

(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
2024 年 10 月 31 日 (31.10.2024)



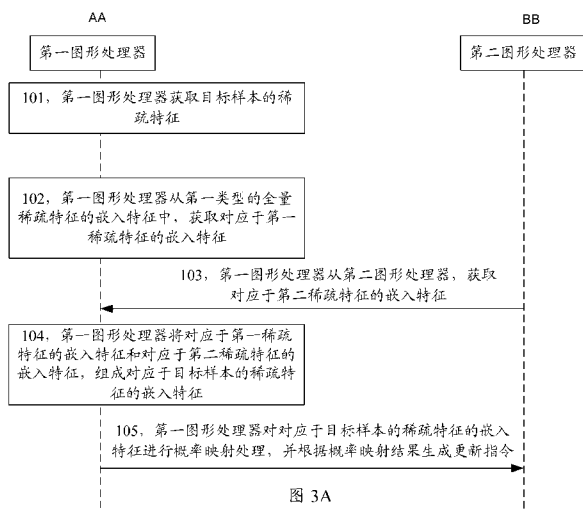
(10) 国际公布号
WO 2024/221925 A1

- (51) 国际专利分类号:
G06F 9/54 (2006.01) **G06N 3/098** (2023.01)
G06F 9/50 (2006.01) **G06N 3/0985** (2023.01)
- (21) 国际申请号: PCT/CN2023/135891
- (22) 国际申请日: 2023 年 12 月 1 日 (01.12.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202310492698.2 2023年4月28日 (28.04.2023) CN
- (71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。
- (72) 发明人: 弓静 (GONG, Jing); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

- (74) 代理人: 北京派特恩知识产权代理有限公司 (CHINA PAT INTELLECTUAL PROPERTY OFFICE); 中国北京市海淀区苏州街 3 号大恒科技大厦南座五层 503, Beijing 100080 (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,

(54) Title: DATA PROCESSING METHOD AND APPARATUS, AND ELECTRONIC DEVICE, COMPUTER-READABLE STORAGE MEDIUM AND COMPUTER PROGRAM PRODUCT

(54) 发明名称: 数据处理方法、装置、电子设备、计算机可读存储介质及计算机程序产品



101 The first graphics processing unit acquires sparse features of a target sample
102 The first graphics processing unit acquires, from among embedding features of full-scale sparse features of a first type, embedding features corresponding to first sparse features
103 The first graphics processing unit acquires, from the second graphics processing unit, embedding features corresponding to second sparse features
104 The first graphics processing unit combines the embedding features corresponding to the first sparse features and the embedding features corresponding to the second sparse features, so as to form embedding features corresponding to the sparse features of the target sample
105 The first graphics processing unit performs probability mapping processing on the embedding features corresponding to the sparse features of the target sample, and generates an update instruction according to a probability mapping result
AA First graphics processing unit
BB Second graphics processing unit

(57) Abstract: Provided in the present application are an artificial-intelligence-based data processing method and apparatus, and an electronic device, a computer-readable storage medium and a computer program product. The data processing method is applied to a first graphics processing unit. The first graphics processing unit stores embedding features of full-scale sparse features of a first type, and a second graphics processing unit stores embedding features of full-scale sparse features of a second type, wherein the first type is different from the second type. The method comprises: a first graphics processing unit acquiring sparse features of a target sample, wherein the sparse features of the target sample include first sparse features of a first type and second sparse features of a second type; the first graphics processing unit acquiring, from among embedding features of full-scale sparse features of the first type, embedding features corresponding to the first sparse features; the first graphics processing unit acquiring, from a second graphics processing unit, embedding features that correspond to the second sparse features and are obtained by querying embedding features of full-scale sparse features of the second type; the first graphics processing unit combining the embedding features corresponding to

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,
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))。

the first sparse features and the embedding features corresponding to the second sparse features, so as to form embedding features corresponding to the sparse features of the target sample; and the first graphics processing unit executing probability mapping processing on the embedding features corresponding to the sparse features of the target sample, and generating an update instruction according to a probability mapping result, wherein the update instruction is used for instructing the second graphics processing unit to update the embedding features of the full-scale sparse features of the second type.

(57) 摘要: 本申请提供了一种基于人工智能的数据处理方法、装置、电子设备、计算机可读存储介质及计算机程序产品; 数据处理方法应用于第一图形处理器, 第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征, 第二图形处理器存储有第二类型的全量稀疏特征的嵌入特征, 第一类型与第二类型不同; 方法包括: 第一图形处理器获取目标样本的稀疏特征, 目标样本的稀疏特征包含第一类型的第一稀疏特征和第二类型的第二稀疏特征; 第一图形处理器从第一类型的全量稀疏特征的嵌入特征中, 获取对应于第一稀疏特征的嵌入特征; 第一图形处理器从第二图形处理器获取从第二类型的全量稀疏特征的嵌入特征查询得到的对应于第二稀疏特征的嵌入特征; 第一图形处理器将对应于第一稀疏特征的嵌入特征和对应于第二稀疏特征的嵌入特征, 组成对应于目标样本的稀疏特征的嵌入特征; 第一图形处理器对对应于目标样本的稀疏特征的嵌入特征执行概率映射处理, 并根据概率映射结果生成更新指令, 更新指令用于指示第二图形处理器更新第二类型的全量稀疏特征的嵌入特征。

数据处理方法、装置、电子设备、计算机可读存储介质及计算机程序产品

相关申请的交叉引用

本申请基于申请号为 202310492698.2、申请日为 2023 年 04 月 28 日的中国
5 专利申请提出，并要求中国专利申请的优先权，中国专利申请的全部内容在此引入本申请作为参考。

技术领域

本申请涉及人工智能技术，尤其涉及一种数据处理方法、装置、电子设备、计算机可读存储介质及计算机程序产品。

10 背景技术

人工智能 (Artificial Intelligence, AI) 是计算机科学的一个综合技术，通过研究各种智能机器的设计原理与实现方法，使机器具有感知、推理与决策的功能。基于医疗图像以及医疗文本的分析是人工智能领域的重要应用之一，医疗分析系统是指利用计算机对医疗图像和医疗文本进行处理、分析和理解，以
15 识别出各种不同模式的目标和对象的系统。

相关技术中基于 Parameter Server 方式对嵌入特征进行学习，通常稀疏特征的嵌入特征表示(embedding)会存放在中央处理器(CPU, Central Processing Unit)上，因此 embedding 训练也会在 CPU 内完成。但是由于 embedding 运算的规模较大，从而导致训练速度较慢。

20 发明内容

本申请实施例提供一种基于人工智能的数据处理方法、装置、电子设备、计算机可读存储介质及计算机程序产品，能够在实现大规模嵌入特征存储的前提下通过图形处理器提高更新处理速度。

本申请实施例的技术方案是这样实现的：

本申请实施例提供一种基于人工智能的数据处理方法，所述数据处理方法应用于第一图形处理器，所述第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征，第二图形处理器存储有第二类型的全量稀疏特征的嵌入特征，所述
5 述第一类型与所述第二类型不同，所述方法包括：

所述第一图形处理器获取目标样本的稀疏特征，所述目标样本的稀疏特征包含所述第一类型的第一稀疏特征和所述第二类型的第二稀疏特征；

所述第一图形处理器从所述第一类型的全量稀疏特征的嵌入特征中，获取对应于所述第一稀疏特征的嵌入特征；

10 所述第一图形处理器从所述第二图形处理器获取从所述第二类型的全量稀疏特征的嵌入特征查询得到的对应于所述第二稀疏特征的嵌入特征；

所述第一图形处理器将对应于所述第一稀疏特征的嵌入特征和对应于所述第二稀疏特征的嵌入特征，组成对应于所述目标样本的稀疏特征的嵌入特征；

所述第一图形处理器对所述对应于所述目标样本的稀疏特征的嵌入特征执行概率映射处理，并根据概率映射结果生成更新指令，所述更新指令用于指示
15 所述第二图形处理器更新所述第二类型的全量稀疏特征的嵌入特征。

本申请实施例提供一种基于人工智能的数据处理装置，所述数据处理方法应用于第一图形处理器，第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征，所述第二图形处理器存储有第二类型的全量稀疏特征的嵌入特征，所述
20 述第一类型与所述第二类型不同，所述装置包括：

第一接收模块，配置为所述第一图形处理器获取目标样本的稀疏特征，所述目标样本的稀疏特征包含所述第一类型的第一稀疏特征和所述第二类型的第二稀疏特征；

第一获取模块，配置为所述第一图形处理器从所述第二图形处理器获取从
25 所述第二类型的全量稀疏特征的嵌入特征查询得到的对应于所述第二稀疏特征的嵌入特征；

第一返回模块，配置为所述第一图形处理器从所述第二图形处理器获取从

所述第二类型的全量稀疏特征的嵌入特征查询得到的对应于所述第二稀疏特征的嵌入特征;

第一确定模块,配置为所述第一图形处理器将对应于所述第一稀疏特征的嵌入特征和对应于所述第二稀疏特征的嵌入特征,组成对应于所述目标样本的
5 稀疏特征的嵌入特征;

第一更新模块,配置为所述第一图形处理器对所述对应于所述目标样本的稀疏特征的嵌入特征执行概率映射处理,并根据概率映射结果生成更新指令,所述更新指令用于指示所述第二图形处理器更新所述第二类型的全量稀疏特征的嵌入特征。

10 本申请实施例提供一种数据处理方法,所述数据处理方法应用于第二图形处理器,第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征,所述第二图形处理器存储有第二类型的全量稀疏特征的嵌入特征,所述第一类型与所述第二类型不同;

所述方法包括:

15 所述第二图形处理器接收所述第一图形处理器发送的,所述第二类型的第二稀疏特征;

所述第二图形处理器从所述第二类型的全量稀疏特征的嵌入特征中查询得到对应于所述第二稀疏特征的嵌入特征;

20 所述第二图形处理器将对应于所述第二稀疏特征的嵌入特征传输至所述第一图形处理器,以使所述第一图形处理器对对应于所述第一类型的第一稀疏特征的嵌入特征和对应于所述第二稀疏特征的嵌入特征,执行概率映射处理,其中,对应于所述第一稀疏特征的嵌入特征是所述第一图形处理器从所述第一类型的全量稀疏特征的嵌入特征中获取得到的;

25 所述第二图形处理器从第一图形处理器接收根据概率映射结果生成的更新指令,并基于所述更新指令更新所述第二类型的全量稀疏特征的嵌入特征。

本申请实施例提供一种数据处理装置,所述数据处理方法应用于第二图形处理器,第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征,所述第

二图形处理器存储有第二类型的全量稀疏特征的嵌入特征，所述第一类型与所述第二类型不同；

所述装置包括：

第二接收模块，配置为所述第二图形处理器接收所述第一图形处理器发送的，所述第二类型的第二稀疏特征；

查询模块，配置为所述第二图形处理器从所述第二类型的全量稀疏特征的嵌入特征中查询得到对应于所述第二稀疏特征的嵌入特征；

第二返回模块，配置为所述第二图形处理器将对应于所述第二稀疏特征的嵌入特征传输至所述第一图形处理器，以使所述第一图形处理对对应于所述第一类型的第一稀疏特征的嵌入特征和对应于所述第二稀疏特征的嵌入特征，执行概率映射处理，其中，对应于所述第一稀疏特征的嵌入特征是所述第一图形处理器从所述第一类型的全量稀疏特征的嵌入特征中获取得到的；

第二更新模块，配置为所述第二图形处理器从第一图形处理器接收根据概率映射结果生成的更新指令，并基于所述更新指令更新所述第二类型的全量稀疏特征的嵌入特征。

本申请实施例提供一种图形处理器，所述图形处理器用于执行本申请实施例提供的数据处理方法。

本申请实施例提供一种电子设备，所述电子设备包括：

存储器，用于存储计算机可执行指令；

图形处理器，用于执行所述存储器中存储的计算机可执行指令时，实现本申请实施例提供的数据处理方法。

本申请实施例提供一种计算机可读存储介质，存储有计算机可执行指令，所述计算机可执行指令被处理器执行时实现本申请实施例提供的数据处理方法，所述处理器是中央处理器或者图形处理器。

本申请实施例提供一种计算机程序产品，包括计算机可执行指令，所述计算机可执行指令被处理器执行时实现本申请实施例提供的数据处理方法，所述处理区是中央处理器或者图形处理器。

本申请实施例具有以下有益效果：

按照类型将嵌入特征存储在不同的图形处理器中，可以实现嵌入特征的分布式存储，第一图形处理器获取目标样本的不同类型的稀疏特征，第一图形处理器从本地获取第一类型的嵌入特征，并从第二图形处理器，获取第二类型的嵌入特征，按照类型划分方便从对应图形处理器中获取嵌入特征，可以提高获取效率，并且便于特征管理，第一图形处理器对两个类型的嵌入特征进行概率映射处理，并根据概率映射结果向第二图形处理器传输更新指令，更新指令用于指示第二图形处理器更新第二类型的嵌入特征，从而在分布式存储的前提下实现了嵌入特征的更新，充分利用图形处理器计算的高效性，在满足存储需求的同时极大提升嵌入特征更新速度。

