



(12) 发明专利

(10) 授权公告号 CN 113657577 B

(45) 授权公告日 2023. 08. 18

(21) 申请号 202110827079.5

(22) 申请日 2021.07.21

(65) 同一申请的已公布的文献号
申请公布号 CN 113657577 A

(43) 申请公布日 2021.11.16

(73) 专利权人 阿里巴巴达摩院(杭州)科技有限
公司

地址 310023 浙江省杭州市余杭区五常街
道文一西路969号3幢5层516室

(72) 发明人 杨健邦 陈榕 王磊 尹强
于文渊 周靖人

(74) 专利代理机构 北京太合九思知识产权代理
有限公司 11610

专利代理师 刘戈 张爱

(51) Int. Cl.

G06N 3/04 (2023.01)

G06N 3/08 (2023.01)

G06V 10/774 (2022.01)

G06V 10/82 (2022.01)

(56) 对比文件

US 2020372355 A1, 2020.11.26

WO 2020182989 A1, 2020.09.17

CN 110751275 A, 2020.02.04

CN 111177433 A, 2020.05.19

CN 112101538 A, 2020.12.18

CN 112381216 A, 2021.02.19

CN 112383516 A, 2021.02.19

CN 112508181 A, 2021.03.16

CN 112651488 A, 2021.04.13

CN 112862093 A, 2021.05.28

CN 112966763 A, 2021.06.15

US 2017061276 A1, 2017.03.02

US 2019355084 A1, 2019.11.21

US 2021200806 A1, 2021.07.01

WO 2021082681 A1, 2021.05.06

呼延康; 樊鑫; 余乐天; 罗钟铤. 图神经网络
回归的人脸超分辨率重建. 软件学报. 2017, (第
04期), 全文.

审查员 钱坤

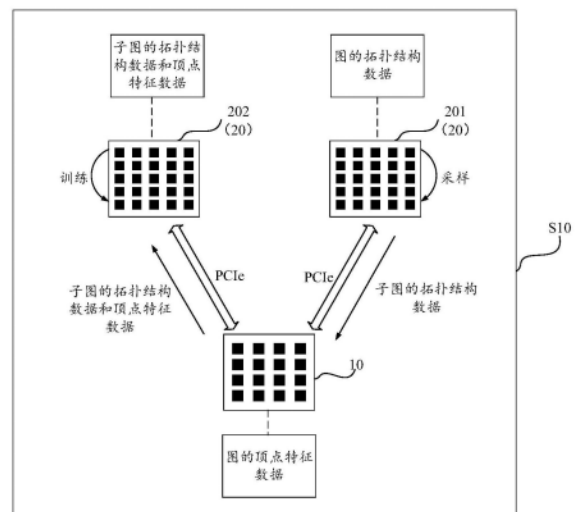
权利要求书3页 说明书17页 附图4页

(54) 发明名称

模型训练方法及计算系统

(57) 摘要

本申请实施例提供一种模型训练方法及计算系统。在本申请实施例中,将数据量较小的图拓扑结构数据存放到采样用并行处理单元,而数据量较大的图顶点特征数据存放到处理单元的内存中。利用并行处理单元的高速计算性能对图拓扑结构数据进行采样,提高了图采样效率,处理单元可基于采样用并行处理单元采样的子图的拓扑结构数据中的顶点标识,从内存存储的图顶点特征数据中获取子图的顶点特征数据;训练用并行处理单元可利用子图的拓扑结构数据和顶点特征数据进行图神经网络训练。对于训练用并行处理单元来说,图采样效率提高,可降低训练用并行处理单元等待采样结果的时间,有助于提高训练用并行处理单元利用率。



CN 113657577 B

1. 一种计算系统,其特征在于,包括:处理单元和多个并行处理单元;所述处理单元与所述多个并行处理单元之间通信连接;

所述处理单元,用于将获取的图数据集存储至所述处理单元的内存中,并控制所述多个并行处理单元中的第一并行处理单元从所述处理单元的内存中读取所述图数据集中的图拓扑结构数据;

所述第一并行处理单元,用于对所述图拓扑结构数据进行采样,得到子图的拓扑结构数据;

所述处理单元,还用于控制所述多个并行处理单元中的第二并行处理单元获取所述子图的拓扑结构数据;以及,基于所述子图的拓扑结构数据中的顶点标识,从所述图数据集中的顶点特征数据中获取所述子图的顶点特征数据;并控制所述第二并行处理单元从所述处理单元的内存中读取所述子图的顶点特征数据;

所述第二并行处理单元,用于以所述子图的拓扑结构数据和顶点特征数据为样本,进行图神经网络训练。

2. 根据权利要求1所述的系统,其特征在于,所述处理单元具体用于:

从所述第一并行处理单元中读取所述子图的拓扑结构数据,并存储至所述处理单元的内存中;

控制所述第二并行处理单元从所述处理单元的内存中读取所述子图的拓扑结构数据。

3. 根据权利要求1所述的系统,其特征在于,所述多个并行处理单元之间通过通信连接;

所述处理单元,具体用于:控制所述第二并行处理单元从所述第一并行处理单元中读取所述子图的拓扑结构数据。

4. 根据权利要求1-3任一项所述的系统,其特征在于,所述处理单元还用于:

在所述第一并行处理单元从所述处理单元的内存中读取所述图拓扑结构数据之后,释放与所述图拓扑结构数据对应的处理单元的内存区域。

5. 根据权利要求4所述的系统,其特征在于,所述处理单元为至少两个,并且其中:

所述至少两个处理单元中的第一处理单元执行以下数据控制操作:控制第一并行处理单元从处理单元的内存中读取图数据集中的图拓扑结构数据;控制所述第一并行处理单元对所述图拓扑结构数据采样,得到子图的拓扑结构数据;以及,控制第二并行处理单元获取所述子图的拓扑结构数据和顶点特征数据;控制所述第二并行处理单元以所述子图的拓扑结构数据和顶点特征数据为样本进行图神经网络训练;

所述至少两个处理单元中的第二处理单元执行基于所述子图的拓扑结构数据中的顶点标识从所述图数据集中的顶点特征数据中获取所述子图的顶点特征数据的计算操作。

6. 一种模型训练方法,其特征在于,包括:

将获取的图数据集存储至处理单元的内存中;

控制第一并行处理单元从处理单元的内存中读取图数据集中的图拓扑结构数据;

控制所述第一并行处理单元对所述图拓扑结构数据采样,得到子图的拓扑结构数据;

基于所述子图的拓扑结构数据中的顶点标识,从所述处理单元的内存存储的图数据集中的顶点特征数据中获取所述子图的顶点特征数据;

控制第二并行处理单元获取所述子图的拓扑结构数据和顶点特征数据;

控制所述第二并行处理单元以所述子图的拓扑结构数据和顶点特征数据为样本进行图神经网络训练。

7. 根据权利要求6所述的方法,其特征在在于,包括:

启动多个线程;利用所述多个线程循环执行以下步骤,直至所述图神经网络满足设定的条件:

利用第一线程控制所述第一并行处理单元对所述图拓扑结构数据采样,得到子图的拓扑结构数据;

利用所述第一线程和第二线程基于所述子图的拓扑结构数据中的顶点标识,从所述处理单元的内存存储的顶点特征数据中获取所述子图的顶点特征数据;

利用所述第二线程控制所述第二并行处理单元获取所述子图的拓扑结构数据和顶点特征数据;

利用第三线程控制所述第二并行处理以所述子图的拓扑结构数据和顶点特征数据为样本进行图神经网络训练。

8. 根据权利要求7所述的方法,其特征在在于,所述顶点标识为顶点索引;所述利用所述第一线程和第二线程基于所述子图的拓扑结构数据中的顶点标识,从所述处理单元的内存中存储的图数据集中的顶点特征数据中获取所述子图的顶点特征数据,包括:

利用所述第一线程控制所述第一并行处理单元按照从0递增的方式对所述子图的拓扑结构数据中的第一顶点索引进行重新编号,得到所述子图的第二顶点索引;

利用所述第一线程控制所述第一并行处理单元建立所述第一顶点索引与所述第二顶点索引之间的对应关系;

利用第二线程根据所述对应关系,从所述处理单元的内存存储的顶点特征数据中获取所述子图的顶点特征数据。

9. 根据权利要求8所述的方法,其特征在在于,还包括:

利用所述第二线程按照所述对应关系中所述第二顶点索引的顺序,将所述子图的顶点特征数据存储至所述处理单元的连续的内存区域。

10. 根据权利要求9所述的方法,其特征在在于,所述利用第二线程控制所述第二并行处理单元获取所述子图的拓扑结构数据和所述子图的顶点特征数据,包括:

利用所述第二线程控制所述第二并行处理单元读取所述子图的拓扑结构数据;并存储至所述第二并行处理单元的内存中;

利用所述第二线程控制所述第二并行处理单元从所述连续的内存区域中读取存储的子图的顶点特征数据,并存储至所述第二并行处理单元的内存中。

11. 根据权利要求6-10任一项所述的方法,其特征在在于,还包括:

在所述第一并行处理单元从所述处理单元的内存中读取图数据集中的图拓扑结构数据之后,释放所述图拓扑结构数据对应的处理单元的内存区域。

12. 根据权利要求11所述的方法,其特征在在于,所述处理单元为至少两个;所述方法包括:

