模块将训练结果写入到报告文件中,报告文件可以记录用户提供的所有参数的配置组合的训练结果,训练结果由用户在参数配置文件中提供的评估指标进行衡量。使用目标模型训练框架进行模型训练时,目标模型训练框架可以在训练的最后一个步骤中启动一个进程,从标准输出中通过日志和正则表达式获取得到每个配置组合的评估指标值(如一个步骤的训练时间、显存占用、MFU(Model FLOPs Utilization,模型算力利用率)等)。其中,部分评估指标值如显存占用等可以通过例如NVTOP(NVidia TOP,英伟达显卡监控工具)等监控工具直接获取,如果是多显卡的情况还需要计算平均值;而对于无法直接获取的评估指标值,如TFLOPS(Tera Floating Point Operations Per Second,每

秒浮点运算多少万亿次)、MFU等,可以由目标模型训练框架进行计算得到;对于用户自定义(例如,在参数配置文件中自定义)的评估指标,其评估指标值可以通过目标模型训练框架调用可以实现的指标计算方式进行计算得到。

[0083] 可以理解的是,枚举完所有参数的配置组合之后,需要将训练结果的报告输出给用户,如果直接将所有配置组合的训练流程的训练结果输出给用户,报告过于冗长,难以从众多配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合,因此,还需要确定需保留的最优参数组合的数量。基于需保留的最优参数组合的数量,目标模型训练框架可以根据用户感兴趣的评估指标,先对所有配置组合的训练流程的训练结果进行排序和过滤,得到筛选后的最优配置组合集,再将最优配置组合集的训练结果的报告输出给用户,由此,算法用户可以基于最优参数组合的数量和每个配置组合的评估指标值,最终确定最优参数组合。

[0084] 举例来说,可以根据一个步骤的训练时间对所有配置组合的训练流程的训练结果进行排序,并根据显存占用大小或者硬件利用率对其进行进一步的过滤,从而筛选出符合要求的配置组合的训练流程的训练结果,进而确定最优参数组合。对于用户自定义的指标,用户可以在实现指标计算方式的同时,对训练结果进行过滤和排序。

[0085] 此外,在一些实施例中,在启动所有配置组合的训练流程时,还包括:记录训练启动失败的配置组合。

[0086] 也就是说,如果因参数太多或者并行策略配置不合理导致机器显存不足等,使得配置组合的训练流程启动失败,可以将训练配置启动失败的配置组合记录下来,便于后续筛选配置,且对于导致机器显存不足或其他训练失败的参数,可以直接进行剔除。

[0087] 进一步地,在一些实施例中,在获取参数配置文件之后,还包括:识别在参数配置文件中的多个参数中未给出参数区间的目标参数;获取目标参数的默认参数区间,并将默认参数区间作为目标参数的参数区间。

[0088] 需要说明的是,若某个参数用户未提供其参数区间或者未在该参数配置文件中列出,则可以对多个参数中未给出参数区间的目标参数进行识别,并获取用户启动训练时的目标参数的默认参数区间,将默认参数区间作为目标参数的参数区间即可。

[0089] 进一步地,在一些实施例中,在根据大模型训练框架名称确定目标模型训练框架之后,还包括:利用目标模型训练框架,校验多个参数中是否存在不满足预设兼容条件的不兼容参数;若多个参数中存在不满足预设兼容条件的不兼容参数,则针对不兼容参数进行报错提醒。

[0090] 可以理解的是,不同大模型训练框架的参数配置文件中的参数名称、格式之间存在差异,为了兼容多个主流的大模型训练框架,在根据大模型训练框架名称确定目标模型训练框架之后,可以利用目标模型训练框架对多个参数中是否存在不满足预设兼容条件的不兼容参数进行校验,其中,在定义和实施预设兼容条件时,确保这些兼容条件是合理且必要的,随着技术和业务需求的变化,预设兼容条件可以随时进行调整与更新,此处不做具体限定。当多个参数中存在不满足预设兼容条件的不兼容参数时,可以针对不兼容参数进行报错提醒,便于用户及时了解情况。

[0091] 举例来说,零冗余优化器ZeRO的阶段zero-stage参数为2时,与流水线并行pipeline-model-parallel-size参数不兼容,当设置了不兼容的参数时应当在框架层面提

前报错以提醒用户。

[0092] 进一步地,在一些实施例中,在针对不兼容参数进行报错提醒之后,还包括:接收用户针对不兼容参数反馈的参数修改指令;基于参数修改指令修改不兼容参数。

[0093] 也就是说,在针对不兼容参数进行报错提醒之后,用户可以在启动枚举训练参数前,基于不兼容参数发出参数修改指令,以对不兼容参数和期望指标进行修改,例如,如果设置了零冗余优化器的阶段zero-stage参数为2,那么流水线并行度参数pipeline-parallel-size只能设置为0,表示不开启流水线并行。