附图说明

图 1A 是本申请实施例提供的基于人工智能的数据处理系统的结构示意图；

图 1B 是本申请实施例提供的数据处理系统的架构图；

图 2 是本申请实施例提供的电子设备的结构示意图；

图 3A-图 3D 是本申请实施例提供的基于人工智能的数据处理方法的流程图；

图 4 是本申请实施例提供的推荐系统的框架示意图；

图 5 是本申请实施例提供的 Parameter Server 的物理架构示意图；

图 6 是本申请实施例提供的单工作节点的示意图；

图 7 是本申请实施例提供的基于人工智能的数据处理方法的数据流转示意图；

图 8 是本申请实施例提供的基于人工智能的数据处理方法的数据流转示意图。

具体实施方式

为了使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申

请作进一步地详细描述，所描述的实施例不应视为对本申请的限制，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它实施例，都属于本申请保护的范围。

在以下的描述中，涉及到“一些实施例”，其描述了所有可能实施例的子集，
5 但是可以理解，“一些实施例”可以是所有可能实施例的相同子集或不同子集，并且可以在不冲突的情况下相互结合。

在以下的描述中，所涉及的术语“第一\第二\第三”仅仅是是区别类似的对象，不代表针对对象的特定排序，可以理解地，“第一\第二\第三”在允许的情况下可以互换特定的顺序或先后次序，以使这里描述的本申请实施例能够以除
10 了在这里图示或描述的以外的顺序实施。

除非另有定义，本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中所使用的术语只是为了描述本申请实施例的目的，不是旨在限制本申请。

对本申请实施例进行进一步详细说明之前，对本申请实施例中涉及的名词
15 和术语进行说明，本申请实施例中涉及的名词和术语适用于如下的解释。

1) 嵌入特征 (Embedding): Embedding 就是用一个数值向量表示一个对象的方法，这里说的对象可以是一个词、一个物品，也可以是一部电影等等。推荐场景中的类别、ID 型特征非常多，大量使用独热编码会导致样本特征向量极度稀疏，而深度学习的结构特点又不利于稀疏特征向量的处理，因此几乎所有
20 深度学习推荐模型都会由嵌入层负责将稀疏高维特征向量转换成稠密低维特征向量，也即嵌入特征 Embedding。

2) TensorFlow: 是一个基于数据流编程的符号数学系统，被广泛应用于各类机器学习算法的编程实现，其拥有多层级结构，可部署于各类服务器、PC 终端和网页，并支持 GPU 和 TPU 高性能数值计算，被广泛应用于产品开发和各
25 领域的科学研究。

3) 参数服务器 (Parameter Server): 是个编程框架，用于方便分布式并行程序的编写，其中重点是对大规模参数的分布式存储和协同的支持，参数服务

器的集群中的节点可以分为计算节点和参数服务节点两种。其中，计算节点负责对分配到自己本地的训练数据（块）计算学习，并更新对应的参数；参数服务节点采用分布式存储的方式，各自存储全局参数的一部分，并作为服务方接受计算节点参数查询和更新请求。

5 4) 稀疏特征与稠密特征：稀疏特征是那些在数据集中不连续出现的特征，并且大多数值为零，例如账号标识，如果存在 100 个账号，那么对于每个账号标识而言，稀疏特征中会有 99 位数值为 0，只有 1 位数值为 1，之所以称为稀疏特征，是因为它们在数据集中只有很少的非零值，全量稀疏特征表征所有样本的稀疏特征，例如，对于账号类型而言，所有样本总计 100 个账号，则这 100
10 个账号分别对应的 100 个稀疏特征即为账号类型的全量稀疏特征。稠密特征一般是相对稀疏特征来说的，稠密特征中数值 0 的占比较少，甚至没有。

5) 类型：本申请实施例提供的类型可以是内容类型或者是数据格式类型，内容类型是基于稀疏特征所表征的语义进行区分的，数据格式类型是基于稀疏特征的数据格式进行区分的。

15 相关技术中基于 Parameter Server 方式对嵌入特征进行学习，通常稀疏特征的嵌入特征表示（embedding）会存放在中央处理器（CPU，Central Processing Unit）上，embedding 训练也在 CPU 内完成。但是由于 embedding 运算的规模较大，如果 embedding 训练在 CPU 内完成，会限制训练速度。申请人在实施本申请实施例时发现如果将 embedding 训练部署在图形处理器（GPU，graphic
20 s processing unit）中，可以有效提高训练速度，但是 GPU 的存储能力无法满足大规模 embedding 的存储要求。

本申请实施例提供一种基于人工智能的数据处理方法、装置、电子设备和计算机可读存储介质，能够在实现大规模嵌入特征存储的前提下通过图形处理器提高更新处理速度。

25 本申请实施例所提供的数据处理方法，可以由终端/服务器独自实现；也可以由终端和服务服务器协同实现，例如终端或者服务器独自承担下文所述的数据处理方法，或者，终端向服务器发送数据样本，服务器根据接收的数据样本执行

数据处理方法。

本申请实施例提供的用于数据处理的电子设备可以是各种类型的终端或服务
服务器，其中，服务器可以是独立的物理服务器，也可以是多个物理服务器构成
的服务器集群或者分布式系统，还可以是提供云计算服务的云服务器；终端可
5 以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表、
车载终端等，但并不局限于此。终端以及服务器可以通过有线或无线通信方式
进行直接或间接地连接，本申请在此不做限制。

以服务器为例，例如可以是部署在云端的服务器集群，向用户开放人工智
能云服务（AI as a Service, AiaaS），平台会把几类常见的 AI 服务进行拆分，
10 并在云端提供独立或者打包的服务，这种服务模式类似于一个 AI 主题商城，所
有的用户都可以通过应用程序编程接口的方式来接入使用 AiaaS 平台提供的一
种或者多种人工智能服务。

作为示例，其中的一种人工智能云服务可以为数据处理服务，即云端的服
务器封装有本申请实施例提供的数据处理的程序。使用者通过终端调用云服务
15 中的数据处理服务，以使部署在云端的服务器调用封装的数据处理的程序。

参见图 1A，图 1A 是本申请实施例提供的数据处理系统的应用场景示意图，
终端 400 通过网络 300 连接服务器 200，网络 300 可以是广域网或者局域网，
又或者是二者的组合，服务器 200 中部署有 4 个 GPU，下面以 GPU1 和 GPU2
进行说明。

20 数据样本可以是推荐模型的数据样本，例如针对新闻推荐模型，单个数据
样本可以包括登录新闻客户端的账号的相关对象数据，这里的相关对象数据包
括账号标识（账号 ID）、账号年龄、表征账号兴趣的属性标签等等。全量数据
样本中对应第一类型（例如账号标识类型）的稀疏特征的嵌入特征保存在服务
器 200 内部署的 GPU1 内，全量数据样本中对应第二类型（例如年龄段类型）
25 的稀疏特征的嵌入特征保存在服务器 200 内部署的 GPU2 内。

终端 400 将匹配请求发送至服务器 200，服务器 200 调用推荐模型，任意
一个 GPU（例如 GPU1）中部署有推荐模型，推荐模型的输入是目标对象的账

号标识类型的嵌入特征以及年龄段类型的嵌入特征，因此 GPU1 需要从 GPU2 中基于年龄段类型的稀疏特征查询目标对象的年龄段类型的嵌入特征，推荐模型的输出是目标对象点击某个新闻的推荐概率，当推荐概率大于概率阈值时，服务器 200 向终端 400 返回该新闻，新闻客户端向目标对象推送该新闻。为了
5 能够提高推荐准确度，下面介绍各个类型的嵌入特征更新的过程。

参见图 1B，图 1B 是本申请实施例提供的数据处理系统的架构图。服务器 200 接收目标样本 A，这里的目标样本 A 被输入至 GPU1，GPU1 获取目标样本 A 的稀疏特征，目标样本 A 的稀疏特征包含目标样本 A 在账号标识类型的稀疏特征和目标样本 A 在年龄段类型的稀疏特征，GPU1 从账号标识类型的全量稀疏特征的嵌入特征中，获取目标样本 A 在账号标识类型的稀疏特征的嵌入特征；
10 GPU1 从 GPU2，获取目标样本 A 在年龄段类型的稀疏特征的嵌入特征，目标样本 A 在年龄段类型的稀疏特征的嵌入特征是 GPU2 从年龄段类型的全量稀疏特征的嵌入特征查询得到的；GPU1 根据目标样本 A 在账号标识类型的稀疏特征的嵌入特征和目标样本 A 在年龄段类型的稀疏特征的嵌入特征，确定目标样本 A 的稀疏特征的嵌入特征；GPU1 根据目标样本 A 的稀疏特征的嵌入特征正
15 向传播，并根据概率映射结果向 GPU2 传输更新指令，更新指令用于指示 GPU 2 更新年龄段类型的全量稀疏特征的嵌入特征。

在一些实施例中，服务器 200 可以是独立的物理服务器，也可以是多个物理服务器构成的服务器集群或者分布式系统，还可以是提供云服务、云数据库、
20 云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN、以及大数据和人工智能平台等基础云计算服务的云服务器。终端 400 可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表等，但并不局限于此。终端以及服务器可以通过有线或无线通信方式进行直接或间接地连接，本发明实施例中不做限制。

25 在一些实施例中，终端或服务器可以通过运行计算机程序来实现本申请实施例提供的基于人工智能的数据处理方法。举例来说，计算机程序可以是操作系统中的原生程序或软件模块；可以是本地（Native）应用程序（APP，Applic

ation), 即需要在操作系统中安装才能运行的程序, 如直播 APP 或者即时通信 APP; 也可以是小程序, 即只需要下载到浏览器环境中就可以运行的程序; 还可以是能够嵌入至任意 APP 中的小程序。总而言之, 上述计算机程序可以是任意形式的应用程序、模块或插件。

5 下面说明本申请实施例提供的用于数据处理的电子设备的结构, 参见图 2, 图 2 是本申请实施例提供的用于数据处理的电子设备的结构示意图, 以电子设备是服务器 200 为例说明, 图 2 所示的用于数据处理的服务器 200 包括: 至少一个处理器 210(处理器 210 可以是图形处理器或者中央处理器)、存储器 250、至少一个网络接口 220。服务器 200 中的各个组件通过总线系统 240 耦合在一
10 起。可理解, 总线系统 240 用于实现这些组件之间的连接通信。总线系统 240 除包括数据总线之外, 还包括电源总线、控制总线和状态信号总线。但是为了清楚说明起见, 在图 2 中将各种总线都标为总线系统 240。

处理器 210 可以是一种集成电路芯片, 具有信号的处理能力, 例如通用处理器、数字信号处理器 (DSP, Digital Signal Processor), 或者其他可编程逻辑
15 器件、分立门或者晶体管逻辑器件、分立硬件组件等, 其中, 通用处理器可以是微处理器或者任何常规的处理器等。

存储器 250 包括易失性存储器或非易失性存储器, 也可包括易失性和非易失性存储器两者。其中, 非易失性存储器可以是只读存储器 (ROM, Read Only Memory), 易失性存储器可以是随机存取存储器 (RAM, Random Access Memory)。本申请实施例描述的存储器 250 旨在包括任意适合类型的存储器。存
20 储器 250 可选地包括在物理位置上远离处理器 210 的一个或多个存储设备。

在一些实施例中, 存储器 250 能够存储数据以支持各种操作, 这些数据的示例包括程序、模块和数据结构或者其子集或超集, 下面示例性说明。

操作系统 251, 包括用于处理各种基本系统服务和执行硬件相关任务的系统程序, 例如框架层、核心库层、驱动层等, 用于实现各种基础业务以及处理
25 基于硬件的任务。

网络通信模块 252, 用于经由一个或多个 (有线或无线) 网络接口 220 到

达其他电子设备, 示例性的网络接口 220 包括: 蓝牙、无线相容性认证(WiFi)、和通用串行总线(USB, Universal Serial Bus)等。

在一些实施例中, 本申请实施例提供的数据处理装置可以采用软件方式实现, 例如, 可以是上文所述的终端中的数据处理插件, 可以是上文所述的服务
5 器中数据处理服务。当然, 不局限于此, 本申请实施例提供的数据处理装置可以提供为各种软件实施例, 包括应用程序、软件、软件模块、脚本或代码在内的各种形式。图 2 示出了存储在存储器 250 中的数据处理装置 255-1, 其可以是程序和插件等形式的软件, 例如图像处理插件, 并包括一系列的模块, 包括第一接收模块 2551、第一获取模块 2552、第一返回模块 2553、第一确定模块 25
10 54 以及第一更新模块 2555。图 2 还示出了存储在存储器 250 中的数据处理装置 255-2, 其可以是程序和插件等形式的软件, 例如图像处理插件, 并包括一系列的模块, 包括第二接收模块 2556、查询模块 2557、第二返回模块 2558、第二更新模块 2559。