所述至少两个处理单元中的第一处理单元执行以下数据控制操作:

控制第一并行处理单元从处理单元的内存中读取图数据集中的图拓扑结构数据;

控制所述第一并行处理单元对所述图拓扑结构数据采样,得到子图的拓扑结构数据;

以及,控制第二并行处理单元获取所述子图的拓扑结构数据和顶点特征数据;

控制所述第二并行处理单元以所述子图的拓扑结构数据和顶点特征数据为样本进行图神经网络训练;

所述至少两个处理单元中的第二处理单元执行所述基于所述子图的拓扑结构数据中的顶点标识,从所述处理单元的内存存储的图数据集中的顶点特征数据中获取所述子图的顶点特征数据的计算操作。

模型训练方法及计算系统

技术领域

[0001] 本申请涉及计算机技术领域,尤其涉及一种模型训练方法及计算系统。

背景技术

[0002] 图神经网络是一种将图结构数据和深度学习技术相结合的神经网络模型,在点分类以及用于个性化推荐的边预测等领域应用广泛。

[0003] 在实际中,用于图神经网络训练的图数据的规模比较大,一般用于图神经网络训练的硬件设备包括通用处理单元和并行处理单元。通用处理单元一般是进行常规串行处理操作的计算单元,例如,中央处理器CPU。并行处理单元常具有较大的并行处理能力,一般用于处理大规模数据流,例如,图形计算或图计算。并行处理单元的示例可以是图形处理单元,人工智能加速单元,等具有高速并行计算能力的计算单元,并广泛应用于深度学习模型训练中。

[0004] 本案的研究者注意到,现有图神经网络模型训练技术中,考虑到并行处理单元的内存相对较小但并行计算能力强而通用处理单元的内存相对较大但并行计算能力相对弱,一般采用通用处理单元对规模大的原始图进行采样得到规模较小的子图;采用并行处理单元对子图进行训练。但由于通用处理单元的并行计算性能比并行计算单元弱,导致并行计算单元的大部分时间在等待通用处理单元的采样计算结果的空闲状态,并行计算单元利用率较低。

发明内容

[0005] 本申请的多个方面提供一种模型训练方法及计算系统,用以提高图采样效率,有助于提高并行处理单元利用率。

[0006] 本申请实施例提供一种计算系统,包括:中央处理器处理单元和多个图形处理器并行处理单元;所述处理单元与所述多个并行处理单元之间通信连接;

[0007] 所述处理单元,用于将获取的图数据集存储至处理单元的内存中;并控制所述多个并行处理单元中的第一并行处理单元从至少一个处理单元的内存中读取将所述图数据集中的图拓扑结构数据;

[0008] 所述第一并行处理单元,用于对所述图拓扑结构数据进行采样,得到子图的拓扑结构数据;

[0009] 所述至少一个处理单元,还用于控制所述多个并行处理单元中的第二并行处理单元获取所述子图的拓扑结构数据;以及,基于所述子图的拓扑结构数据中的顶点标识,从所述图数据集中的顶点特征数据中获取所述子图的顶点特征数据;并控制所述第二并行处理单元从至少一个处理单元的内存中读取所述子图的顶点特征数据;

[0010] 所述第二并行处理单元,用于以所述子图的拓扑结构数据和顶点特征数据为样本,进行图神经网络训练。

[0011] 本申请实施例还提供一种模型训练方法,适用于处理单元,包括:

- [0012] 将获取的图数据集存储至所述处理单元的内存中；
- [0013] 控制第一并行处理单元从处理单元的内存中读取图数据集中的图拓扑结构数据；
- [0014] 控制所述第一并行处理单元对所述图拓扑结构数据采样，得到子图的拓扑结构数据；
- [0015] 基于所述子图的拓扑结构数据中的顶点标识，从所述处理单元的内存存储的图数据集中的顶点特征数据中获取所述子图的顶点特征数据；
- [0016] 控制第二并行处理单元获取所述子图的拓扑结构数据和所述子图的顶点特征数据；
- [0017] 控制所述第二并行处理单元以所述子图的拓扑结构数据和顶点特征数据为样本进行图神经网络训练。
- [0018] 在本申请实施例中，将数据量较小的图拓扑结构数据存放到采样用并行处理单元，而数据量较大的图顶点特征数据存放到处理单元的内存中。采样用并行处理单元可对图拓扑结构数据进行采样，得到子图的拓扑结构数据；处理单元可根据子图的拓扑结构数据中的顶点标识，从内存存储的图顶点特征数据中获取子图的顶点特征数据；并将子图的拓扑结构数据和子图的顶点特征数据存储至训练用并行处理单元；训练用并行处理单元可以子图的拓扑结构数据和子图的顶点特征数据为样本进行图神经网络训练。在本申请实施例中，利用处理单元的内存大的性能，对数据量较大的图顶点数据进行存储，利用并行处理单元的高速计算性能对图拓扑结构数据进行采样，提高了图采样效率。对于训练用并行处理单元来说，图采样效率提高，可降低训练用并行处理单元等待采样结果的时间，有助于提高训练用并行处理单元利用率。

附图说明

- [0019] 此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：
- [0020] 图1a和图1b为本申请实施例提供的计算系统的结构示意图；
- [0021] 图1c为本申请实施例提供的模型训练过程示意图；
- [0022] 图2为本申请实施例提供的模型训练过程时序示意图；
- [0023] 图3和图4为本申请实施例提供的模型训练方法的流程示意图。

具体实施方式

- [0024] 为使本申请的目的、技术方案和优点更加清楚，下面将结合本申请具体实施例及相应的附图对本申请技术方案进行清楚、完整地描述。显然，所描述的实施例仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。
- [0025] 为了解决上述并行计算单元利用率低的技术问题，在本申请一些实施例中，将图拓扑结构数据存放到采样用并行处理单元，而图顶点特征数据存放到通用处理单元（以下简称：处理单元）的内存中。采样用并行处理单元可对图拓扑结构数据进行采样，得到子图的拓扑结构数据；处理单元可根据子图的拓扑结构数据中的顶点标识，从内存存储的图顶点特征数据中获取子图的顶点特征数据；并将子图的拓扑结构数据和子图的顶点特征数据

存储至训练用并行处理单元;训练用并行处理单元可以子图的拓扑结构数据和子图的顶点特征数据为样本进行图神经网络训练。在本申请实施例中,利用处理单元的内存大的性能对数据量较大的图顶点数据进行存储,利用并行处理单元的高速计算性能对图拓扑结构数据进行采样,提高了图采样效率。对于训练用并行处理单元来说,图采样效率提高,可降低训练用并行处理单元等待采样结果的时间,有助于提高训练用并行处理单元利用率。

[0026] 以下结合附图,详细说明本申请各实施例提供的技术方案。

[0027] 应注意到:相同的标号在下面的附图以及实施例中表示同一物体,因此,一旦某一物体在一个附图或实施例中定义,则在随后的附图和实施例中不需要对其进行进一步讨论。

[0028] 图1a和图1b为本申请实施例提供的计算系统的结构示意图。如图1a和图1b所示,该计算系统S10包括:处理单元10和多个并行处理单元20。在本实施例中,不限定处理单元10的数量。处理单元10可以为至少1个,即1个或多个;每个处理单元10可以为单核处理单元,也可为多核处理单元。上述多个是指2个或2个以上。图1a和图1b中仅以处理单元10的数量为1个;并行处理单元20的数量为2个进行图示,但不构成限定。

[0029] 在本实施例中,处理单元10可以为任何具有处理计算能力的处理器件。处理单元10可为串行处理单元,也可为并行处理单元。例如,处理单元10可以为通用处理器,如CPU等。并行处理单元20是指可并行计算处理的处理器件。例如,并行处理单元20可以为GPU、FPGA等。在本实施例中,处理单元10的内存大于并行处理单元20的内存。

[0030] 其中,处理单元10与多个并行处理单元20之间通信连接。对于多个并行处理单元20之间,在一些实施例中,如图1a所示,多个并行处理单元20之间不直接通信连接,不同并行处理单元20之间的通信可通过处理单元10进行数据转发。在另一些实施例中,如图1b所示,多个并行处理单元20之间可直接通信连接。不同并行处理单元20之间可直接进行数据传输,也可通过处理单元10进行数据转发。在一些实施例中,处理单元10与多个并行处理单元20设置于同一物理机,处理单元10与多个并行处理单元20可通过通信接口连接。在该实施例中,对于不同并行处理单元20之间直接通信连接的情况,不同并行处理单元20之间可通过通信接口连接。在另一些实施例中,处理单元10与多个并行处理单元20可设置于不同的物理机,处理单元10与多个并行处理单元20之间可通过网络通信连接。多个并行处理单元20之间也可通过网络通信连接。例如,处理单元10与多个并行处理单元20可设置于云服务器中,通过网络通信连接;等等。图1a和图1b仅以处理单元10与多个并行处理单元20设置于同一物理机进行图示,但不构成限定。

[0031] 上述通信接口可为PCIe接口、PCI接口或USB接口等。优选地,通信接口为PCIe接口,可提高处理单元10与并行处理单元20之间以及不同并行处理单元20之间的数据传输速率。对于不同并行处理单元20之间的通信接口还可为NVLINK接口。图1a和图1b仅以通信接口为PCIe接口进行图示,但不构成限定。