[0094] 根据本发明实施例提出的用于大模型训练的参数自动搜索方法,通过获取参数配置文件,其包括大模型训练框架名称、多个参数和每个参数的参数区间,可以根据大模型训练框架名称确定目标模型训练框架,根据目标模型训练框架、多个参数和每个参数的参数区间确定所有配置组合的训练流程,启动所有配置组合的训练流程,并基于评估指标从所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合。由此,通过目标模型训练框架对参数配置组合进行枚举训练,即可得到最优参数配置组合,解决了当前确定最优参数配置的过程繁琐耗时,导致模型开发周期较长的问题,提高用户确定最优参数配置的效率,降低开发成本。

[0095] 其次参照附图描述根据本发明实施例提出的用于大模型训练的参数自动搜索装置。

[0096] 图3是本发明一个实施例的用于大模型训练的参数自动搜索装置的方框示意图。

[0097] 如图3所示,该用于大模型训练的参数自动搜索装置10包括:获取模块100、第一确定模块200和第二确定模块300。

[0098] 其中,获取模块100,用于获取参数配置文件,其中,参数配置文件包括大模型训练框架名称、用于排列组合的多个参数和每个参数的参数区间,其中,参数包括模型结构参数和并行训练参数;

[0099] 第一确定模块200,用于根据大模型训练框架名称确定目标模型训练框架,并根据目标模型训练框架、多个参数和每个参数的参数区间确定所有配置组合的训练流程;

[0100] 第二确定模块300,用于启动所有配置组合的训练流程,并基于评估指标从所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合。

[0101] 进一步地,在一些实施例中,在获取参数配置文件之后,获取模块100,还用于:

[0102] 识别在参数配置文件中的多个参数中未给出参数区间的目标参数;

[0103] 获取目标参数的默认参数区间,并将默认参数区间作为目标参数的参数区间。

[0104] 进一步地,在一些实施例中,在根据大模型训练框架名称确定目标模型训练框架之后,第一确定模块200,还包括:

[0105] 校验单元,用于利用目标模型训练框架,校验多个参数中是否存在不满足预设兼容条件的不兼容参数;

[0106] 报错单元,用于在多个参数中存在不满足预设兼容条件的不兼容参数时,针对不兼容参数进行报错提醒。

[0107] 进一步地,在一些实施例中,在针对不兼容参数进行报错提醒之后,报错单元,还用于:

[0108] 接收用户针对不兼容参数反馈的参数修改指令;

[0109] 基于参数修改指令修改不兼容参数。

[0110] 进一步地,在一些实施例中,第一确定模块200,具体用于:

[0111] 从参数配置文件中获取每次训练的迭代次数;

[0112] 根据多个参数和每个参数的参数区间确定所有参数的配置组合;

[0113] 基于迭代次数和所有参数的配置组合确定所有配置组合的训练流程。

[0114] 进一步地,在一些实施例中,第二确定模块300,具体用于:

[0115] 从参数配置文件中获取需保留的最优参数组合的数量;

[0116] 基于训练结果获取每个配置组合的评估指标值;

[0117] 基于最优参数组合的数量和每个配置组合的评估指标值,确定最优参数组合。

[0118] 进一步地,在一些实施例中,在启动所有配置组合的训练流程时,第二确定模块300,还用于:

[0119] 记录训练启动失败的配置组合。

[0120] 需要说明的是,前述对用于大模型训练的参数自动搜索方法实施例的解释说明也适用于该实施例的用于大模型训练的参数自动搜索装置,此处不再赘述。

[0121] 根据本发明实施例提出的用于大模型训练的参数自动搜索装置,通过获取参数配置文件,其包括大模型训练框架名称、多个参数和每个参数的参数区间,可以根据大模型训练框架名称确定目标模型训练框架,根据目标模型训练框架、多个参数和每个参数的参数区间确定所有配置组合的训练流程,启动所有配置组合的训练流程,并基于评估指标从所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合。由此,通过目标模型训练框架对参数配置组合进行枚举训练,即可得到最优参数配置组合,解决了当前确定最优参数配置的过程繁琐耗时,导致模型开发周期较长的问题,提高用户确定最优参数配置的效率,降低开发成本。

[0122] 图4为本发明实施例提供的电子设备的结构示意图。该电子设备可以包括:

[0123] 存储器401、处理器402及存储在存储器401上并可在处理器402上运行的计算机程序。

[0124] 处理器402执行程序时实现上述实施例中提供的用于大模型训练的参数自动搜索方法。

[0125] 进一步地,电子设备还包括:

[0126] 通信接口403,用于存储器401和处理器402之间的通信。

[0127] 存储器401,用于存放可在处理器402上运行的计算机程序。

[0128] 存储器401可能包含高速RAM(Random Access Memory,随机存取存储器)存储器,也可能还包括非易失性存储器,例如至少一个磁盘存储器。

[0129] 如果存储器401、处理器402和通信接口403独立实现,则通信接口403、存储器401和处理器402可以通过总线相互连接并完成相互间的通信。总线可以是ISA(Industry Standard Architecture,工业标准体系结构)总线、PCI(Peripheral Component Interconnect,外部设备互连)总线或EISA(Extended Industry Standard Architecture,扩展工业标准体系结构)总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示,图4中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0130] 可选的,在具体实现上,如果存储器401、处理器402及通信接口403,集成在一块芯

片上实现,则存储器401、处理器402及通信接口403可以通过内部接口完成相互间的通信。

[0131] 处理器402可能是一个CPU(Central Processing Unit,中央处理器),或者是ASIC(Application Specific Integrated Circuit,特定集成电路),或者是被配置成实施本发明实施例的一个或多个集成电路。

[0132] 本发明实施例还提供一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现如上的用于大模型训练的参数自动搜索方法。

[0133] 此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本发明的描述中,“多个”的含义是至少两个,例如两个,三个等,除非另有明确具体的限定。

[0134] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不是必须针对的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0135] 尽管上面已经示出和描述了本发明的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本发明的限制,本领域的普通技术人员在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

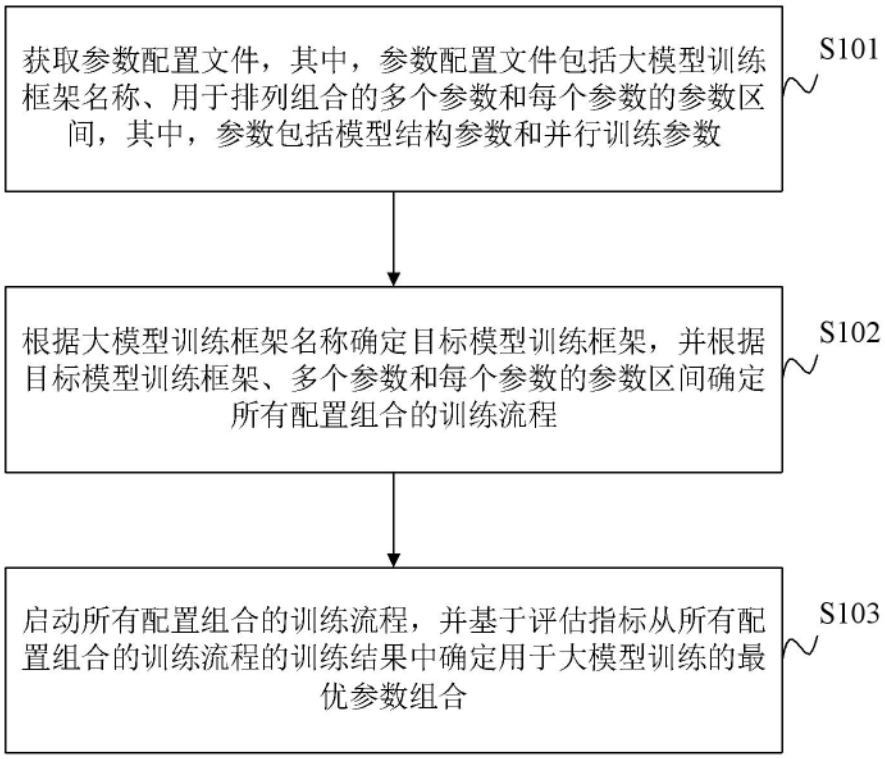

获取参数配置文件，其中，参数配置文件包括大模型训练框架名称、用于排列组合的多个参数和每个参数的参数区间，其中，参数包括模型结构参数和并行训练参数
S101
根据大模型训练框架名称确定目标模型训练框架，并根据目标模型训练框架、多个参数和每个参数的参数区间确定所有配置组合的训练流程
S102
启动所有配置组合的训练流程，并基于评估指标从所有配置组合的训练流程的训练结果中确定用于大模型训练的最优参数组合
S103