如前, 本申请实施例提供的数据处理方法可以由各种类型的电子设备实施。
15 例如包括多个图形处理器的电子设备, 电子设备中部署有第一图形处理器和第二图形处理器(第一图形处理器和第二图形处理器还可以部署在不同的电子设备上), 例如包括多个中央处理器的电子设备, 电子设备中部署有第一中央处理器和第二中央处理器, 其中, 图形处理器的计算能力优于中央处理器。

在一些实施例中, 第一图形处理器存储有第一类型的全量稀疏特征的嵌入
20 特征, 第二图形处理器存储有第二类型的全量稀疏特征的嵌入特征, 第一类型与第二类型不同。通过本申请实施例可以将多个类型的全量稀疏特征的嵌入特征划分到不同的图形处理器中存储, 从而可以缓解每个图形处理器的存储压力。

作为示例, 若这里的类型是基于稀疏特征所表征的语义进行区分的内容类型, 可以获取稀疏特征所表征的语义的字段(字段比如可以是关键词), 当字段
25 与预设字段匹配时, 则将预设字段对应的内容类型作为稀疏特征的内容类型。例如, 稀疏特征所表征的语义的字段可以是“账号”, 预设字段可以是账号类型, 那么这里的字段与预设字段匹配, 稀疏特征的类型即为账号类型。按照内容类

型的区分进行分布式存储的方式可以有助于快速获取不同内容类型的稀疏特征的嵌入特征，从而在对特征多样性以及丰富度有要求的推荐场景下可以更快查询到相应内容类型的嵌入特征，提高分布式场景下训练过程中嵌入特征的获取效率。

5 作为示例，这里的类型还可以是基于稀疏特征的数据格式进行区分的数据格式类型，数据格式类型可以是文本类型、图像类型、语音类型等等，按照数据格式类型的区分进行分布式存储的方式可以有助于快速获取不同数据格式类型的稀疏特征的嵌入特征，从而可以更快查询到相应数据格式类型的嵌入特征，提高分布式场景下训练过程中嵌入特征的获取效率。

10 本申请实施例所提供的类型不局限于上述示例，在推荐场景下，稀疏特征是用于描述用户的数据，那么类型可以是基于不同用户进行划分得到的对象类型，例如有 10 个稀疏特征均是用于描述用户 A，那么这 10 个稀疏特征所属的类型即为对象类型 A，有 10 个稀疏特征均是用于描述用户 B，那么这 10 个稀疏特征所属的类型即为对象类型 B，即每个用户对应一种对象类型，将属于相
15 同对象类型的稀疏特征的嵌入特征存储在相同图形处理器内，按照这种分布式存储方式，在推荐场景下，且参与训练的目标样本来源于不同用户时，可以更快查询到相应用户的所有相关嵌入特征，提高分布式场景下训练过程中嵌入特征的获取效率。本申请实施例提供的类型还可以是基于稀疏特征的来源进行划分得到的来源类型等等，本申请实施例提供的类型还可以是基于稀疏特征的获取
20 时段进行划分得到的时段类型，在此不再进行赘述。

下文主要以内容类型为例进行说明，第一图形处理器和第二图形存储器存储有多个内容类型的所有稀疏特征，合计存储有 100 个稀疏特征，例如这里的内容类型可以是账号类型、年龄段类型等等，这里的第一类型可以是账号类型，这里的第二类型可以是年龄段类型，那么第一图形处理器中存储有身份类型的全量稀疏特征的嵌入特征表征第一图形处理器存储有所有样本的属于身份类型的
25 稀疏特征，那么第二图形处理器中存储有年龄段类型的稀疏特征的全量嵌入特征，表征第二图形处理器存储有所有样本的属于年龄段类型的稀疏特征。

在一些实施例中，第一图形处理器和第二图形处理器属于多个图形处理器，多个图形处理器存储有多个类型的全量稀疏特征的嵌入特征，第一类型包含多个类型中的至少一个类型，第二类型包含多个类型中的至少一个类型。通过本申请实施例可以按照类型来将多个类型的全量稀疏特征的嵌入特征划分到不同的图形处理器中存储，从而在查询的时候可以具有明确的指向性，而不必从所有图形处理器中进行查询。

承接上述示例，图形处理器系统包括多个图形处理器，例如包括 4 个图形处理器，其中，第一图形处理器和第二图形处理器是 4 个图形处理器中的 2 个图形处理器，100 个稀疏特征对应的 100 个嵌入特征会分布式存储在 4 个图形处理器中，这里的第一类型可以是账号类型和兴趣类型，这里的第二类型可以是年龄段类型和性别类型，那么第一图形处理器中存储有身份类型的所有稀疏特征的嵌入特征以及兴趣类型的所有稀疏特征的嵌入特征，那么第二图形处理器中存储有年龄段类型的所有稀疏特征的嵌入特征以及性别类型的所有稀疏特征的嵌入特征。

在一些实施例中，第一类型的全量稀疏特征的嵌入特征的存储位置和第二类型的全量稀疏特征的嵌入特征的存储位置，是在多个类型的全量稀疏特征的嵌入特征的数据量大于阈值的情况下划分的。通过本申请实施例仅在数据量大于阈值的情况下进行分布式存储，从而可以提高图形处理器的存储资源利用率，避免单个图形处理器的存储资源的负载过大。在多个类型的全量稀疏特征的嵌入特征的数据量不大于阈值的情况下，多个类型的全量稀疏特征的嵌入特征保存在多个图形处理器中的每个图形处理器中。通过本申请实施例可以在数据量不大于阈值的情况利用单个图形处理器来存储所有稀疏特征的嵌入特征，从而可以后续训练阶段不需要各个图形处理器不需要再从其他图形处理器获取嵌入特征，可以提高训练效率。

作为示例，这里的数据量可以是嵌入特征的总数目，在多个类型的全量稀疏特征的嵌入特征的总数目大于阈值的情况下，第一类型的全量稀疏特征的嵌入特征以及第二类型的全量稀疏特征的嵌入特征会以分布式的方式分别保存在

两个图形处理器中，在多个类型的全量稀疏特征的嵌入特征的总数目不大于阈值的情况下，多个类型的全量稀疏特征的嵌入特征，即所有的嵌入特征会被保存在每个图形处理器中，即每个图形处理器中均保存有全量的嵌入特征。

作为示例，在实际业务场景中，通常稀疏特征的总量较多，需要的处理较为复杂。本申请实施例针对稀疏特征使用 tfra 组件的动态 embedding 数据结构存储。当稀疏特征参数量较大，会导致对应的嵌入特征的数量较大（例如，大于阈值），GPU 单卡显存无法容纳时，可以通过按照类型划分的方式将稀疏特征对应的 embedding 划分到多张 GPU 卡的显存中存放；当稀疏特征的总数量较少，则意味着对应的嵌入特征的总数量较小（例如不大于阈值），GPU 单卡显存可以容纳时，使用 Replica 的方式，每张 GPU 卡的显存都放置一份全量的嵌入特征。

作为示例，在推荐系统领域，嵌入特征（embedding）已成为处理稀疏特征的常用手段，作为一种“函数映射”，embedding 层通常将高维稀疏特征映射为低维稠密向量，这里所得到的低维稠密向量即为嵌入特征，这里的嵌入特征与对应的稀疏特征所表示的信息含义相同，但是所占据的存储空间更小且数据类型更低，再进行模型端到端训练。本申请实施例利用组件 tfra.dynamic_embedding 作为 embedding 的存储方式，该组件用 tf.lookup.MutableHashTable 保存多个参数（fullweights），也即保存嵌入特征，并且可以复用 tensorflow 原生的优化器。

作为示例，针对每个稀疏特征的映射处理的过程如下，将稀疏特征的类型编号与某个设定数值进行相除处理，得到对应的余数，基于稀疏特征的类型编号得到的余数即为稀疏特征的处理器标识，从而将稀疏特征保存在对应处理器标识所指示的图形处理器中。

通过本申请实施例可以将稀疏特征的嵌入特征按照类型区分保存在相应的图形处理器，合理分配多个图形处理器的存储资源。并且，相比于每个图形处理器存储所有类型的稀疏特征的嵌入特征这一技术方案，在本申请中，对于每个图形处理器来说，只需要负责所分配的特定类型的稀疏特征的嵌入特征保存

与更新，降低存储和更新开销，降低成本。通过本申请实施例可以灵活存储嵌入特征，在每个图形处理器的存储资源利用率以及图形处理器之间的交互资源占用率之间取得平衡。

参见图 3A，图 3A 是本申请实施例提供的基于人工智能的数据处理方法的
5 流程示意图，将结合图 3A 示出的步骤 101 至步骤 105 进行说明。

在步骤 101 中，第一图形处理器获取目标样本的稀疏特征，目标样本的稀疏特征包含第一类型的第一稀疏特征和第二类型的第二稀疏特征。

作为示例，服务器内可以部署有多个图形处理器，这里步骤 101 可以是由
10 第一图形处理器实现，同时服务器可以是单个机器或者是由多个机器组成的服务器集群，即多个图形处理器可以部署在一个服务器内，也可以部署在多个服务器内。

作为示例，在进行训练时，目标样本被输入至第一图形处理器，第一图形处理器从目标样本中获取目标样本的稀疏特征，例如表征账号标识的稀疏特征以及表征年龄段的稀疏特征，即这里的目标样本的稀疏特征是包括第一稀疏特征
15 和第二稀疏特征，这里第一稀疏特征是上文所述的第一类型的全量稀疏特征中对应目标样本的稀疏特征，这里的第二稀疏特征是上文所述的第二类型的全量稀疏特征中对应目标样本的稀疏特征，例如，目标样本是用户 A，第一类型是账号类型，第二类型是年龄段类型，第一类型的全量稀疏特征是所有用户的账号类型的稀疏特征，如果用户的数目是 100，每个用户有一个账号，那么第一类型的全量稀疏特征是 100 个用户的账号类型的稀疏特征，第二类型的全量
20 稀疏特征是所有用户的年龄段类型的稀疏特征，总共有 50 个年龄段，那么第二类型的全量稀疏特征是 50 个年龄段类型的稀疏特征，这里第一稀疏特征是用用户 A 在账号类型的稀疏特征（表征一个账号的稀疏特征）以及用户 A 在年龄段类型的稀疏特征（表征一个年龄段的稀疏特征）。

25 由于嵌入特征是被预先保存在第一图形处理器以及第二图形处理器中，而嵌入特征又是对稀疏特征进行嵌入压缩处理得到的，因此下面具体说明在存储之前获取多个数据样本的稀疏特征的过程，即获取全量稀疏特征的过程。

在一些实施例中，获取登录推荐客户端的多个对象账号以及每个对象账号的对象数据，将多个对象账号的对象数据作为多个数据样本，对多个对象账号的对象数据进行特征解析处理，得到多个对象账号的对象特征，并将多个对象账号的对象特征作为多个数据样本的稀疏特征。

5 作为示例，推荐客户端可以是新闻客户端、视频客户端、购物客户端等等，以新闻客户端为例进行说明，对象账号是之前登录过新闻客户端的账号，账号由用户持有（下文统一以对象来代替用户）。例如，对象 A 持有对象账号 A，对象 B 持有对象账号 B，可以获取对象账号 A 的属性数据以及操作数据作为对象账号 A 的对象数据。针对对象账号 A 而言，可以将对象账号 A 的对象数据
10 作为一个数据样本，多个数据样本可以由所有对象的对象数据构成的。

通过本申请实施例可以获取准确描述对象的稀疏特征，从而使得嵌入特征可以有效表征对象以提高嵌入特征的对象表征能力。

在一些实施例中，上述对多个对象账号的对象数据进行特征解析处理，得到多个对象账号的对象特征，可以通过以下技术方案实现：针对每个对象账号
15 执行以下处理：从对象账号的对象数据中获取以下至少之一：对应对象账号的账号数据、对应对象账号的生物数据、对应对象账号的位置数据、对象账号的兴趣数据；对获取的数据进行独热编码处理，并将得到的编码结果作为对象账号的对象特征。

作为示例，针对对象 A 而言，可以获取对象账号 A 的属性数据以及操作数
20 据作为对象账号 A 的对象数据，从属性数据中获取账号标识（例如账号 ID）、生物标识（例如，性别）以及位置标识（例如长期居住地），从操作数据中获取兴趣标识，操作数据可以是浏览操作收藏操作等等，兴趣标识可以是被操作的信息的兴趣标签，例如被操作的信息是体育新闻，则兴趣标签可以是体育。通过独热编码处理对这些标识进行编码处理，例如，将性别男编码为（0，1），将
25 性别女编码为（1，0），通过独热编码得到的编码结果可以作为对象特征。在存储过程中，对于部分对象而言，即便对象不同，他们也具有相同的性别或者相同的年龄，因此对于部分类型的稀疏特征而言，稀疏特征的数目是固定的。