[0032] 在本实施例中,如图1a和图1b所示,处理单元10可获取图数据集,并将获取的图数据集存储至处理单元10的内存中。其中,用于存储图数据集的处理单元可为处理单元10中的全部处理单元,也可为部分处理单元,如可以为处理单元10中的某一个处理单元等。在本实施例中,不限定处理单元10获取图数据集的实现形式。在一些实施例中,处理单元10可从外部存储介质中读取图数据集,或者,从图数据库中读取图数据集,或者,接收其它设备发

送的图数据集等等。

[0033] 在本实施例中,图数据集可为原始图的数据集。其中,图(Graph)是用于表示对象之间关联关系的一种抽象数据结构,使用顶点(Vertex)和边(Edge)进行描述。顶点表示对象,边表示对象之间的关系。图数据集则是可抽象成用图描述的数据,可包括:图拓扑结构数据和图的顶点特征数据。其中,图拓扑结构数据是指描述图的结构特征的数据,例如表示顶点之间关系的数据、点边关系的数据以及边的类型和权重等。图拓扑结构数据可实现为邻接矩阵、邻接链表、压缩稀疏行(Compressed Sparse Row,CSR)矩阵、坐标(Coordinate,C00)矩阵或压缩稀疏列(Compressed Sparse Column,CSC)矩阵等数据结构。

[0034] 顶点特征数据是指描述顶点代表的对象的特征的数据,一般为高维度数据,可以向量或矩阵等数据结构表示。例如,在一些实施例中,顶点表示用户,则顶点特征数据为描述用户特征的数据,可包括用户个人信息、行为习惯等等。

[0035] 在本申请实施例中,图神经网络是一种将图结构数据和深度学习技术相结合的神经网络模型,在点分类以及用于个性化推荐的边预测等领域应用广泛。在实际应用中,在使用图神经网络之前,需要以大量的图数据为样本对图神经网络进行模型训练。但是,用于图神经网络训练的图数据的规模较大,顶点和边的规模达到了数亿甚至数十亿以上。训练如此大规模的图数据需要大量的计算,训练时间较长。传统的深度学习训练一般采用并行处理单元来进行训练加速。但是,由于对大规模的图数据进行训练需要使用大量的内存,而并行处理单元的内存比较小,无法对大规模的图数据进行训练。

[0036] 图采样训练是用来解决大规模图数据训练的一种常用的方法。图采样训练可先通过图采样从原始的大图中得到一个规模较小的子图,这个小规模的子图是可存储在并行处理单元的内存中进行训练的。因此,图采样训练是一种适合使用并行处理单元加速的图神经网络训练方法。

[0037] 目前,现有的针对大规模图数据的图采样训练方法考虑到并行处理单元的内存相对较小,而处理单元的内存相对较大,采用处理单元采样与并行处理单元训练相结合的采样训练的训练方案。但由于处理单元的计算性能比并行处理单元弱,处理单元上的采样计算占据了整个训练的主要的时间,并行处理单元的大部分时间在等待处理单元的采样计算结果,利用率不高。

[0038] 为了解决上述问题,在本实施例中,针对包括多个并行处理单元20的计算系统,将部分并行处理单元20用于图采样;部分并行处理单元20用于图神经网络训练。在本申请实施例中,为了便于描述和区分,将用于图采样的并行处理单元称为采样用并行处理单元201;将用于训练的并行处理单元称为训练用并行处理单元202。其中,采样用并行处理单元201和训练用并行处理单元202都可为1个或多个。图1a和图1b仅以采样用并行处理单元201和训练用并行处理单元202为1个进行图示,但不构成限定。

[0039] 由于并行处理单元20的内存较小,无法将图数据集存储至并行处理单元20的内存中。考虑到图拓扑结构数据的数据量较小,顶点特征数据的数据量较大,因此,在本实施例中,处理单元10可控制采样用并行处理单元201从至少一个处理单元10的内存中读取图数据集中的图拓扑结构数据;采样用并行处理单元201将读取的图拓扑结构数据存储至采样用并行处理单元201的内存中。

[0040] 值得说明的是,上述存储图数据集的处理单元10与控制采样用并行处理单元201

读取图拓扑结构数据的处理单元可以为同一处理单元,也可为不同的处理单元,在本申请实施例中不进行限定。当然,上述采样用并行处理单元201读取图拓扑结构数据的处理单元为至少一个处理单元10存储有图数据集的处理单元。

[0041] 可选地,图数据集中的数据携带有数据类型标识;数据类型标识用于标识该数据属于拓扑结构数据,还是顶点特征数据。相应地,处理单元10可根据图数据集中的数据类型标识,从图数据集中识别出图拓扑结构数据和顶点特征数据。进一步,处理单元10可向采样用并行处理单元201发送读数据指令1;该读数据指令1可包括待读取数据的存储位置及数据长度。其中,处理单元的标识是指存储有图数据集的处理单元的标识。待读取的图拓扑结构数据的存储位置可包括:处理单元的标识及待读取数据在处理单元的内存空间位置。在该实施例中,待读取数据为图结构数据。待读取数据在处理单元的内存空间位置为图结构数据在处理单元的内存空间位置;待读取数据的数据长度为图拓扑结构数据的数据长度。采样用并行处理单元201,可响应于读数据指令1,从处理单元的标识对应的处理单元10中从待读取数据在处理单元的内存空间位置开始,读取上述数据长度的数据,作为图拓扑结构数据。

[0042] 对于处理单元10来说,在采样用并行处理单元201从处理单元10的内存中读取图数据集中的图拓扑结构数据之后,可继续保存图拓扑结构数据,也可释放与图拓扑结构数据对应的处理单元的内存区域。在本实施例中,处理单元10继续保存处理单元10的内存中的图数据集中的顶点特征数据。

[0043] 对于处理单元10在采样用并行处理单元201从处理单元10的内存中读取图数据集中的图拓扑结构数据之后,释放图拓扑结构数据对应的处理单元的内存区域的实施方式,可实现图拓扑结构数据与图的顶点特征数据分离存储的方法,可解决大规模图数据的数据量大而无法全部放进并行处理单元的问题,可直接利用并行处理单元来加速采样计算的过程。

[0044] 对于采样用并行处理单元201来说,如图1c所示,可对图拓扑结构数据进行采样,得到子图的拓扑结构数据。处理单元10可启动一线程,利用该线程控制采样用并行处理单元201对图拓扑结构数据进行采样,得到子图的拓扑结构数据。子图的规模大小可根据训练用并行处理单元202的内存以及实际需求确定。即子图的拓扑结构数据和顶点特征数据的总数据量小于或等于训练用并行处理单元202的内存,这样在后续训练用并行处理单元202才可利用子图进行图神经网络训练。

[0045] 在本实施例中,不限定采样用并行处理单元201可对图拓扑结构数据进行采样的具体实施方式。下面以几种实施方式为例,对采样用并行处理单元201可对图拓扑结构数据进行采样进行示例性说明。

[0046] 实施方式1:采样用并行处理单元201可从图拓扑结构数据中的部分顶点出发,利用随机采样算法对图拓扑结构数据进行采样,得到子图的拓扑结构数据。具体地,采样用并行处理单元201可随机选择M个顶点A;并从M个顶点A出发,从顶点A的邻居顶点中随机选择N个邻居顶点B;之后,从N个邻居顶点B出发,从N个邻居顶点的邻居顶点中随机选择K个邻居顶点C,以此往外拓展至遍历到图拓扑结构数据中无邻居的顶点,或者至选择出的子图的数据量达到设定数据量。其中,上述M、N和K均为大于或等于1的整数。可选地, $M < N < K$ 。

[0047] 实施方式2:对于加权图,采样用并行处理单元201可从图拓扑结构数据中的部分

顶点出发,利用权重采样算法对图拓扑结构数据进行采样,得到子图的拓扑结构数据。具体地,具体地,采样用并行处理单元201可随机选择P个顶点A;并从P个顶点A出发,从顶点A的邻居顶点中选择与顶点A之间的边的权重大于设定的第一权重阈值的Q个邻居顶点B;之后,从Q个邻居顶点B出发,从Q个邻居顶点B的邻居顶点中与顶点B之间的边的权重大于设定的第二权重阈值的R个邻居顶点,以此往外拓展,直至遍历到图拓扑结构数据中无邻居的顶点为止,或者遍历至无满足权重要求的顶点为止,或者遍历至选择出的子图的数据量达到设定数据量为止。其中,上述P、Q和R均为大于或等于1的整数。

[0048] 采样用并行处理单元201在得到子图的拓扑结构数据之后,处理单元10可控制训练用并行处理单元202获取子图的拓扑结构数据。在一些实施例中,如图1a所示,不同并行处理单元20之间无直接通信连接,处理单元10可从采样用并行处理单元201中读取子图的拓扑结构数据,并存储至处理单元10的内存中;并控制训练用并行处理单元202从处理单元10中读取子图的拓扑结构数据。在该实施例中,处理单元10可根据子图的拓扑结构数据在采样用并行处理单元201的存储位置,从采样用并行处理单元201中读取子图的拓扑结构数据。处理单元10启动一线程控制训练用并行处理单元202从处理单元10的内存中读取子图的拓扑结构数据。训练用并行处理单元202可将读取到的子图的拓扑结构数据存储至内存中。