图1

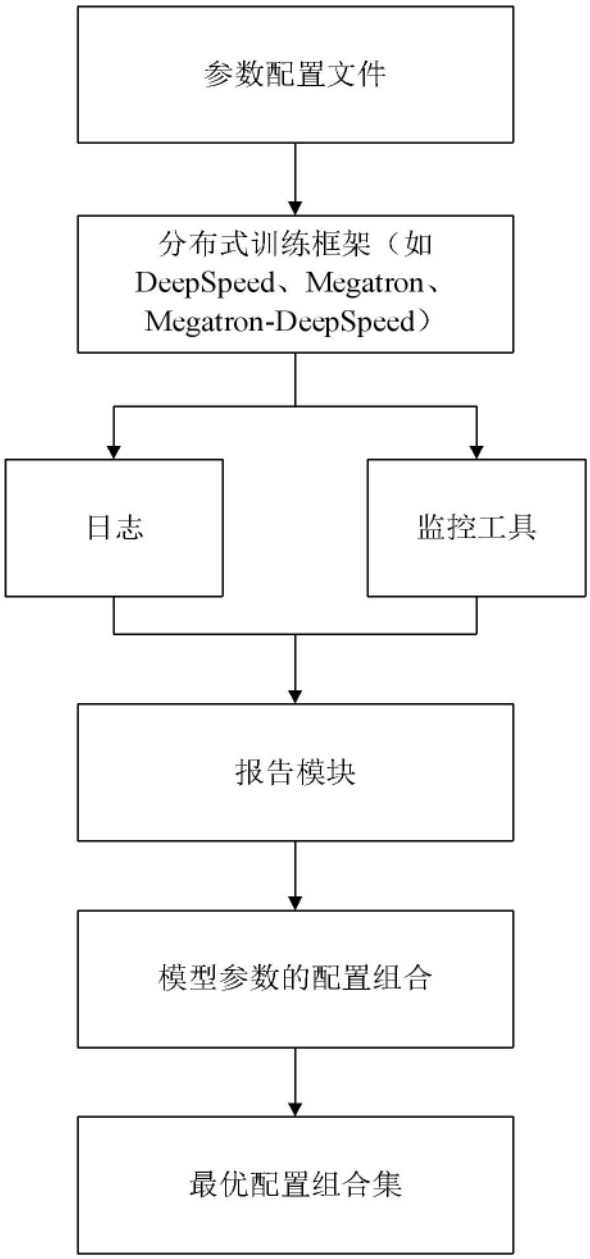

参数配置文件
分布式训练框架（如DeepSpeed、Megatron、Megatron-DeepSpeed）
日志
监控工具
报告模块
模型参数的配置组合
最优配置组合集

图2

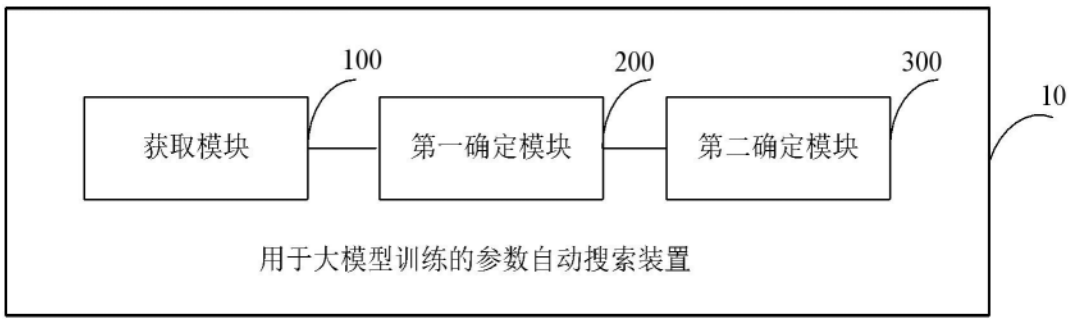

100
200
300
10
获取模块
第一确定模块
第二确定模块
用于大模型训练的参数自动搜索装置

图3

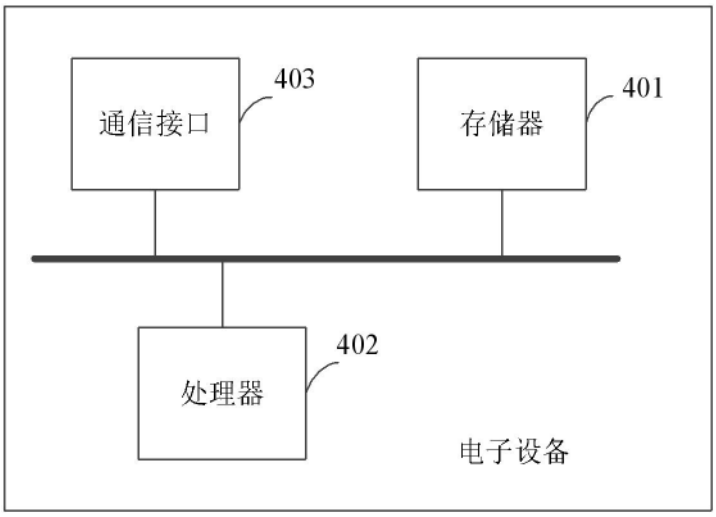


图4