通过本申请实施例可以获取准确描述对象的稀疏特征，并且稀疏特征可以全面表征对象，从而使得嵌入特征可以有效表征对象以提高嵌入特征的对象表征能力。

5 在一些实施例中，执行步骤 101 之前，获取第一图形处理器前一次从多个数据样本中获取到的历史目标样本；将多个数据样本中除历史目标样本的数据样本作为其他数据样本；从其他数据样本中随机获取至少一个数据样本作为目标样本。

作为示例，为了不断提高嵌入特征的表达能力，由图形处理器构成的图形处理器系统会进行多次训练，每次训练过程可以均是针对多个数据样本进行的，10 也可以针对多个数据样本中的部分数据样本进行，对此不进行限制，但是为了提高训练效率，本申请实施例提供多个图形处理器，从而每个图形处理器可以分担部分数据样本，例如，本次训练过程中会涉及到 10 个数据样本，这 10 个数据样本可以是多个数据样本，也可以是多个数据样本中的部分数据样本，第一图形处理器 A 会接收 10 个数据样本中的 2 个数据样本。针对每次训练过程，15 获取第一图形处理器上一次训练过程中接收到的数据样本（历史目标样本），这个获取的过程可以是第一图形处理器本身来执行，例如，上一次训练过程中接收到的数据样本是数据样本 A 和数据样本 B，那么本次训练过程就会从多个数据样本中随机获取除数据样本 A 和数据样本 B 之外的任意数目的数据样本，可以获取两个数据样本，或者获取其他数目的数据样本。

20 通过本申请实施例可以避免多次训练过程中单个图形处理器重复处理相同的数据样本，提高单个图形处理器的训练效率。

在一些实施例中，执行步骤 101 之前，基于图形处理器的数目对多个数据样本进行平均划分处理，得到多个样本集合，其中，每个样本集合包括至少一个数据样本，样本集合的集合数目是与图形处理器的数目相同；确定多个样本25 集合与多个图形处理器之间的一一对应关系；将与第一图形处理器之间具有对应关系的样本集合中的数据样本，作为对应第一图形处理器的目标样本。通过本申请实施例可以通过平均划分的方式来保证每个图形处理器的负载均衡。通

过本申请实施例可以避免多次训练过程中单个图形处理器重复处理相同的数据样本，从而优化训练效果。

作为示例，除了按照上述思路来避免多次训练过程中单个图形处理器重复处理相同的数据样本，还可以通过平均划分的方式来保证每个图形处理器的负载均衡，例如多个数据样本包括 10 个数据样本，图形处理器系统中包括 5 个图形处理器，通过平均划分处理可以得到 5 个样本集合，每个样本集合中包括 2 个数据样本；样本集合和图形处理器之间的对应关系可以是预先保存好的，直接通过获取的方式来确定，除此之外还可以确定出样本集合和图形处理器之间的匹配度，按照任意随机的顺序依次为每个图形处理器确定出与每个样本集合的匹配度，例如针对第一图形处理器 A，确定出第一图形处理器 A 与每个样本集合的匹配度，再将最高匹配度对应的样本集合作为与该图形处理器具有对应关系的样本集合，匹配度与样本集合与第一图形处理器的历史目标样本数据的重复度负相关。

在步骤 102 中，第一图形处理器从第一类型的全量稀疏特征的嵌入特征中，获取对应于第一稀疏特征的嵌入特征。

作为示例，第一图形处理器从第一类型的全量稀疏特征的嵌入特征中，获取目标样本在第一类型的第一稀疏特征的嵌入特征。

作为示例，全量表征所有的含义，每个稀疏特征对应有一个嵌入特征，相当于所有样本（包括目标样本）在第一类型的稀疏特征的嵌入特征均存储在第一图形处理器，因此这里相当于是从全量的嵌入特征中进行查询，第一图形处理器从第一类型的所有的稀疏特征分别对应的嵌入特征中查询出目标样本在第一类型的第一稀疏特征的嵌入特征，相当于是从第一类型的所有的嵌入特征中查询出目标样本在第一类型的嵌入特征。通过本申请实施例可以直接从本地获取目标样本在第一类型的第一稀疏特征的嵌入特征，提高嵌入特征的获取效率。

例如，第一类型是账号标识类型，这里的全量表征所有的含义，第一图形处理器存储有属于账号标识类型的所有稀疏特征分别对应的嵌入特征，例如，存储有 100 个账号类型的稀疏特征分别对应的 100 个嵌入特征，第一图形处理

器从所有属于账号标识类型的嵌入特征中查询出目标样本的嵌入特征，例如对象 A 在账号标识类型的嵌入特征，这里查询的时候是根据目标样本的稀疏特征进行查询，例如，目标样本（对象 A）的稀疏特征是（0，0，0，0，0，0，0，0，1，0，0，0）以稀疏特征为键（key）从哈希表中查询到对应的嵌入特征，
5 例如，这里的嵌入特征可以是数值 9，这里的嵌入特征相当于是值（value），稀疏特征和嵌入特征以键值对的方式存储在哈希表中。

在步骤 103 中，第一图形处理器从第二图形处理器获取从第二类型的全量稀疏特征的嵌入特征查询得到的对应于第二稀疏特征的嵌入特征。

作为示例，第一图形处理器从第二图形处理器，获取目标样本在第二类型的第二稀疏特征的嵌入特征，目标样本在第二类型的第二稀疏特征的嵌入特征是第二图形处理器从第二类型的全量稀疏特征的嵌入特征查询得到的。
10

在一些实施例中，参见图 3B，步骤 103 中第一图形处理器从第二图形处理器获取从第二类型的全量稀疏特征的嵌入特征查询得到的对应于第二稀疏特征的嵌入特征可以通过图 3B 示出的步骤 1031 至步骤 1033 实现。

15 在步骤 1031 中，第一图形处理器向第二图形处理器传输第二稀疏特征。

在步骤 1032 中，第二图形处理器从第二类型的全量稀疏特征的嵌入特征中查询出对应于第二稀疏特征的嵌入特征。

全量表征所有的含义，每个稀疏特征对应有一个嵌入特征，相当于所有样本（包括目标样本）在第二类型的稀疏特征的嵌入特征均存储在第二图形处理器，因此这里相当于是从全量的嵌入特征中进行查询，第二图形处理器从第二类型的所有的稀疏特征分别对应的嵌入特征中查询出对应于目标样本的第二稀疏特征的嵌入特征，相当于是从第二类型的所有的嵌入特征中查询出目标样本在第二类型的嵌入特征。
20

在步骤 1033 中，第一图形处理器接收，对应于第二稀疏特征的嵌入特征。

25 作为示例，第一类型是账号标识类型，第二类型是年龄段类型，第一图形处理器将目标样本的年龄段类型的稀疏特征传输至第二图形处理器，第二图形处理器从所有属于年龄段类型的所有嵌入特征中查询出目标样本的年龄段类型

的嵌入特征，第二图形处理器将目标样本的年龄段类型的嵌入特征传输至第一图形处理器，第一图形处理器接收目标样本的年龄段类型的嵌入特征。通过本申请实施例可以实现嵌入特征的交换，对于第一图形处理器而言，可以从第二图形处理器获取第一图形处理器本地未存储的嵌入特征，可以在分布式存储的前提下实现嵌入特征的共享，提高嵌入特征的利用率。

在步骤 104 中，第一图形处理器将对应于第一稀疏特征的嵌入特征和对应于第二稀疏特征的嵌入特征，组成对应于目标样本的稀疏特征的嵌入特征。

作为示例，第一图形处理器将目标样本在第一类型的稀疏特征的嵌入特征以及目标样本在第二类型的稀疏特征的嵌入特征组成目标样本的稀疏特征的嵌入特征，例如，第一图形处理器将目标样本在账号标识类型的稀疏特征的嵌入特征以及目标样本在年龄段类型的稀疏特征的嵌入特征组成目标样本的稀疏特征的嵌入特征，作为后续模型推理的输入。

在步骤 105 中，第一图形处理器对对应于目标样本的稀疏特征的嵌入特征执行概率映射处理，并根据概率映射结果生成更新指令，更新指令用于指示第二图形处理器更新第二类型的全量稀疏特征的嵌入特征。

作为示例，概率映射是指将嵌入特征转化为概率的映射处理，概率映射处理可以是线性变换处理或者非线性变换处理，即概率映射处理实际上是对嵌入特征进行线性变换或者非线性变换，以将嵌入特征变换为一个概率值，概率映射的方式较多，本申请不具体限定。

作为示例，针对线性变换的概率映射处理，会涉及到权重参数（权重参数的数据形式是列向量）以及偏置参数（标量），将权重参数与嵌入特征（嵌入特征的数据形式是行向量）进行点乘处理，再将点乘结果（这里的点乘结果是标量）与偏置参数相加，最后将相加得到的数值作为映射得到的概率（概率映射结果），也可以对相加得到的数值进行归一化处理，将归一化得到的数值作为概率映射结果。

作为示例，概率映射的过程具体还可以参见公式（1）：

$$h(x) = \frac{1}{1 + \exp(-W^T x + b)} \quad (1);$$

其中, x 是嵌入特征, $-W^T$ 是权重参数, b 是偏置参数, 可以直接将 $-W^T x + b$ 作为概率映射得到的概率, 还可以根据公式 (1) 对 $-W^T x + b$ 执行归一化处理, 得到 $h(x)$, 并将 $h(x)$ 作为概率映射结果。

5 在一些实施例中, 参见图 3C, 步骤 105 中第一图形处理器对对应于目标样本的稀疏特征的嵌入特征执行概率映射处理, 可以通过执行图 3C 示出的步骤 1051 实现, 步骤 105 中根据概率映射结果生成更新指令可以通过执行图 3C 示出的步骤 1052 至步骤 1053 实现。

在步骤 1051 中, 以对应于目标样本的稀疏特征的嵌入特征作为推荐模型的
10 输入, 通过推荐模型对对应于目标样本的稀疏特征的嵌入特征执行概率映射处理, 得到对应目标样本的推荐概率作为概率映射结果。

作为示例, 推荐模型可以是深度神经网络 (DNN, Deep Neural Networks), 对于任何一个目标样本, 可以对应多个类型的嵌入特征, 将这些类型的嵌入特征组成对应于目标样本的嵌入特征, 以对应于目标样本的嵌入特征为推荐模
15 型的输入, 推荐模型至少包括全连接层, 通过全连接层可以对对应于目标样本的嵌入特征执行概率映射处理 (可以是线性变换或者是非线性变换), 得到每个目标样本的推荐概率。

在步骤 1052 中, 第一图形处理器确定对应目标样本的推荐概率与对应目标样本的标签值之间的误差, 并基于误差确定目标样本的第二类型的全量稀疏特
20 征的嵌入特征的梯度。

作为示例, 这里的误差是将推荐概率与标签值代入损失函数得到的理, 通过误差计算目标样本的第二类型的全量稀疏特征的嵌入特征的梯度, 计算方式包括但不限于以下几种: 数值计算法 (通过函数求导进行计算)、解析计算法。

在步骤 1053 中, 第一图形处理器生成携带有第二类型的全量稀疏特征的嵌
25 入特征的梯度的更新指令。

作为示例, 这里首先生成未携带任何信息的更新指令, 用于指示第二图形

处理器对第二类型的全量稀疏特征的嵌入特征进行更新，由于更新需要基于第二类型的全量稀疏特征的嵌入特征的梯度执行，因此将第二类型的全量稀疏特征的嵌入特征的梯度添加至上述未携带任何信息的更新指令。

5 作为示例，第一图形处理器将更新指令传输至第二图形处理器，第二图形处理器基于第二类型的全量稀疏特征的嵌入特征的梯度对第二类型的全量稀疏特征的嵌入特征进行更新。

通过本申请实施例可以使得第二图形处理器对本地存储的嵌入特征进行更新，相当于实现了分布式存储下的嵌入特征更新，突破了分布式存储造成的信息隔离的壁垒，将训练过程也拆分到不同图形处理器以分布式的方式完成，可以
10 以有效提高训练效率。

在一些实施例中，第一图形处理器根据概率映射结果更新第一类型的全量稀疏特征的嵌入特征。具体而言，概率映射结果是推荐概率，第一图形处理器确定出推荐概率与标签值之间的误差，并基于误差确定出第一类型的全量稀疏特征的嵌入特征的梯度，再根据第一类型的全量稀疏特征的嵌入特征的梯度更新
15 第一类型的全量稀疏特征的嵌入特征。通过本申请实施例可以使得第一图形处理器对本地存储的嵌入特征进行更新，由于这里的概率映射结果是由第一图形处理器得到的，因此第一图形处理器可以直接基于概率映射结果更新本地存储的嵌入特征，提高训练效率。