[0049] 可选地,处理单元10可向训练用并行处理单元202发送另一读数据指令2;该读数据指令2可包括待读取数据的存储位置及数据长度。其中,待读取数据的存储位置可包括:处理单元的标识及待读取数据在处理单元的内存的内存空间位置。在该实施例中,待读取数据为子图的结构数据。待读取数据在处理单元的内存的内存空间位置为子图的拓扑结构数据在处理单元的内存空间位置;待读取数据的数据长度为子图的拓扑结构数据的数据长度。训练用并行处理单元202,可响应于读数据指令2,从处理单元10中从待读取数据在处理单元的内存的内存空间位置开始,读取上述数据长度的数据,作为子图的拓扑结构数据。

[0050] 在另一些实施例中,如图1b所示,不同并行处理单元10之间通信连接,则处理单元10可控制训练用并行处理单元202从采样用并行处理单元201中读取子图的拓扑结构数据。

[0051] 可选地,处理单元10可向训练用并行处理单元202发送读数据指令2;该读数据指令可包括待读取数据的存储位置及数据长度。其中,待读取数据的存储位置可包括:采样用并行处理单元201的标识及待读取数据在采样用并行处理单元201的内存的内存空间位置。在该实施例中,待读取数据为子图的结构数据。待读取数据在采样用并行处理单元201的内存的内存空间位置为:子图的拓扑结构数据在采样用并行处理单元201的内存空间位置;待读取数据的数据长度为子图的拓扑结构数据的数据长度。训练用并行处理单元202,可响应于读数据指令3,从采样用并行处理单元201中从待读取数据在采样用并行处理单元201的内存空间位置开始,读取上述数据长度的数据,作为子图的拓扑结构数据。

[0052] 值得说明的是,上述控制训练用并行处理单元201读取子图的拓扑结构数据的处理单元,与上述控制采样用并行处理单元201从处理单元10的内存中读取图拓扑结构数据的处理单元以及存储图数据集的处理单元可为同一处理单元,也可为不同的处理单元。

[0053] 对于图神经网络训练来说,不仅需要拓扑结构数据,还需要图的顶点特征数据。因此,处理单元10还根据子图的拓扑结构数据中的顶点标识,从其内存存储的顶点特征数据中获取子图的顶点特征数据。在一些实施例中,图拓扑结构数据中的顶点标识可为顶点编

号。顶点特征数据包括：顶点编号与顶点特征数据之间的对应关系，基于该对应关系，处理单元10可将子图的拓扑结构数据中的顶点编号在顶点编号与顶点特征数据之间的对应关系中进行匹配，得到与子图的拓扑结构数据中的顶点编号对应的顶点特征数据，作为子图的顶点特征数据。可选地，顶点编号与顶点特征数据之间的对应关系可以键值(Key-Value, KV)表形成存储。

[0054] 在另一些实施例中，图拓扑结构数据中的顶点标识为顶点索引，可提供指向该顶点的特征数据的指针。基于此，处理单元10可获取子图的拓扑结构数据中的顶点索引指向的顶点特征数据，作为子图的顶点特征数据。

[0055] 但是，由于子图的拓扑结构数据是采样用并行处理单元201采样得到的，得到的子图的拓扑结构数据中的顶点索引可能不连续，导致处理单元10获取到的子图的顶点特征数据不连续，即不同顶点索引指向的子图的顶点特征数据之间存在空白，这不仅造成对子图的顶点特征数据存储空间的浪费，还加大了处理单元10与训练用并行处理单元202之间传输的数据量，而且还会增加对训练用并行处理单元202内存的占用量。

[0056] 为了解决上述问题，在本实施例中，针对图拓扑结构数据中的顶点标识为顶点索引的实施例，采样用并行处理单元201在得到子图的拓扑结构数据之后，如图1c所示，还可按照从0递增的方式对子图的拓扑结构数据中的原始顶点索引进行重新编号，得到子图的新顶点索引；并可建立子图的原始顶点索引与新顶点索引之间的对应关系。可选地，采样用并行处理单元201可按照从0递增，增加步进为1的方式对子图的原始顶点索引进行重新编号。

[0057] 具体地，采样用并行处理单元201可按照从0递增，增加步进为1的方式首先对采样的M个起始顶点A进行随机编号 $0, 1, 2, \dots, (M-1)$ ；再按照从M递增，增加步进为1的方式对起始点A的邻居顶点B进行随机编号；依次往外拓展至子图的顶点重新编号完成。

[0058] 进一步，如图1c所示，处理单元10可从采样用并行处理单元201中读取子图的原始顶点索引与新顶点索引之间的对应关系。处理单元10可将该对应关系存储至内存中，并根据原始顶点索引与新顶点索引之间的对应关系，从处理单元的内存存储的顶点特征数据中获取子图的顶点特征数据。具体地，处理单元10根据原始顶点索引与新顶点索引之间的对应关系，从处理单元的内存存储的顶点特征数据中获取原始顶点索引指向的顶点特征数据，作为原始顶点索引对应的新顶点索引指向的顶点特征数据。

[0059] 进一步，处理单元10可按照原始顶点索引与新顶点索引之间的对应关系中新顶点索引的顺序，将子图的顶点特征数据存储至连续的内存区域。这样，新顶点索引可指向其对应的原始顶点索引指向的顶点特征数据。新顶点索引为从0开始的连续指针，因此，连续的内存区域中的子图的顶点特征数据的存储空间也为连续的，不再存在间断的空白区域，可降低子图的顶点特征数据的内存消耗；而且可减少处理单元10与训练用并行处理单元202之间传输的数据量，还会降低对训练用并行处理单元202内存的占用量。

[0060] 进一步，如图1a-图1c所示，处理单元10控制训练用并行处理单元202从处理单元10的内存中读取子图的顶点特征数据。相应地，训练用并行处理单元202可将子图的顶点特征数据存储至内存中，并以子图的拓扑结构数据和顶点特征数据为样本，进行图神经网络训练。

[0061] 进一步，处理单元10可从训练用并行处理单元202中读取图神经网络模型的参数。

进一步,处理单元10可判断当前图神经网络模型是否满足设定的条件;若判断结果为是,图神经网络训练完成;若判断结果为否,控制采样用并行处理单元201再次执行上述对图拓扑结构数据采样处理、子图的顶点特征数据提取以及训练用并行处理单元202对采样得到的子图的拓扑结构数据和顶点特征数据对图神经网络进行训练的过程,直至图神经网络满足设定条件,如图神经网络参数收敛等。

[0062] 值得说明的是,上述执行子图的顶点特征数据提取的处理单元与上述存储图数据集的处理单元可为同一处理单元;该处理单元与上述控制训练用并行处理单元202读取子图的顶点特征数据、控制训练用并行处理单元202读取子图的拓扑结构数据的处理单元、以及上述控制采样用并行处理单元201从处理单元10的内存中读取图拓扑结构数据的处理单元同一处理单元,也可为不同的处理单元。

[0063] 可选地,上述执行控制采样用并行处理单元201和采样用并行处理单元202进行数据读取的处理单元为同一处理单元;与执行子图的顶点特征数据提取的处理单元与上述存储图数据集的处理单元为不同的处理单元。

[0064] 针对执行上述数据控制操作的处理单元与上述执行子图的顶点特征数据提取的处理单元为不同的处理单元的实施例,计算系统可包括:至少2个处理单元10。其中,至少2个处理单元10中的第一处理单元可执行上述数据控制操作,第二处理单元可执行上述将获取的图数据集存储到第二存储单元的内存中的操作等。例如,第一处理单元可控制第一并行处理单元201从第二处理单元的内存中读取所述图数据集中的图拓扑结构数据;并控制第二并行处理单元202获取子图的拓扑结构数据。第二处理单元基于子图的拓扑结构数据中的顶点标识,从所述图数据集中的顶点特征数据中获取所述子图的顶点特征数据。进一步,第一处理单元可控制第二并行处理单元从第二处理单元的内存中读取子图的顶点特征数据等。

[0065] 在本实施例中,将数据量较小的图拓扑结构数据存储到采样用并行处理单元,而将数据量较大的图顶点特征数据存储到处理单元的内存中。采样用并行处理单元可对图拓扑结构数据进行采样,得到子图的拓扑结构数据;处理单元可根据子图的拓扑结构数据中的顶点标识,从内存存储的图顶点特征数据中获取子图的顶点特征数据;并将子图的拓扑结构数据和子图的顶点特征数据存储至训练用并行处理单元;训练用并行处理单元可以子图的拓扑结构数据和子图的顶点特征数据为样本进行图神经网络训练。在本申请实施例中,利用处理单元的内存大的性能对数据量较大的图顶点数据进行存储,利用并行处理单元的高速计算性能对图拓扑结构数据进行采样,提高了图采样效率。对于训练用并行处理单元来说,图采样效率提高,可降低训练用并行处理单元等待采样结果的时间,有助于提高训练用并行处理单元利用率。

[0066] 另一方面,可将图拓扑结构数据常驻在采样用并行处理单元201中,使得剩余的并行处理单元的内存可以存放图拓扑结构数据以及满足采样计算所需要的内存,图拓扑结构数据可以常驻并行处理单元中进行采样计算,避免了图拓扑结构数据的换入换出。