作为示例，TensorFlow 提供的是声明式的编程接口，不需要关心求导细节，
20 只需要定义好模型得到损失 loss 方程，然后使用 TensorFlow 实现的各种优化器（Op，Optimizer）来进行运算即可更新得到新嵌入特征。TensorFlow 的 Python 代码层面，自动求导的部分是靠各种各样的 Op 串起来的，TensorFlow 的做法是每一个 Op 在建图的时候就同时包含了它的梯度计算公式，构成前向计算图的时候会自动建立反向部分的计算图，前向计算出来的输入输出会保留下来，
25 留到后向计算的时候用完了才删除，然后在最后加上一个 Op（例如 GradientDescentOptimizer、AdamOptimizer）。Op 都继承自 Optimizer 这个类，这个类的方法非常多，几个重要方法是 minimize、compute_gradients、apply_gradients、sl

ot 系列。具体梯度如何更新到变量，由 `_apply_dense`、`_resource_apply_dense`、`_apply_sparse`、`_resource_apply_spars` 这四个方法实现。

在一些实施例中，本申请实施例提供的数据处理方法，应用于第二图形处理器，第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征，所述第二图形处理器存储有第二类型的全量稀疏特征的嵌入特征，所述第一类型与所述第二类型不同。

参见图 3D，图 3D 是本申请实施例提供的数据处理方法的流程示意图。

在步骤 201 中，第二图形处理器接收第一图形处理器发送的第二类型的第二稀疏特征

第二图形处理器接收第一图形处理器发送的，目标样本的第二类型的第二稀疏特征。

在步骤 202 中，第二图形处理器从第二类型的全量稀疏特征的嵌入特征中查询得到对应于第二稀疏特征的嵌入特征。

在步骤 203 中，第二图形处理器将对应于第二稀疏特征的嵌入特征传输至第一图形处理器。

在步骤 204 中，所述第一图形处理器对对应于第一稀疏特征的嵌入特征和对应于第二稀疏特征的嵌入特征，执行概率映射处理。

作为示例，对应于第一稀疏特征的嵌入特征是所述第一图形处理器根据目标样本的第一类型的第一稀疏特征从第一类型的全量稀疏特征的嵌入特征中获取到的。

作为示例，将对对应于第一稀疏特征的嵌入特征和对应于第二稀疏特征的嵌入特征组成对应于目标样本的稀疏特征的嵌入特征，对对应于目标样本的稀疏特征的嵌入特征执行概率映射处理，得到概率映射结果。

在步骤 205 中，第二图形处理器从第一图形处理器接收根据概率映射结果生成的更新指令，并基于更新指令更新第二类型的全量稀疏特征的嵌入特征。

在一些实施例中，第一图形处理器和第二图形处理器属于多个图形处理器，多个图形处理器存储有多个类型的全量稀疏特征的嵌入特征，第一类型包含多

个类型中的至少一个类型，第二类型包含多个类型中的至少一个类型。

在一些实施例中，第一类型的全量稀疏特征的嵌入特征的存储位置和第二类型的全量稀疏特征的嵌入特征的存储位置，是在多个类型的全量稀疏特征的嵌入特征的数据量大于阈值的情况下划分的。

5 在一些实施例中，在多个类型的全量稀疏特征的嵌入特征的数据量不大于阈值的情况下，多个类型的全量稀疏特征的嵌入特征保存在多个图形处理器中的每个图形处理器中。

按照类型将嵌入特征存储在不同的图形处理器中，可以实现嵌入特征的分布式存储，第一图形处理器获取目标样本的不同类型的稀疏特征，第一图形处
10 理器从本地获取第一类型的嵌入特征，并从第二图形处理器，获取第二类型的嵌入特征，按照类型划分方便从对应图形处理器中获取嵌入特征，可以提高获取效率，并且便于特征管理，第一图形处理器对两个类型的嵌入特征进行概率映射处理，并根据概率映射结果向第二图形处理器传输更新指令，更新指令用于指示第二图形处理器更新第二类型的嵌入特征，从而在分布式存储的前提下
15 实现了嵌入特征的更新，充分利用 GPU 计算的高效性，在满足存储需求的同时极大提升嵌入特征更新速度。

下面，将说明本申请实施例在一个实际的应用场景中的示例性应用。

在一些实施例中，为了使得新闻客户端可以向用户提供符合用户兴趣的新闻推荐，需要训练并部署新闻推荐模型（下文简称推荐模型），推荐模型的输入
20 是目标对象的嵌入特征，推荐模型的输出是目标对象点击某个新闻的概率，当概率大于概率阈值时，新闻客户端向目标对象推送该新闻，在上述过程中嵌入特征是通过账号 ID 等稀疏特征进行映射处理得到的，因此嵌入特征对于稀疏特征的表征能力极大影响推荐准确度，所以我们需要对嵌入特征进行更新以使得嵌入特征具有优秀的表达能力。

25 下面介绍嵌入特征的更新过程，数据样本可以是推荐模型的数据样本，例如针对新闻推荐模型，单个数据样本可以包括登录新闻客户端的账号的相关对象数据，这里的相关对象数据包括账号标识（账号 ID）、账号年龄、表征账号

兴趣的属性标签等等。全量数据样本中对应第一类型的稀疏特征的嵌入特征保存在服务器 200 内部署的 GPU1 内，全量数据样本中对应第二类型的稀疏特征的嵌入特征保存在服务器 200 内部署的 GPU2 内。

服务器 200 接收目标样本 A，这里的目标样本 A 被输入至 GPU1，GPU1
5 获取目标样本 A 的稀疏特征，目标样本 A 的稀疏特征包含目标样本 A 在账号标识类型的稀疏特征和目标样本 A 在年龄段类型的稀疏特征，GPU1 从账号标识类型的全量稀疏特征的嵌入特征中，获取目标样本 A 在账号标识类型的稀疏特征的嵌入特征；GPU1 从 GPU2，获取目标样本 A 在年龄段类型的稀疏特征的嵌入特征，目标样本 A 在年龄段类型的稀疏特征的嵌入特征是 GPU2 从年龄
10 段类型的全量稀疏特征的嵌入特征查询得到的；GPU1 根据目标样本 A 在账号标识类型的稀疏特征的嵌入特征和目标样本 A 在年龄段类型的稀疏特征的嵌入特征，确定目标样本 A 的稀疏特征的嵌入特征；GPU1 根据目标样本 A 的稀疏特征的嵌入特征正向传播，并根据概率映射结果向 GPU2 传输更新指令，更新指令用于指示 GPU2 更新年龄段类型的全量稀疏特征的嵌入特征。

15 推荐系统会帮用户挑选满意的商品、感兴趣的新闻、最适合的课程、符合用户兴趣的短视频、符合用户需求的音乐，参见图 4，在获知“用户信息”、“物品信息”以及“场景信息”的基础上，推荐系统要处理的问题可以较形式化地定义为：对于用户 U，在特定场景 C 下，针对海量物品信息，构建函数 $f(U, I, C)$ ，预测用户对特定候选物品 I 的喜好程度，再根据喜好程度对所有候选物
20 品进行排序，生成推荐列表。

当采用参数服务（Parameter Server）方式训练时，推荐场景下的 Embedding 存放在 CPU 上，embedding 更新也在 CPU 内完成。由 CPU 进行 Embedding 层的复杂大型运算操作，这既限制了训练的速度，又导致实际生产中无法使用比较复杂的模型，因为使用复杂模型会导致 CPU 对给定输入计算时间过长，无法及时响应请求。GPU 训练已在图像识别、文字处理等应用上取得巨大成功。
25 GPU 训练以其在卷积等数学运算上的独特效率优势，极大地提升了训练深度神经网络的速度。然而在推荐场景中，由于存在大量的稀疏样本（例如用户 ID），

每个用户 ID 在输入推荐模型之前需要被映射为对应的 embedding 向量, 因此完整的推荐模型通常体积巨大以至于单 GPU 无法进行存储。相关技术中要么为了利用高吞吐电子设备做计算而受限于电子设备的存储容量(比如 GPU 的显存), 导致 embedding 规模难以形成大规模, 要么为了容纳大规模的 embedding (比如 TB 级), 不得不使用 CPU 训练, 无法发挥数据并行训练模式的性能优势。

针对推荐系统工程化的特点, 分布式可扩展的参数服务(Parameter Server) 方案被提出, 几乎完美地解决了机器学习模型的分布式训练问题, Parameter Server 不仅被直接应用在机器学习平台上, 也被集成在 TensorFlow、MXNet 等主流的深度学习框架中, 作为机器学习分布式训练重要的解决方案。参见图 5, 可以看出, Parameter Server 由服务器节点(server 节点)和工作节点(worker 节点)组成, server 节点的主要功能是保存模型参数、接受 worker 节点计算出的局部梯度、汇总计算全局梯度, 并更新模型参数。worker 节点的主要功能是保存部分训练数据, 从 server 节点拉取最新的模型参数, 根据训练数据计算局部梯度, 上传给 server 节点。参见图 6, TensorFlow 的单机训练是在单个 worker 节点上进行的, worker 节点内部按照任务关系图的方式在不同 GPU 和 CPU 节点间进行并行计算。

下面介绍大规模 embedding, 深度学习网络中 embedding 层的作用是将稀疏输入向量转换成稠密向量, 但 embedding 层的存在往往会拖慢整个神经网络的收敛速度。embedding 层的参数数量巨大, 假设输入层的维度是 100000, embedding 层输出维度是 32, 再加 5 层 32 维的全连接层, 最后输出层维度是 10, 那么输入层到 embedding 层的参数数量是 3200000, 其余所有层的参数总数是 4416, embedding 层的权重总数占比是 99.86%, 也就是说 embedding 层的权重占了整个网络权重的绝大部分。那么训练过程可想而知, 大部分的训练时间和计算开销都被 embedding 层占据; 由于输入特征过于稀疏, 在随机梯度下降的过程中, 只有与非零特征相连的 embedding 层权重(嵌入特征)会被更新, 这进一步降低了 embedding 层的收敛速度。

大规模稀疏离散特征的使用, 导致深度模型的 embedding 层的数据量急剧

膨胀，数 TB 大小的推荐模型一度流行于业界各大头部业务场景。在大多数情况下，由于推荐类的任务对延时非常敏感，粗排场景的模型结构通常比较“矮胖”，有大量的特征输入然后映射为 embedding 向量，神经网络部分的深度相对较低。因此，特征样本内容的解析和 embedding 在整体耗时中的占比相对较高。

- 5 当采用 Parameter Server 方式训练时，推荐场景的 embedding 存放在 CPU 上，embedding 训练也在 CPU 端完成，既限制了训练速度，又因为 embedding 在整体耗时中的占比相对较高导致，从而无法支持深度较高的模型的计算，因此实际生产中无法使用比较复杂的模型，因为使用复杂模型会导致 CPU 对给定输入计算时间过长，无法及时响应请求。

- 10 本申请实施例提供的数据处理方法支持大规模 embedding 的 GPU 训练，本申请实施例提供基于 GPU 系统的训练方案，可以选用机型为 NVIDIA V100，服务器的硬件拓扑包含 8 张 GPU，并且充分考虑硬件的特性以充分发挥性能的优势。GPU 系统主要包括三个核心模块：数据模块、计算模块以及通信模块。数据模块依托 Tensorflow 的数据处理 API 完成预读取（prefetch）以及转储（dumpfile）等逻辑；计算模块用于控制每张 GPU 启动一个 TensorFlow 训练进程执行训练；在通信模块中，使用 Horovod 进程来做分布式训练的卡间通信，在每个节点上启动一个 Horovod 进程来执行对应的通信任务。

- 参见图 7，GPU0 和 GPU1 分别通过各自的 IO 接口执行 prefetch 和 dumpfile 的处理逻辑，从而 GPU0 获取输入的样本 0，GPU1 获取输入的样本 1，GPU0 对样本 0 进行特征解析，得到第一类型的稀疏特征 0 和第二类型的稀疏特征 1，GPU1 对样本 1 进行特征解析，得到第一类型的稀疏特征 0 和第二类型的稀疏特征 1，GPU0 中本地存储有所有样本的第一类型的稀疏特征 0，GPU1 中本地存储有所有样本的第二类型的稀疏特征 1，因此 GPU0 将样本 0 的第二类型的稀疏特征 1 传输至 GPU1，GPU1 将样本 1 的第一类型的稀疏特征 0 传输至 GPU0。GPU0 基于样本 0 的第一类型的稀疏特征 0 以及样本 1 的第一类型的稀疏特征 0 在哈希表 0 中进行查找，得到样本 0 的第一类型的嵌入特征 0 以及样本 1 的第一类型的嵌入特征 0，GPU1 基于样本 0 的第二类型的稀疏特征 1 以及样

本 1 的第二类型的稀疏特征 1 在哈希表 1 中进行查找，得到样本 0 的第二类型的嵌入特征 1 以及样本 1 的第二类型的嵌入特征 1，上述稀疏特征是键 (key)，例如，key1、key2 和 key3，嵌入特征的值 (value)，例如，value1、value2、value3)。GPU0 将样本 1 的第一类型的稀疏特征 0 传输至 GPU1，GPU1 将样本 0 的第二类型的稀疏特征 1 传输至 GPU0，从而 GPU0 根据样本 0 的第一类型的嵌入特征 0 和样本 0 的第二类型的嵌入特征 1 执行基于深度模型的推理处理，GPU1 根据样本 1 的第一类型的嵌入特征 0 和样本 1 的第二类型的嵌入特征 1 执行基于深度模型的推理处理。

本申请实施例的整个训练流程涉及 embedding 存储、稀疏特征处理与 embedding 切分、卡间通信等几个关键部分。

首先介绍 embedding 参数存储，在推荐系统领域，embedding 已成为处理身份类稀疏特征的常用手段，作为一种“函数映射”，Embedding 通常将高维稀疏特征映射为低维稠密向量，再进行模型端到端训练。在 TensorFlow 框架中，全部以稠密 Tensor 为基本数据单元对数据进行计算、存储以及传输。TensorFlow 也基于稠密 Tensor 提供了静态 Embedding 机制，用于存储 Embedding 的 Tensorshape 固定为[vocabulary_size,embedding_dimension], vocabulary_size 通常由身份空间决定。本申请实施例利用组件 tf.nn.dynamic_embedding 作为 embedding 的存储方式，该组件用 tf.nn.embedding_lookup_sparse 保存多个参数 (fullweights)，并且可以复用 tensorflow 原生的优化器。

接下来介绍稀疏特征处理与 embedding 切分，在实际业务场景中，通常稀疏特征的总量较多，需要的处理较为复杂。本申请实施例针对稀疏特征使用 tf.nn.dynamic_embedding 组件的动态 embedding 数据结构存储。当稀疏特征参数量较大，会导致对应的嵌入特征的数量较大 (例如，大于阈值)，GPU 单卡显存无法容纳时，可以通过按照类型划分的方式将稀疏特征对应的 embedding 划分到多张 GPU 卡的显存中存放；当稀疏特征的总数量较少，则意味着对应的嵌入特征的总数量较小 (例如不大于阈值)，GPU 单卡显存可以容纳时，使用 Replica 的方式，每张 GPU 卡的显存都放置一份全量的嵌入特征。

最后介绍卡间通信，稀疏特征（例如，身份标识类特征）由于规模较大，因此使用哈希表存储稀疏特征，由于每张 GPU 的输入样本数据不同，因此输入的稀疏特征对应的 embedding 特征向量，可能存放在其他 GPU 卡上。在训练的前向传播过程中，通过 GPU 间 AllToAll 通信，每张卡在内部的哈希表中查询稀疏特征对应的 embedding 特征向量。然后再通过 GPU 间 AllToAll 通信，将第一次 AllToAll 从其他 GPU 上拿到的特征向量原路返回，通过两次 GPU 间 AllToAll 通信，每张 GPU 的样本输入的稀疏特征都可以拿到对应的嵌入特征。在训练的反向更新过程中，会再次通过 GPU 间 AllToAll 通信，将嵌入特征的梯度返回至其他 GPU 中，每张 GPU 获取到各自的梯度后再通过优化器执行优化处理，完成大规模嵌入特征的优化。

参见图 8，GPU0 对样本 0 进行特征解析，得到第一类型的稀疏特征 0 和第二类型的稀疏特征 1，GPU1 对样本 1 进行特征解析，得到第一类型的稀疏特征 0 和第二类型的稀疏特征 1，GPU0 中本地存储有所有样本的第一类型的稀疏特征 0，GPU1 中本地存储有所有样本的第二类型的稀疏特征 1，因此 GPU0 将样本 0 的第二类型的稀疏特征 1 传输至 GPU1，GPU1 将样本 1 的第一类型的稀疏特征 0 传输至 GPU0。GPU0 基于样本 0 的第一类型的稀疏特征 0 以及样本 1 的第一类型的稀疏特征 0 在哈希表 0 中进行查找，得到样本 0 的第一类型的嵌入特征 0 以及样本 1 的第一类型的嵌入特征 0，GPU1 基于样本 0 的第二类型的稀疏特征 1 以及样本 1 的第二类型的稀疏特征 1 在哈希表 1 中进行查找，得到样本 0 的第二类型的嵌入特征 1 以及样本 1 的第二类型的嵌入特征 1。GPU0 将样本 1 的第一类型的稀疏特征 0 传输至 GPU1，GPU1 将样本 0 的第二类型的稀疏特征 1 传输至 GPU0，从而 GPU0 根据样本 0 的第一类型的嵌入特征 0 和样本 0 的第二类型的嵌入特征 1 执行基于深度模型的推理处理，得到概率映射结果，基于概率映射结果与标签值的误差确定出对应第一类型的嵌入特征 0 的梯度 0 以及对应第二类型的嵌入特征 1 的梯度 1，GPU1 根据样本 1 的第一类型的嵌入特征 0 和样本 1 的第二类型的嵌入特征 1 执行基于深度模型的推理处理，得到概率映射结果，基于概率映射结果与标签值的误差确定出对应第一类型的

嵌入特征 0 的梯度 0 以及对应第二类型的嵌入特征 1 的梯度 1。GPU0 将对应第二类型的嵌入特征 1 的梯度 1 传输至 GPU1, GPU1 的优化器基于对应第二类型的嵌入特征 1 的梯度 1 更新第二类型的嵌入特征 1, GPU1 将对应第一类型的嵌入特征 0 的梯度 0 传输至 GPU0, GPU0 的优化器基于对应第一类型的嵌入特征 0 的梯度 0 更新第一类型的嵌入特征 0。

现有的推荐系统工程技术方案要么为了利用高吞吐设备做计算, 而受限于设备的存储容量 (比如 GPU 的显存), 导致 embedding 规模难以 scale。要么为了容纳大规模的 embedding (比如 TB 级), 不得不使用 CPU 训练, 无法发挥数据并行训练模式的性能优势。

本申请实施例提出了一种能支持大规模 embedding 的 GPU 训练方案, 具有如下优点: 使用 GPU 存储 embedding 参数, 充分利用 GPU 训练的高效, 极大提升模型的训练速度; 支持大规模 embedding 参数存储, 当 embedding 参数单卡 GPU 显存放不下时, 通过高效的 GPUNCCCL 通信, 实现 embedding 拆分存储逻辑; 不失易用性, 可以无缝的对接第三方框架, 如 TensorFlow、Pytorch 等主流框架。

可以理解的是, 在本申请实施例中, 涉及到用户信息等相关的数据, 当本申请实施例运用到具体产品或技术中时, 需要获得用户许可或者同意, 且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。

下面继续说明本申请实施例提供的基于人工智能的数据处理装置 255-1 的实施为软件模块的示例性结构, 数据处理方法应用于第二图形处理器, 第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征, 第二图形处理器存储有第二类型的全量稀疏特征的嵌入特征, 第一类型与第二类型不同, 在一些实施例中, 如图 2 所示, 存储在存储器 250 的基于人工智能的数据处理装置 255-1 中的软件模块可以包括: 第一接收模块 2551, 配置为第一图形处理器获取目标样本的稀疏特征, 目标样本的稀疏特征包含所述第一类型的第一稀疏特征和所述第二类型的第二稀疏特征; 第一获取模块 2552, 配置为所述第一图形处理器从所述第二图形处理器获取从所述第二类型的全量稀疏特征的嵌入特征查询得

到的对应于所述第二稀疏特征的嵌入特征；第一返回模块 2553，配置为所述第一图形处理器从所述第二图形处理器获取从所述第二类型的全量稀疏特征的嵌入特征查询得到的对应于所述第二稀疏特征的嵌入特征；第一确定模块 2554，配置为所述第一图形处理器将对应于所述第一稀疏特征的嵌入特征和对应于所述
5 所述第二稀疏特征的嵌入特征，组成对应于所述目标样本的稀疏特征的嵌入特征；第一更新模块 2555，配置为所述第一图形处理器对所述对应于所述目标样本的稀疏特征的嵌入特征执行概率映射处理，并根据概率映射结果生成更新指令，所述更新指令用于指示所述第二图形处理器更新所述第二类型的全量稀疏特征的嵌入特征。

10 在一些实施例中，第一更新模块 2555，还配置为第一图形处理器根据概率映射结果更新第一类型的全量稀疏特征的嵌入特征。

在一些实施例中，第一图形处理器和第二图形处理器属于多个图形处理器，多个图形处理器存储有多个类型的全量稀疏特征的嵌入特征，第一类型包含多个类型中的至少一个类型，第二类型包含多个类型中的至少一个类型。

15 在一些实施例中，第一类型的全量稀疏特征的嵌入特征的存储位置和第二类型的全量稀疏特征的嵌入特征的存储位置，是在多个类型的全量稀疏特征的嵌入特征的数据量大于阈值的情况下划分的。

20 在一些实施例中，在多个类型的全量稀疏特征的嵌入特征的数据量不大于阈值的情况下，多个类型的全量稀疏特征的嵌入特征保存在多个图形处理器中的每个图形处理器中。

在一些实施例中，第一返回模块 2553，还配置为：所述第一图形处理器向所述第二图形处理器传输所述第二稀疏特征，使得所述第二图形处理器从所述第二类型的全量稀疏特征的嵌入特征中查询出对应于所述第二稀疏特征的嵌入特征；所述第一图形处理器接收对应于所述第二稀疏特征的嵌入特征。

25 在一些实施例中，第一返回模块 2553，还配置为：所述第一图形处理器从所述第一类型的全量稀疏特征的嵌入特征中查询出对应所述第一稀疏特征的嵌入特征。

在一些实施例中，在第一图形处理器获取目标样本的稀疏特征之前，第一接收模块 2551，还配置为获取第一图形处理器前一次从多个数据样本中获取到的历史目标样本；将多个数据样本中除历史目标样本的数据样本作为其他数据样本；从其他数据样本中随机获取至少一个数据样本作为目标样本。

5 在一些实施例中，在第一图形处理器获取目标样本的稀疏特征之前，第一接收模块 2551，还配置为基于图形处理器的数目对多个数据样本进行平均划分处理，得到多个样本集合，其中，每个样本集合包括至少一个数据样本，样本集合的集合数目是与图形处理器的数目相同；确定多个样本集合与多个图形处理器之间的一一对应关系；将与第一图形处理器之间具有对应关系的样本集合
10 中的数据样本，作为对应第一图形处理器的目标样本。

在一些实施例中，第一更新模块 2555，配置为以所述对应于所述目标样本的稀疏特征的嵌入特征作为推荐模型的输入，通过所述推荐模型对所述对应于所述目标样本的稀疏特征的嵌入特征执行概率映射处理，得到对应所述目标样本的推荐概率作为所述概率映射结果；所述第一图形处理器确定对应所述目标
15 样本的推荐概率与对应所述目标样本的标签值之间的误差，并基于所述误差确定所述目标样本的第二类型的全量稀疏特征的嵌入特征的梯度；生成携带有所述第二类型的全量稀疏特征的嵌入特征的梯度的更新指令。

下面继续说明本申请实施例提供的基于人工智能的数据处理装置 255-2 的实施为软件模块的示例性结构，数据处理方法应用于第二图形处理器，第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征，第二图形处理器存储有
20 第二类型的全量稀疏特征的嵌入特征，第一类型与第二类型不同，在一些实施例中，如图 2 所示，存储在存储器 250 的基于人工智能的数据处理装置 255-2 中的软件模块可以包括：第二接收模块 2556，配置为所述第二图形处理器接收所述第一图形处理器发送的，所述第二类型的第二稀疏特征；查询模块 2557，
25 配置为所述第二图形处理器从所述第二类型的全量稀疏特征的嵌入特征中查询得到对应于所述第二稀疏特征的嵌入特征；第二返回模块 2558，配置为所述第二图形处理器将对应于所述第二稀疏特征的嵌入特征传输至所述第一图形处理

器，以使所述第一图形处理器对对应于所述第一类型的第一稀疏特征的嵌入特征和对应于所述第二稀疏特征的嵌入特征，执行概率映射处理，其中，对应于所述第一稀疏特征的嵌入特征是所述第一图形处理器从所述第一类型的全量稀疏特征的嵌入特征中获取到的；第二更新模块 2559，配置为所述第二图形处理器从第一图形处理器接收根据概率映射结果生成的更新指令，并基于所述更新指令更新所述第二类型的全量稀疏特征的嵌入特征。

本申请实施例提供一种图形处理器，用于执行本申请实施例上述的基于人工智能的数据处理方法。

本申请实施例提供了一种计算机程序产品，该计算机程序产品包括计算机可执行指令，该计算机可执行指令存储在计算机可读存储介质中。电子设备的处理器从计算机可读存储介质读取该计算机可执行指令，处理器执行该计算机可执行指令，使得该电子设备执行本申请实施例上述的基于人工智能的数据处理方法，处理器是图形处理器或者中央处理器。