[0067] 另一方面,在本实施例中,数据量较大的顶点特征数据存放在处理单元的内存以及训练过程在训练用并行处理单元中进行,使得采样用并行处理单元具有可以存放图拓扑结构数据以及满足采样计算所需要的内存,进而拓扑结构数据可以常驻并行处理单元中进行采样计算,避免了图拓扑结构数据的换入换出,可降低图拓扑结构数据复制的时间开销。

[0068] 在本申请实施例中,为了提高并行处理单元20及处理单元10的利用率,在本申请实施例中,可采用流水线式任务执行方式,即采样用并行处理单元201、处理单元10及训练用并行处理单元202的计算任务同时进行,分别处理不同批次的图计算任务,可降低处理单元10和并行处理单元20处于空闲等待状态的时间,有助于提高资源利用率。例如,如图2所示,采样用并行处理单元201、处理单元10及训练用并行处理单元202在完成当前批次的计算任务之后,无需等待其它处理器对当前批次的计算任务处理完成,即可开启下一批次的图计算任务。这样,如图2所示,采样用并行处理单元201可在处理完第N批次的图采样任务之后,直接执行第(N+1)次的图采样任务,而无需等待处理单元10的第N次顶点特征提取及训练用并行处理单元202第N次图神经网络训练完成,可降低采样用并行处理单元201的空闲等待时间,有助于提高采样用并行处理单元201的利用率。图2中仅以批次1-5等5个批次进行图示,但不构成限定。

[0069] 另一方面,采样用并行处理单元201、处理单元10及训练用并行处理单元202的计算任务同时进行,分别处理不同批次的图计算任务,使得图采样、顶点特征提取和数据复制以及图神经网络训练同时进行,可降低采样计算和数据复制的时间,有助于提高图采样训练的效率。

[0070] 为了实现流水线式任务执行方式,上述模型训练过程的先后顺序,可由处理单元10控制。在一些实施例中,在另一些实施例中,处理单元10可启动多个线程,分别以不同的线程控制采样用并行处理单元201、处理单元10及训练用并行处理单元202循环执行下述步骤S11-S15,直至训练用并行处理单元202训练出的图神经网络满足设定条件,如图神经网络参数收敛等。其中,步骤S11-S15如下:

[0071] S11、处理单元10利用第一线程控制采样用并行处理单元201对图拓扑结构数据进行采样,得到子图的拓扑结构数据。

[0072] S12、处理单元10可利用第一线程和第二线程基于子图的拓扑结构数据中的顶点标识,从其内存存储的图数据集中的顶点特征数据中获取子图的顶点特征数据。

[0073] 可选地,处理单元10还利用第一线程控制采样用并行处理单元201按照从0递增的方式对子图的拓扑结构数据中的原始顶点索引进行重新编号,得到子图的新顶点索引;并利用第一线程控制采样用并行处理单元201建立子图的原始顶点索引与新顶点索引之间的对应关系。

[0074] 进一步,处理单元10可利用第二线程将子图的原始顶点索引与新顶点索引之间的对应关系复制到内存中。具体地,处理单元10利用第二线程从采样用并行处理单元201中读取子图的原始顶点索引与新顶点索引之间的对应关系。处理单元10可将该对应关系存储至内存中。

[0075] 进一步,处理单元10可利用第二线程根据子图的原始顶点索引与新顶点索引之间的对应关系,从处理单元10的内存存储的顶点特征数据中获取子图的顶点特征数据;并利用第二线程按照子图的原始顶点索引与新顶点索引之间的对应关系中新索引顶点的顺序,将子图的顶点特征数据存储至连续的内存区域。

[0076] S13、处理单元10利用第二线程将子图的拓扑结构数据和顶点特征数据复制到训练用并行处理单元202的内存中。

[0077] 可选地,处理单元10可利用第二线程控制训练用并行处理单元202获取子图的拓

扑结构数据;训练用并行处理单元202将子图的拓扑结构数据存储至训练用并行处理单元202的内存中;并利用第二线程控制训练用并行处理单元202从处理单元10内存中的连续的内存区域读取子图的顶点特征数据;训练用并行处理单元202将子图的顶点特征数据存储至训练用并行处理单元202的内存中。

[0078] 其中,处理单元10可利用第二线程控制训练用并行处理单元202获取子图的拓扑结构数据的具体实施方式可参见上述相关内容,在此不再赘述。

[0079] S14、处理单元10利用第三线程控制训练用并行处理单元202以子图的拓扑结构数据和顶点特征数据为样本,进行图神经网络训练。

[0080] S15、处理单元10可判断图神经网络的模型参数是否收敛。若判断结果为是,则图神经网络训练完成;若判断结果为否,则返回执行上述步骤S1;直至图神经网络模型参数收敛。

[0081] 上述实施例中,处理单元10启动不同的线程对控制采样用并行处理单元201、处理单元10及训练用并行处理单元202,实现了采样用并行处理单元201、处理单元10及训练用并行处理单元202之间流水线式的任务执行方式,即采样用并行处理单元201、处理单元10及训练用并行处理单元202的计算任务同时进行,分别处理不同批次的图计算任务,可降低处理单元10和并行处理单元20处于空闲等待状态的时间,有助于提高资源利用率。例如,第一线程可控制采样用并行处理单元201在第N次图采样完成之后,进入下一批次(第N+1批次)图采样处理,无需等待处理单元10和训练用并行处理单元202对第N次图采样得到的子图处理完成,可降低采样用并行处理单元201的空闲等待时间,有助于提高采样用并行处理单元201的资源利用率。

[0082] 在另一些实施例中,处理单元10可启动单线程,异步控制采样用并行处理单元201、处理单元10及训练用并行处理单元202,实现采样用并行处理单元201、处理单元10及训练用并行处理单元202之间流水线式的任务执行方式。具体实施为:处理单元10将图数据集中的图拓扑结构数据存储至采样用并行处理单元201中,并将图数据集中的顶点特征数据存储至处理单元10的内存中。之后处理单元10异步控制采样用并行处理单元201、处理单元10及训练用并行处理单元202循环执行以下操作,直至训练用并行处理单元202训练得到的图神经网络的模型参数收敛:

[0083] S21、处理单元10控制采样用并行处理单元201对图拓扑结构数据进行采样,得到子图的拓扑结构数据。

[0084] S22、处理单元10基于子图的拓扑结构数据中的顶点标识,从其内存存储的顶点特征数据中获取子图的顶点特征数据。

[0085] S23、处理单元10将子图的拓扑结构数据和顶点特征数据复制在训练用并行处理单元202的内存中。

[0086] S24、处理单元10控制训练用并行处理单元202以子图的拓扑结构数据和顶点特征数据为样本,进行图神经网络训练。

[0087] 关于步骤S21-S24的具体实施方式,可参见上述实施例的相关描述,在此不再赘述。

[0088] 值得说明的是,上述计算机设备的结构形式仅为示例性说明。在一些实施例中,计算机设备还可以包括:存储器、通信组件、电源组件、音频组件等组件(图中未示出)。其中,

存储器用于存储计算机程序,并可被配置为存储其它各种数据以支持在其所在设备上的操作。其中,处理器可执行存储器中存储的计算机程序,以实现相应控制逻辑。存储器可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器(SRAM),电可擦除可编程只读存储器(EEPROM),可擦除可编程只读存储器(EPROM),可编程只读存储器(PROM),只读存储器(ROM),磁存储器,快闪存储器,磁盘或光盘。

[0089] 通信组件被配置为便于其所在设备和其他设备之间有线或无线方式的通信。通信组件所在设备可以接入基于通信标准的无线网络,如WiFi,2G或3G,4G,5G或它们的组合。在一个示例性实施例中,通信组件经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,所述通信组件还可基于近场通信(NFC)技术、射频识别(RFID)技术、红外数据协会(IrDA)技术、超宽带(UWB)技术、蓝牙(BT)技术或其他技术来实现。

[0090] 显示组件可以包括液晶显示器(LCD)和触摸面板(TP)。如果显示组件包括触摸面板,显示组件可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。

[0091] 电源组件被配置为其所在设备的各种组件提供电力。电源组件可以包括电源管理系统,一个或多个电源,及其他与为电源组件所在设备生成、管理和分配电力相关联的组件。

[0092] 音频组件可被配置为输出和/或输入音频信号。例如,音频组件包括一个麦克风(MIC),当音频组件所在设备处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器或经由通信组件发送。在一些实施例中,音频组件还包括一个扬声器,用于输出音频信号。例如,对于具有语言交互功能的设备,可通过音频组件实现与用户的语音交互等。

[0093] 除了上述计算机设备之外,本申请实施例还提供模型训练方法,下面分别从处理单元和采样用并行处理单元的角度,对本申请实施例提供的模型训练方法进行示例性说明。

[0094] 图3为本申请实施例提供的模型训练方法的流程示意图。如图3所示,该模型训练方法包括:

[0095] 301、将获取的图数据集存储至处理单元的内存中。

[0096] 302、控制采样用并行处理单元从处理单元的内存中读取图数据集中的图拓扑结构数据。

[0097] 303、控制采样用并行处理单元对图拓扑结构数据采样,得到子图的拓扑结构数据。

[0098] 304、基于子图的拓扑结构数据中的顶点标识,从处理单元的内存存储的顶点特征数据中获取子图的顶点特征数据。

[0099] 305、控制训练用并行处理单元获取子图的拓扑结构数据和顶点特征数据。

[0100] 306、控制训练用并行处理单元以子图的拓扑结构数据和顶点特征数据为样本进行图神经网络训练。

[0101] 在本实施例中,步骤301中,处理单元可将获取的图数据集存储至处理单元的内存

中。考虑到图拓扑结构数据的数据量较小,顶点特征数据的数据量较大,因此,在步骤302中,可控制采样用并行处理单元从处理单元的内存中读取图数据集中的图拓扑结构数据。可选地,对于处理单元来说,在采样用并行处理单元从处理单元的内存中读取图数据集中的图拓扑结构数据之后,可继续保存图拓扑结构数据,也可释放图拓扑结构数据对应的处理单元的内存区域。在本实施例中,处理单元继续保存处理单元的内存中的图数据集中的顶点特征数据。

[0102] 对于处理单元在采样用并行处理单元从处理单元10的内存中读取图数据集中的图拓扑结构数据之后,释放图拓扑结构数据对应的处理单元的内存的实施方式,可实现图拓扑结构数据与图的顶点特征数据分离存储的方法,可解决大规模图数据的数据量大而无法全部放进并行处理单元的问题,可直接利用并行处理单元来加速采样计算的过程。

[0103] 进一步,为了实现图采样,在步骤303中,可控制采样用并行处理单元对图拓扑结构数据进行采样,得到子图的拓扑结构数据。可选地,可启动一线程,利用该线程控制采样用并行处理单元对图拓扑结构数据进行采样,得到子图的拓扑结构数据。在本实施例中,关于采样用并行处理单元对图拓扑结构数据进行采样的具体实施方式可参见上述设备实施例中的实施方式1和2,在此不再赘述。

[0104] 对于图神经网络训练来说,不仅需要拓扑结构数据,还需要图的顶点特征数据。因此,在步骤303中,可基于子图的拓扑结构数据中的顶点标识,从处理单元的内存存储的顶点特征数据中获取子图的顶点特征数据。在一些实施例中,图拓扑结构数据中的顶点标识可为顶点编号。顶点特征数据包括:顶点编号与顶点特征数据之间的对应关系,基于该对应关系,处理单元10可将子图的拓扑结构数据中的顶点编号在顶点编号与顶点特征数据之间的对应关系中进行匹配,得到与子图的拓扑结构数据中的顶点编号对应的顶点特征数据,作为子图的顶点特征数据。可选地,顶点编号与顶点特征数据之间的对应关系可以键值(Key-Value,KV)表形成存储。

[0105] 在另一些实施例中,图拓扑结构数据中的顶点标识为顶点索引,可提供指向该顶点的特征数据的指针。基于此,处理单元可获取子图的拓扑结构数据中的顶点索引指向的顶点特征数据,作为子图的顶点特征数据。

[0106] 但是,由于子图的拓扑结构数据是采样用并行处理单元采样得到的,得到的子图的拓扑结构数据中的顶点索引可能不连续,导致处理单元获取到的子图的顶点特征数据不连续,即不同顶点索引指向的子图的顶点特征数据之间存在空白,这不仅造成对子图的顶点特征数据存储空间的浪费,还加大了处理单元与训练用并行处理单元之间传输的数据量,而且还会增加对训练用并行处理单元的内存的占用量。

[0107] 为了解决上述问题,在本实施例中,针对图拓扑结构数据中的顶点标识为顶点索引的实施例,采样用并行处理单元在读取子图的拓扑结构数据之后,还可控制采样用并行处理单元按照从0递增的方式对子图的拓扑结构数据中的原始顶点索引进行重新编号,得到子图的新顶点索引;并可建立子图的原始顶点索引与新顶点索引之间的对应关系。可选地,采样用并行处理单元可按照从0递增,增加步进为1的方式对子图的原始顶点索引进行重新编号。

[0108] 进一步,处理单元可从采样用并行处理单元中读取子图的原始顶点索引与新顶点索引之间的对应关系。处理单元可将该对应关系存储至内存中,并根据原始顶点索引与新

顶点索引之间的对应关系,从处理单元的内存存储的顶点特征数据中获取子图的顶点特征数据。

[0109] 进一步,处理单元可按照原始顶点索引与新顶点索引之间的对应关系中新顶点索引的顺序,将子图的顶点特征数据存储至连续的内存区域。这样,新顶点索引可指向其对应的原始顶点索引指向的顶点特征数据。新顶点索引为从0开始的连续指针,因此,连续的内存区域中的子图的顶点特征数据的存储空间也为连续的,不再存在间断的空白区域,可降低子图的顶点特征数据的内存消耗;而且可减少处理单元与训练用并行处理单元之间传输的数据量,还会降低对训练用并行处理单元的内存的占用量。

[0110] 进一步,在步骤304中,可控制训练用并行处理单元获取子图的拓扑结构数据和顶点特征数据存储至。可选地,处理单元可控制训练用并行处理单元获取子图的拓扑结构数据。例如,处理单元可控制训练用并行处理单元从采样用并行处理单元中读取子图的拓扑结构数据;或者,处理单元可从采样用并行处理单元中读取子图的拓扑结构数据,并存储至处理单元的内存中;进一步,处理单元可控制训练用并行处理单元从处理单元的内存中读取子图的拓扑结构数据。进一步,处理单元可控制训练用并行处理单元将子图的拓扑结构数据存储至训练用并行处理单元的内存中。相应地,处理单元还可控制训练用并行处理单元从处理单元上述连续的内存区域中读取的子图的顶点特征数据,并控制训练用并行处理单元将子图的顶点特征数据存储至训练用并行处理单元的内存中。

[0111] 进一步,在步骤305中,可控制训练用并行处理单元以子图的拓扑结构数据和顶点特征数据为样本,进行图神经网络训练。

[0112] 进一步,处理单元可从训练用并行处理单元中读取图神经网络模型的参数。处理单元可判断当前图神经网络模型是否满足设定的条件;若判断结果为是,图神经网络训练完成;若判断结果为否,控制采样用并行处理单元再次执行上述对图拓扑结构数据采样处理、子图的顶点特征数据提取以及训练用并行处理单元对采样得到的子图的拓扑结构数据和顶点特征数据对图神经网络进行训练的过程,直至图神经网络满足设定条件,如图神经网络参数收敛等。

[0113] 在本实施例中,将数据量较小的图拓扑结构数据存储到采样用并行处理单元,而将数据量较大的图顶点特征数据存储到处理单元的内存中。采样用并行处理单元可对图拓扑结构数据进行采样,得到子图的拓扑结构数据;处理单元可根据子图的拓扑结构数据中的顶点标识,从内存存储的图顶点特征数据中获取子图的顶点特征数据;并将子图的拓扑结构数据和子图的顶点特征数据存储至训练用并行处理单元;训练用并行处理单元可以子图的拓扑结构数据和子图的顶点特征数据为样本进行图神经网络训练。在本申请实施例中,利用处理单元的内存大的性能对数据量较大的图顶点数据进行存储,利用并行处理单元的高速计算性能对图拓扑结构数据进行采样,提高了图采样效率。对于训练用并行处理单元来说,图采样效率提高,可降低训练用并行处理单元等待采样结果的时间,有助于提高训练用并行处理单元利用率。

[0114] 在本申请实施例中,为了提高并行处理单元及处理单元的利用率,在本申请实施例中,可采用流水线式任务执行方式,即采样用并行处理单元、处理单元及训练用并行处理单元的计算任务同时进行,分别处理不同批次的图计算任务,可降低处理单元和并行处理单元处于空闲等待状态的时间,有助于提高资源利用率。

[0115] 为了实现流水线式任务执行方式,上述模型训练过程的先后顺序,可由处理单元控制。在一些实施例中,在另一些实施例中,处理单元可启动多个线程,并利用多个线程执行上述步骤S11-S15,直至训练用并行处理单元训练出的图神经网络满足设定条件,如图神经网络参数收敛等。其中,步骤S11-S15可参见上述实施例的相关内容,在此不再赘述。

[0116] 上述实施例中,处理单元启动不同的线程对控制采样用并行处理单元、处理单元及训练用并行处理单元,实现了采样用并行处理单元、处理单元及训练用并行处理单元之间流水线式的任务执行方式,即采样用并行处理单元、处理单元及训练用并行处理单元202的计算任务同时进行,分别处理不同批次的图计算任务,可降低处理单元和并行处理单元处于空闲等待状态的时间,有助于提高资源利用率。

[0117] 在另一些实施例中,处理单元可启动单线程,异步控制采样用并行处理单元、处理单元及训练用并行处理单元,实现采样用并行处理单元、处理单元及训练用并行处理单元之间流水线式的任务执行方式。具体实施为:处理单元将图数据集中的图拓扑结构数据存储至采样用并行处理单元中,并将图数据集中的顶点特征数据存储至处理单元的内存中。之后处理单元异步控制采样用并行处理单元、处理单元及训练用并行处理单元循环执行上述步骤S21-S24中的步骤,直至训练用并行处理单元训练得到的图神经网络的模型参数收敛。