本申请实施例提供一种存储有计算机可执行指令的计算机可读存储介质，其中存储有计算机可执行指令，当计算机可执行指令被处理器执行时，将引起处理器执行本申请实施例提供的基于人工智能的数据处理方法，例如，如图 3 A-图 3D 示出的基于人工智能的数据处理方法，处理器是图形处理器或者中央处理器。

在一些实施例中，计算机可读存储介质可以是 FRAM、ROM、PROM、EPROM、EEPROM、闪存、磁表面存储器、光盘、或 CD-ROM 等存储器；也可以是包括上述存储器之一或任意组合的各种设备。

在一些实施例中，计算机可执行指令可以采用程序、软件、软件模块、脚本或代码的形式，按任意形式的编程语言（包括编译或解释语言，或者声明性或过程性语言）来编写，并且其可按任意形式部署，包括被部署为独立的程序或者被部署为模块、组件、子例程或者适合在计算环境中使用的其它单元。

作为示例，计算机可执行指令可以但不一定对应于文件系统中的文件，可以被存储在保存其它程序或数据的文件的一部分，例如，存储在超文本标记

语言（HTML，Hyper Text Markup Language）文档中的一个或多个脚本中，存储在专用于所讨论的程序的单个文件中，或者，存储在多个协同文件（例如，存储一个或多个模块、子程序或代码部分的文件）中。

5 作为示例，计算机可执行指令可被部署为在一个电子设备上执行，或者在位于一个地点的多个电子设备上执行，又或者，在分布在多个地点且通过通信网络互连的多个电子设备上执行。

综上所述，按照类型将嵌入特征存储在不同的图形处理器中，可以实现嵌入特征的分布式存储，第一图形处理器获取目标样本的不同类型的稀疏特征，第一图形处理器从本地获取第一类型的嵌入特征，并从第二图形处理器，获取
10 第二类型的嵌入特征，按照类型划分方便从对应图形处理器中获取嵌入特征，可以提高获取效率，并且便于特征管理，第一图形处理器对两个类型的嵌入特征进行概率映射处理，并根据概率映射结果向第二图形处理器传输更新指令，更新指令用于指示第二图形处理器更新第二类型的嵌入特征，从而在分布式存储的前提下实现了嵌入特征的更新，充分利用 GPU 计算的高效性，在满足存储
15 需求的同时极大提升嵌入特征更新速度。

以上所述，仅为本申请的实施例而已，并非用于限定本申请的保护范围。凡在本申请的精神和范围之内所作的任何修改、等同替换和改进等，均包括在本申请的保护范围之内。

权利要求书

1、一种数据处理方法，所述数据处理方法应用于第一图形处理器，所述第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征，第二图形处理器存储有第二类型的全量稀疏特征的嵌入特征，所述第一类型与所述第二类型不同；

5 所述方法包括：

所述第一图形处理器获取目标样本的稀疏特征，所述目标样本的稀疏特征包含所述第一类型的第一稀疏特征和所述第二类型的第二稀疏特征；

所述第一图形处理器从所述第一类型的全量稀疏特征的嵌入特征中，获取对应于所述第一稀疏特征的嵌入特征；

10 所述第一图形处理器从所述第二图形处理器获取从所述第二类型的全量稀疏特征的嵌入特征查询得到的对应于所述第二稀疏特征的嵌入特征；

所述第一图形处理器将对应于所述第一稀疏特征的嵌入特征和对应于所述第二稀疏特征的嵌入特征，组成对应于所述目标样本的稀疏特征的嵌入特征；

15 所述第一图形处理器对所述对应于所述目标样本的稀疏特征的嵌入特征执行概率映射处理，并根据概率映射结果生成更新指令，所述更新指令用于指示所述第二图形处理器更新所述第二类型的全量稀疏特征的嵌入特征。

2、根据权利要求1所述的方法，其中，所述方法还包括：

所述第一图形处理器根据所述概率映射结果更新所述第一类型的全量稀疏特征的嵌入特征。

20 3、根据权利要求1或2所述的方法，其中，所述第一图形处理器和所述第二图形处理器属于多个图形处理器，所述多个图形处理器存储有多个类型的全量稀疏特征的嵌入特征，所述第一类型包含所述多个类型中的至少一个类型，所述第二类型包含所述多个类型中的至少一个类型。

4、根据权利要求3所述的方法，其中，所述第一类型的全量稀疏特征的嵌入特征的存储位置和所述第二类型的全量稀疏特征的嵌入特征的存储位置，是在所述多个类型的全量稀疏特征的嵌入特征的数据量大于阈值的情况下划分的。

5、根据权利要求3所述的方法，其中，在所述多个类型的全量稀疏特征的

嵌入特征的数据量不大于阈值的情况下，所述多个类型的全量稀疏特征的嵌入特征保存在所述多个图形处理器中的每个图形处理器中。

6、根据权利要求 1 至 5 中任意一项所述的方法，其中，所述第一图形处理器从所述第二图形处理器获取从所述第二类型的全量稀疏特征的嵌入特征查询得到的对应于所述第二稀疏特征的嵌入特征，包括：

所述第一图形处理器向所述第二图形处理器传输所述第二稀疏特征，使得所述第二图形处理器从所述第二类型的全量稀疏特征的嵌入特征中查询出对应于所述第二稀疏特征的嵌入特征；

所述第一图形处理器接收对应于所述第二稀疏特征的嵌入特征。

7、根据权利要求 1 至 6 中任意一项所述的方法，其中，所述第一图形处理器从所述第一类型的全量稀疏特征的嵌入特征中，获取对应于所述第一稀疏特征的嵌入特征，包括：

所述第一图形处理器从所述第一类型的全量稀疏特征的嵌入特征中查询出对应所述第一稀疏特征的嵌入特征。

8、根据权利要求 1 至 7 中任意一项所述的方法，其中，在所述第一图形处理器获取目标样本的稀疏特征之前，所述方法还包括：

获取所述第一图形处理器前一次从所述多个数据样本中获取到的历史目标样本；

将所述多个数据样本中除所述历史目标样本的数据样本作为其他数据样本；
从所述其他数据样本中随机获取至少一个数据样本作为所述目标样本。

9、根据权利要求 1 至 8 中任意一项所述的方法，其中，在所述第一图形处理器获取目标样本的稀疏特征之前，所述方法还包括：

基于所述图形处理器的数目对所述多个数据样本进行平均划分处理，得到多个样本集合，其中，每个所述样本集合包括至少一个数据样本，所述样本集合的集合数目与所述图形处理器的数目相同；

确定所述多个样本集合与所述多个图形处理器之间的一一对应关系；

将与所述第一图形处理器之间具有所述对应关系的样本集合中的数据样本，

作为对应所述第一图形处理器的目标样本。

10、根据权利要求 1 至 9 中任意一项所述的方法，其中，

所述第一图形处理器根据所述对应于所述目标样本的稀疏特征的嵌入特征执行概率映射处理，包括：

5 以所述对应于所述目标样本的稀疏特征的嵌入特征作为推荐模型的输入，通过所述推荐模型对所述对应于所述目标样本的稀疏特征的嵌入特征执行概率映射处理，得到对应所述目标样本的推荐概率作为所述概率映射结果；

所述根据概率映射结果生成更新指令，包括：

10 所述第一图形处理器确定对应所述目标样本的推荐概率与对应所述目标样本的标签值之间的误差，并基于所述误差确定所述目标样本的第二类型的全量稀疏特征的嵌入特征的梯度；

生成携带有所述第二类型的全量稀疏特征的嵌入特征的梯度的更新指令。

11、一种数据处理方法，所述数据处理方法应用于第二图形处理器，第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征，所述第二图形处理器
15 存储有第二类型的全量稀疏特征的嵌入特征，所述第一类型与所述第二类型不同；

所述方法包括：

所述第二图形处理器接收所述第一图形处理器发送的，所述第二类型的第二稀疏特征；

20 所述第二图形处理器从所述第二类型的全量稀疏特征的嵌入特征中查询得到对应于所述第二稀疏特征的嵌入特征；

所述第二图形处理器将对应于所述第二稀疏特征的嵌入特征传输至所述第一图形处理器，以使所述第一图形处理器对对应于第一稀疏特征的嵌入特征和对应于所述第二稀疏特征的嵌入特征，执行概率映射处理，其中，所述对应于
25 第一稀疏特征的嵌入特征是所述第一图形处理器从所述第一类型的全量稀疏特征的嵌入特征中获取得到的；

所述第二图形处理器从第一图形处理器接收根据概率映射结果生成的更新

指令，并基于所述更新指令更新所述第二类型的全量稀疏特征的嵌入特征。

12、根据权利要求 11 所述的方法，其中，所述第一图形处理器和所述第二图形处理器属于多个图形处理器，所述多个图形处理器存储有多个类型的全量稀疏特征的嵌入特征，所述第一类型包含所述多个类型中的至少一个类型，所述
5 第二类型包含所述多个类型中的至少一个类型。

13、根据权利要求 12 所述的方法，其中，所述第一类型的全量稀疏特征的嵌入特征的存储位置和所述第二类型的全量稀疏特征的嵌入特征的存储位置，是在所述多个类型的全量稀疏特征的嵌入特征的数据量大于阈值的情况下划分的。

10 14、根据权利要求 12 所述的方法，其中，在所述多个类型的全量稀疏特征的嵌入特征的数据量不大于阈值的情况下，所述多个类型的全量稀疏特征的嵌入特征保存在所述多个图形处理器中的每个图形处理器中。

15 15、一种数据处理装置，所述数据处理方法应用于第一图形处理器，第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征，所述第二图形处理器存储有第二类型的全量稀疏特征的嵌入特征，所述第一类型与所述第二类型不同，所述装置包括：

第一接收模块，配置为所述第一图形处理器获取目标样本的稀疏特征，所述目标样本的稀疏特征包含所述第一类型的第一稀疏特征和所述第二类型的第二稀疏特征；

20 第一获取模块，配置为所述第一图形处理器从所述第二图形处理器获取从所述第二类型的全量稀疏特征的嵌入特征查询得到的对应于所述第二稀疏特征的嵌入特征；

第一返回模块，配置为所述第一图形处理器从所述第二图形处理器获取从所述第二类型的全量稀疏特征的嵌入特征查询得到的对应于所述第二稀疏特征的
25 嵌入特征；

第一确定模块，配置为所述第一图形处理器将对应于所述第一稀疏特征的嵌入特征和对应于所述第二稀疏特征的嵌入特征，组成对应于所述目标样本的

稀疏特征的嵌入特征;

第一更新模块,配置为所述第一图形处理器对所述对应于所述目标样本的稀疏特征的嵌入特征执行概率映射处理,并根据概率映射结果生成更新指令,所述更新指令用于指示所述第二图形处理器更新所述第二类型的全量稀疏特征
5 的嵌入特征。

16、一种数据处理装置,所述数据处理方法应用于第二图形处理器,第一图形处理器存储有第一类型的全量稀疏特征的嵌入特征,所述第二图形处理器存储有第二类型的全量稀疏特征的嵌入特征,所述第一类型与所述第二类型不同;
10 同;

所述装置包括:

第二接收模块,配置为所述第二图形处理器接收所述第一图形处理器发送的,所述第二类型的第二稀疏特征;

查询模块,配置为所述第二图形处理器从所述第二类型的全量稀疏特征的
15 嵌入特征中查询得到对应于所述第二稀疏特征的嵌入特征;

第二返回模块,配置为所述第二图形处理器将对应于所述第二稀疏特征的嵌入特征传输至所述第一图形处理器,以使所述第一图形处理器对对应于第一稀疏特征的嵌入特征和对应于所述第二稀疏特征的嵌入特征,执行概率映射处理,其中,所述对应于第一稀疏特征的嵌入特征是所述第一图形处理器从所述
20 第一类型的全量稀疏特征的嵌入特征中获取得到的;

第二更新模块,配置为所述第二图形处理器从第一图形处理器接收根据概率映射结果生成的更新指令,并基于所述更新指令更新所述第二类型的全量稀疏特征的嵌入特征。

17、一种图形处理器,所述图形处理器用于执行权利要求 1 至 10 或者权利要求 11 至 14 任一项所述的数据处理方法。
25

18、一种电子设备,所述电子设备包括:

存储器,用于存储计算机可执行指令;

图形处理器，用于执行所述存储器中存储的计算机可执行指令时，实现权利要求 1 至 10 或者权利要求 11 至 14 任一项所述的数据处理方法。

19、一种计算机可读存储介质，存储有计算机可执行指令，所述计算机可执行指令被处理器执行时实现权利要求 1 至 10 或者权利要求 11 至 14 任一项所述的数据处理方法，所述处理区是中央处理器或者图形处理器。