[0118] 值得说明的是,上述步骤301-306也可由同一处理单元执行,也可由不同的处理单元执行。例如,第一处理单元可执行上述步骤301和步骤304图数据集的数据存储及子图的顶点特征提取操作;第二处理单元可执行上述步骤302、303以及步骤305-306中的数据控制操作等。

[0119] 相应地,本申请实施例还提供一种模型训练方法,适用于上述采样用并行处理单元。如图4所示,模型训练方法包括:

[0120] 401、在处理单元的控制下从处理单元的内存中读取的图拓扑结构数据。

[0121] 402、对图拓扑结构数据进行采样,以得到子图的拓扑结构数据,以供训练用并行处理单元获取。

[0122] 403、基于子图的拓扑结构数据,确定子图关联的顶点标识信息,以供处理单元获取并基于顶点标识信息获取子图的顶点特征数据以及控制训练用并行处理单元读取所述子图的顶点特征数据。

[0123] 其中,训练用并行处理单元可以子图的拓扑结构数据和顶点特征数据进行图神经网络训练。

[0124] 在本实施例中,关于步骤401和402的描述,可参见上述设备实施例中,采样用并行处理单元的相关内容,在此不再赘述。

[0125] 在步骤403中,可基于子图的拓扑结构数据,确定子图关联的顶点标识信息。在本实施例中,子图关联的顶点标识信息可为子图的顶点标识或者与顶点标识相关的信息。

[0126] 在一些实施例中,顶点标识为顶点编号。相应地,可从子图的拓扑结构数据中获取子图的顶点编号,作为子图关联的顶点标识信息。

[0127] 在另一些实施例中,顶点标识为顶点索引,可提供指向该顶点的特征数据的指针。基于此,处理单元可获取子图的拓扑结构数据中的顶点索引指向的顶点特征数据,作为子图的顶点特征数据。

[0128] 但是,由于子图的拓扑结构数据是采样用并行处理单元采样得到的,得到的子图的拓扑结构数据中的顶点索引可能不连续,导致处理单元获取到的子图的顶点特征数据不连续,即不同顶点索引指向的子图的顶点特征数据之间存在空白,这不仅造成对子图的顶点特征数据存储空间的浪费,还加大了处理单元与训练用并行处理单元之间传输的数据量,而且还会增加对训练用并行处理单元的内存的占用量。

[0129] 为了解决上述问题,在本实施例中,针对图拓扑结构数据中的顶点标识为顶点索引的实施例,步骤403的一种可选实施方式为:按照从0递增的方式对子图的拓扑结构数据中的原始顶点索引进行重新编号,得到子图的新顶点索引;并可建立子图的原始顶点索引与新顶点索引之间的对应关系。可选地,采样用并行处理单元可按照从0递增,增加步进为1的方式对子图的原始顶点索引进行重新编号。进一步,可将子图的原始顶点索引与新顶点索引之间的对应关系,作为子图关联的顶点标识信息。

[0130] 进一步,处理单元可从采样用并行处理单元中读取子图的顶点标识信息。处理单元可基于子图的顶点标识信息,从内存存储的图顶点特征数据中获取子图的顶点特征数据;并控制训练用并行处理单元从处理单元的内存中读取子图的顶点特征数据。关于采样用并行处理单元将顶点标识信息提供给处理单元,以及处理单元基于子图关联的顶点标识信息获取子图的顶点特征数据的具体实施方式,可参见上述实施例的相关内容,在此不再赘述。

[0131] 进一步,训练用并行处理单元可以子图的拓扑结构数据和处理单元提供的子图顶点特征数据进行图神经网络训练。

[0132] 在本实施例中,将数据量较小的图拓扑结构数据存储到采样用并行处理单元,而将数据量较大的图顶点特征数据存储到处理单元的内存中。采样用并行处理单元可对图拓扑结构数据进行采样,得到子图的拓扑结构数据,并获取子图的顶点标识信息;处理单元可基于子图的顶点标识信息,从内存存储的图顶点特征数据中获取子图的顶点特征数据;并将子图的拓扑结构数据和子图的顶点特征数据存储至训练用并行处理单元;训练用并行处理单元可以子图的拓扑结构数据和子图的顶点特征数据为样本进行图神经网络训练。在本申请实施例中,利用处理单元的内存大的性能对数据量较大的图顶点数据进行存储,利用并行处理单元的高速计算性能对图拓扑结构数据进行采样,提高了图采样效率。对于训练用并行处理单元来说,图采样效率提高,可降低训练用并行处理单元等待采样结果的时间,有助于提高训练用并行处理单元利用率。

[0133] 需要说明的是,上述实施例所提供方法的各步骤的执行主体均可以是同一设备,或者,该方法也由不同设备作为执行主体。比如,步骤301和302的执行主体可以为设备A;又比如,步骤301的执行主体可以为设备A,步骤302的执行主体可以为设备B;等等。

[0134] 另外,在上述实施例及附图中的描述的一些流程中,包含了按照特定顺序出现的多个操作,但是应该清楚了解,这些操作可以不按照其在本文中出现的顺序来执行或并行执行,操作的序号如301、302等,仅仅是用于区分开各个不同的操作,序号本身不代表任何的执行顺序。另外,这些流程可以包括更多或更少的操作,并且这些操作可以按顺序执行或并行执行。

[0135] 相应地,本申请实施例还提供一种存储有计算机指令的计算机可读存储介质,当计算机指令被一个或多个处理器执行时,致使一个或多个处理器执行上述各模型训练方法

中的步骤。

[0136] 需要说明的是,本文中的“第一”、“第二”等描述,是用于区分不同的消息、设备、模块等,不代表先后顺序,也不限定“第一”和“第二”是不同的类型。

[0137] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0138] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0139] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0140] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0141] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0142] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0143] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0144] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要

素的过程、方法、商品或者设备中还存在另外的相同要素。

[0145] 以上所述仅为本申请的实施例而已,并不用于限制本申请。对于本领域技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本申请的权利要求范围之内。

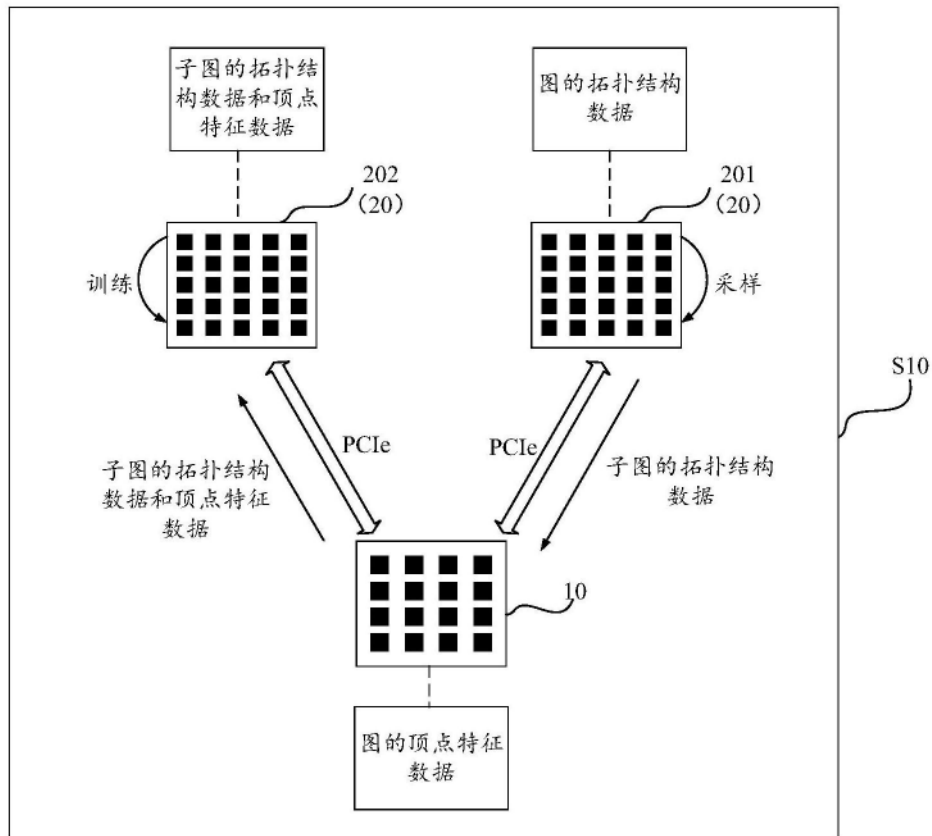


图1a

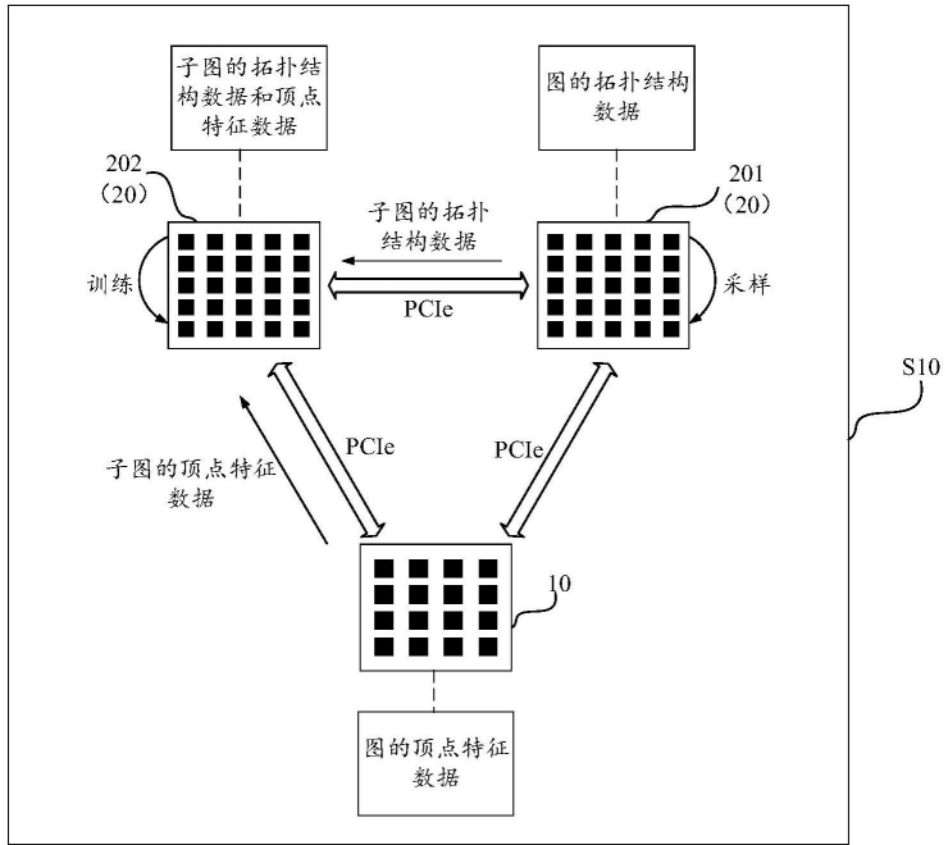


图1b

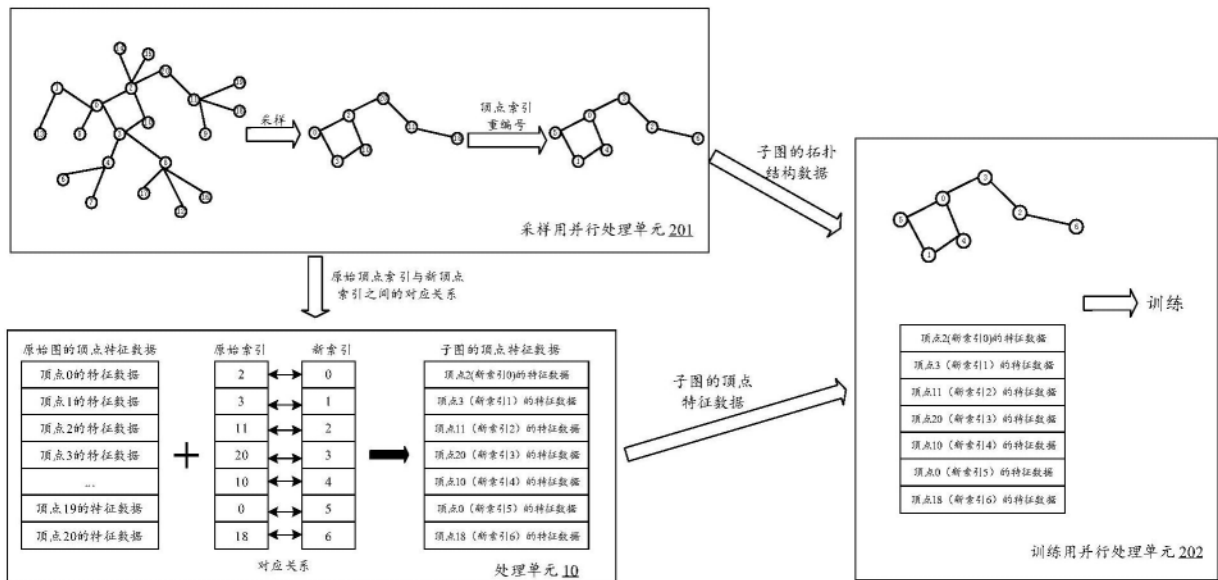


图1c

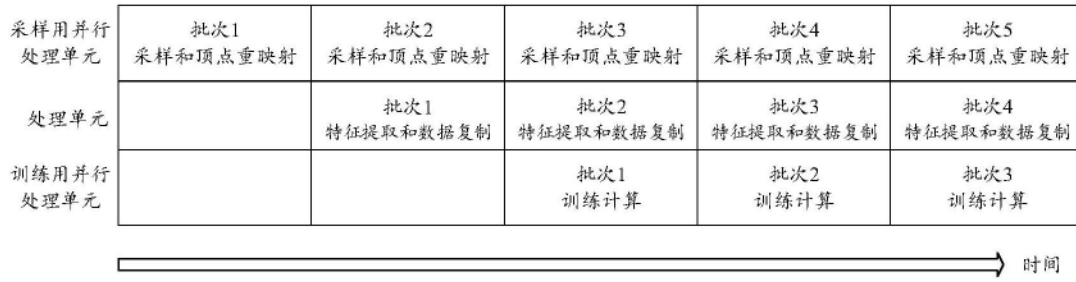


图2

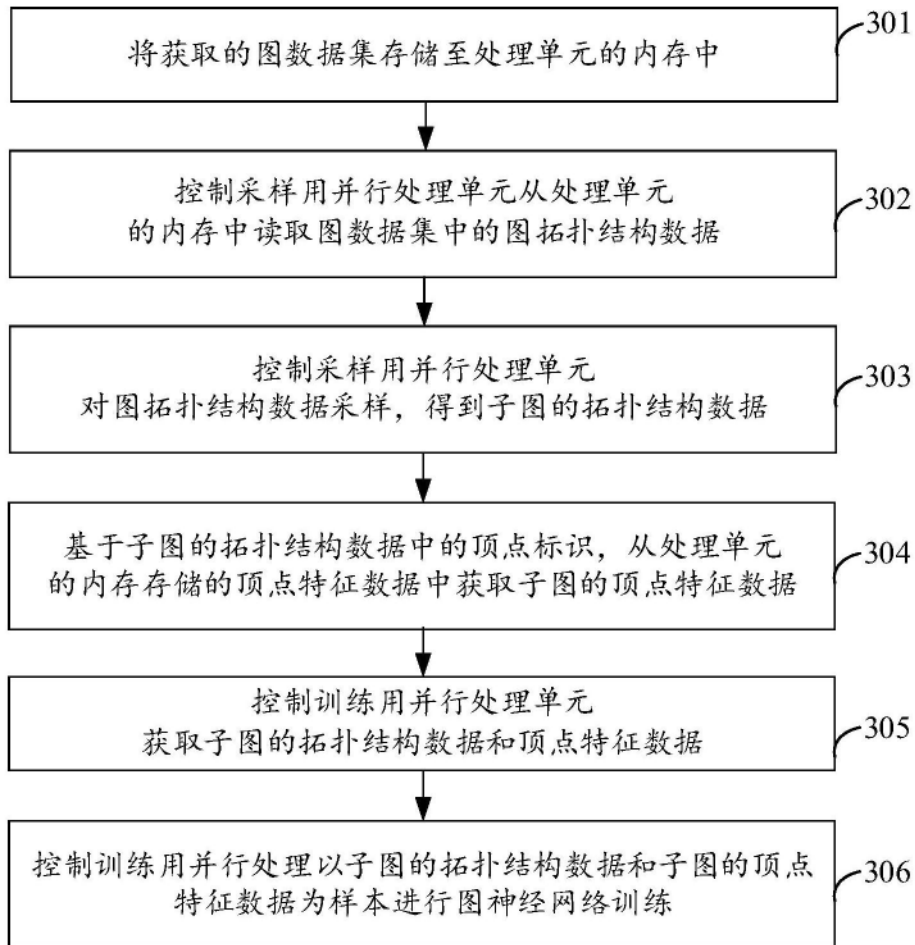


图3

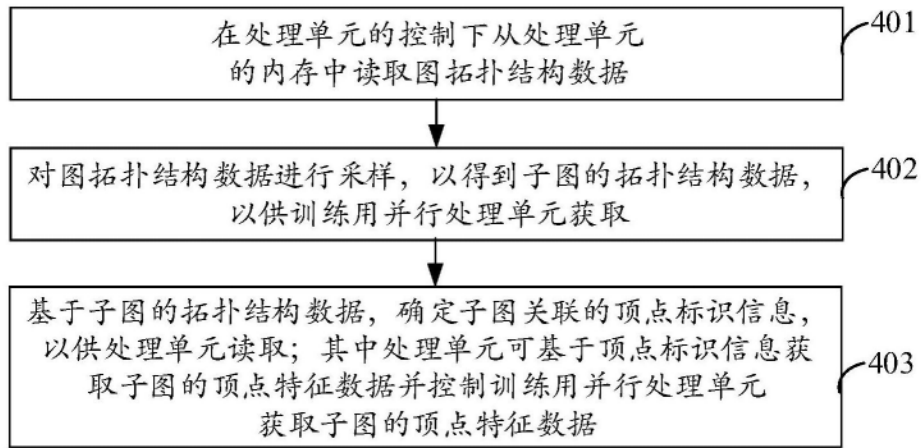


图4