20、一种计算机程序产品，包括计算机可执行指令，所述计算机可执行指令被处理器执行时实现权利要求 1 至 10 或者权利要求 11 至 14 任一项所述的数据处理方法，所述处理区是中央处理器或者图形处理器。

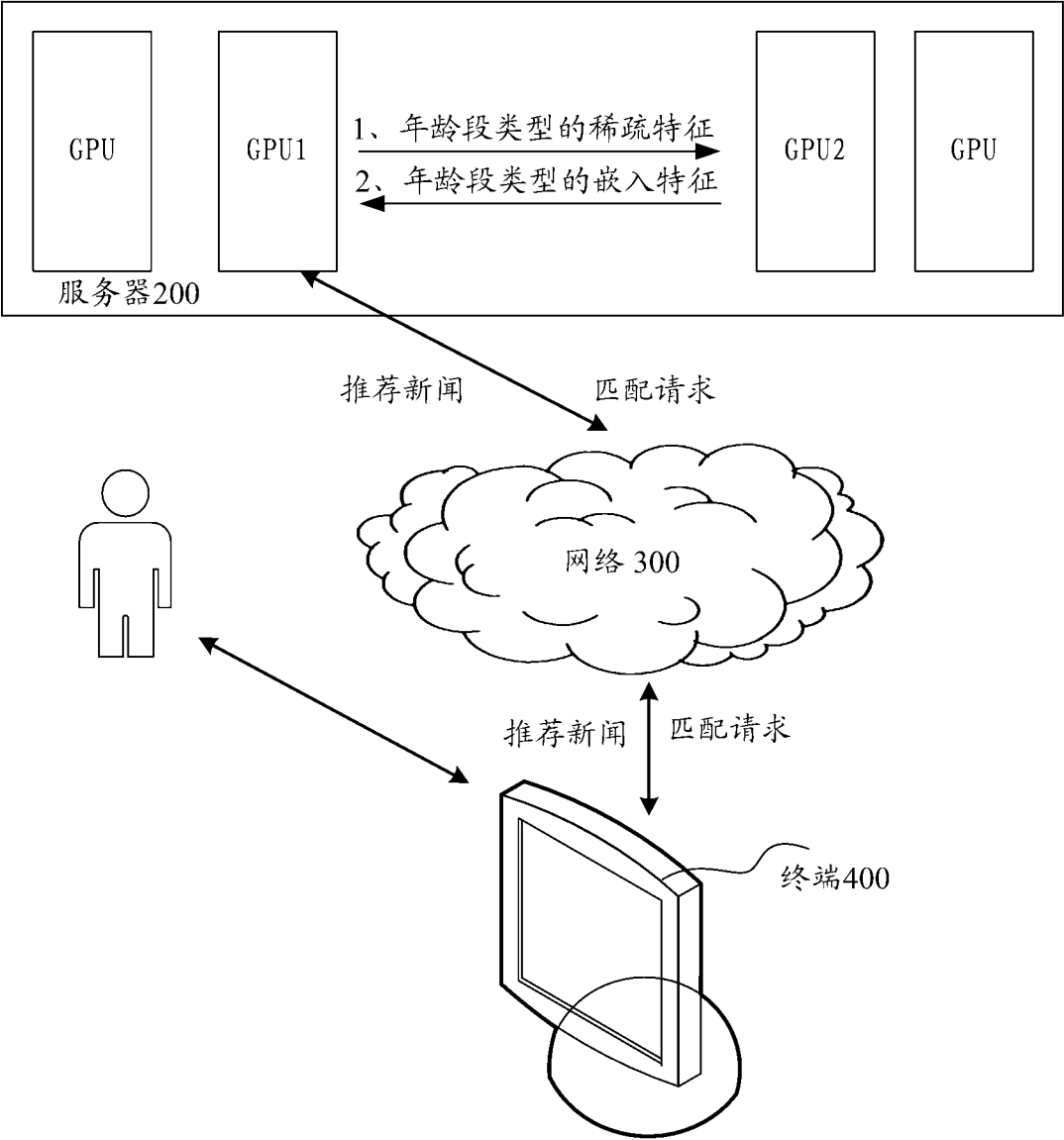


图 1A

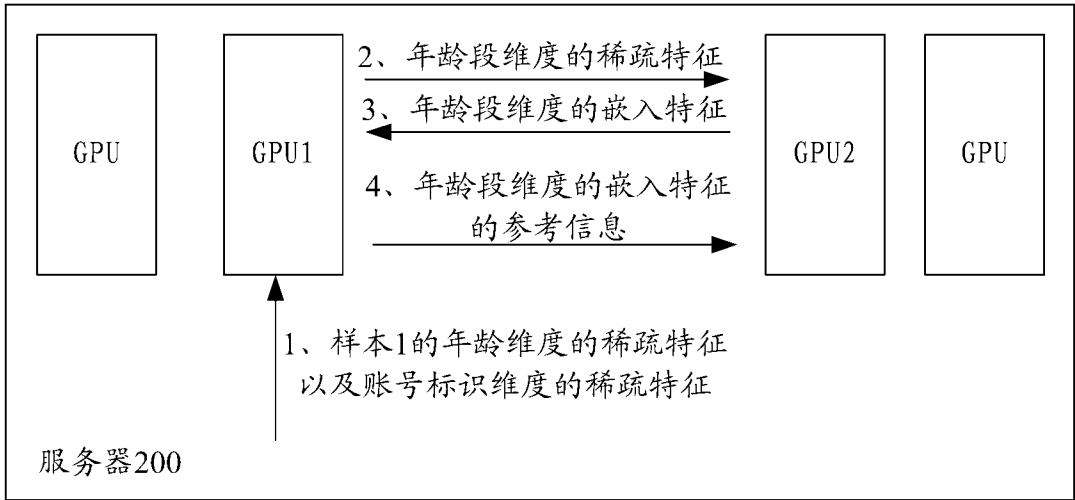


图 1B

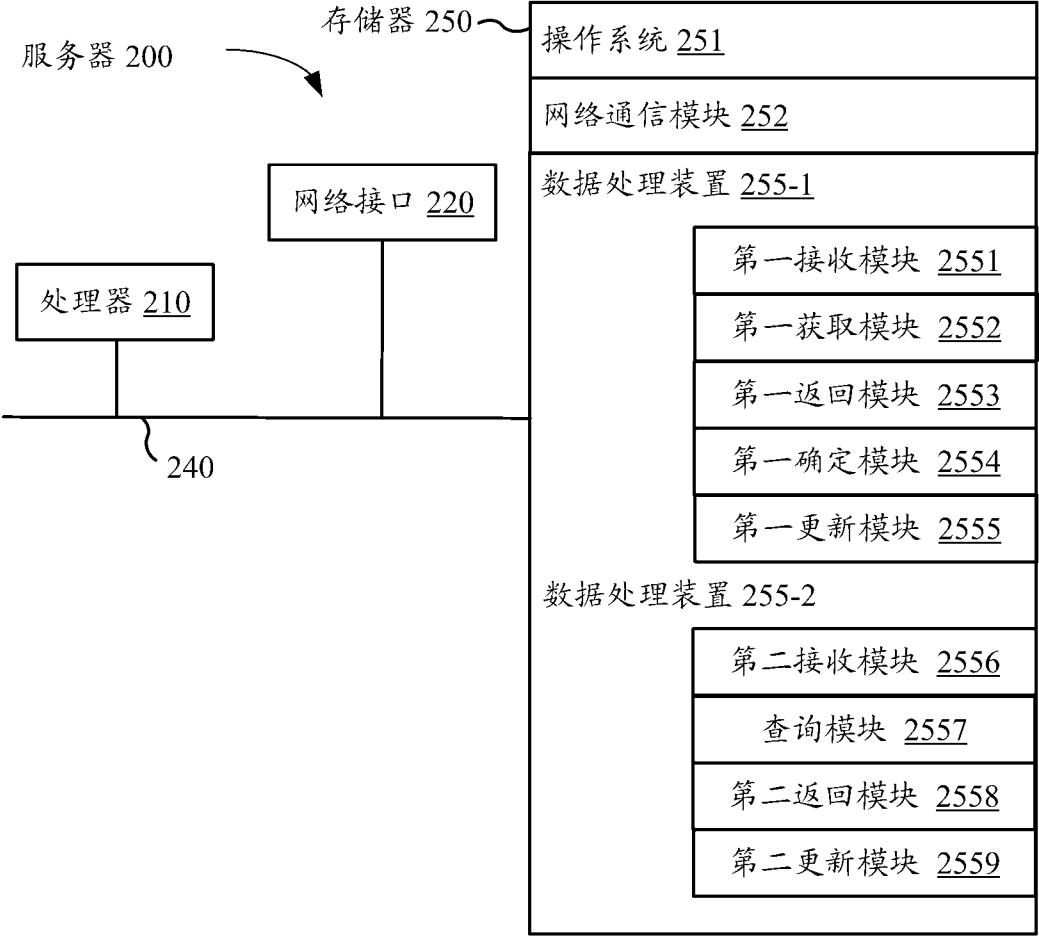


图 2

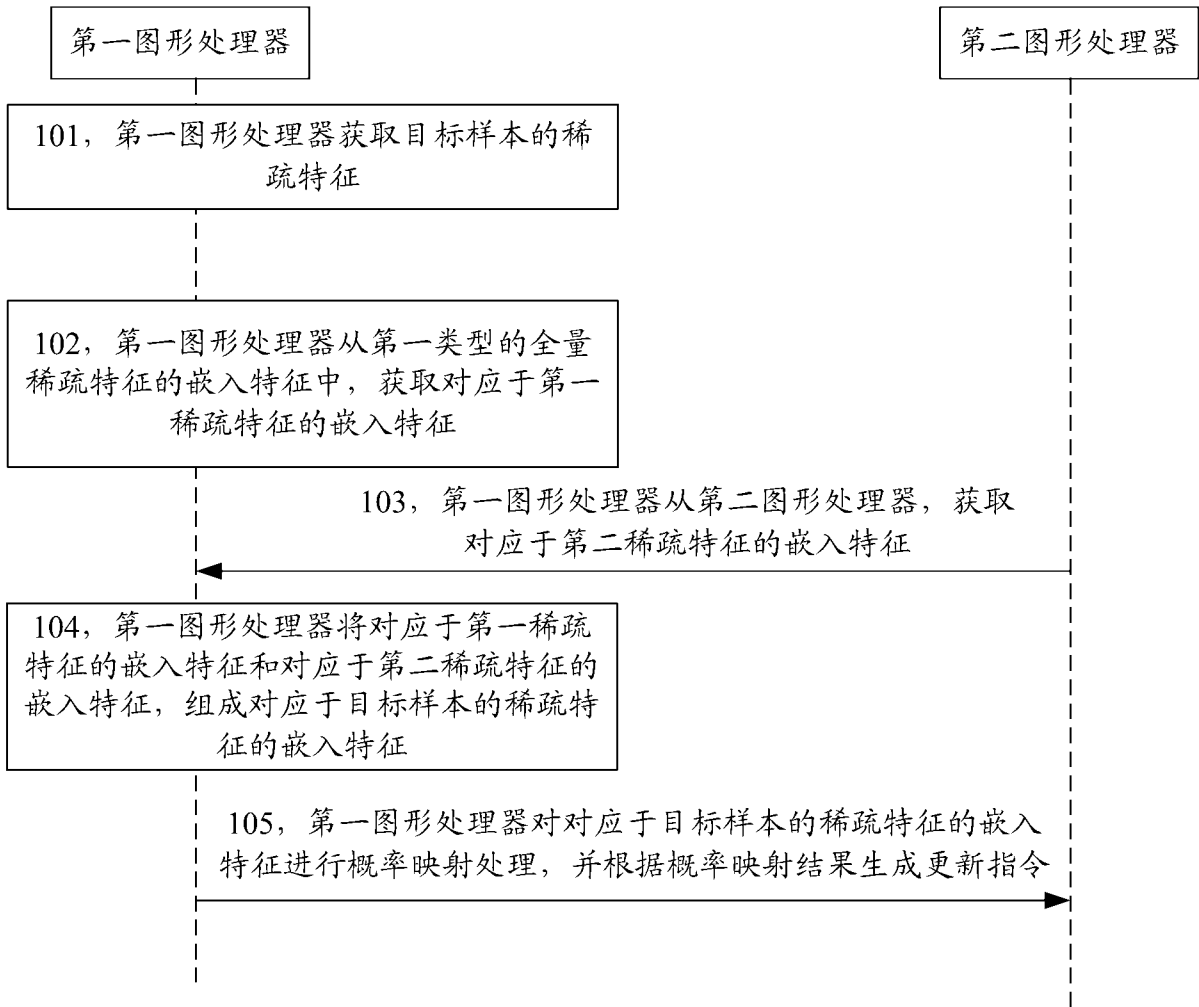


图 3A

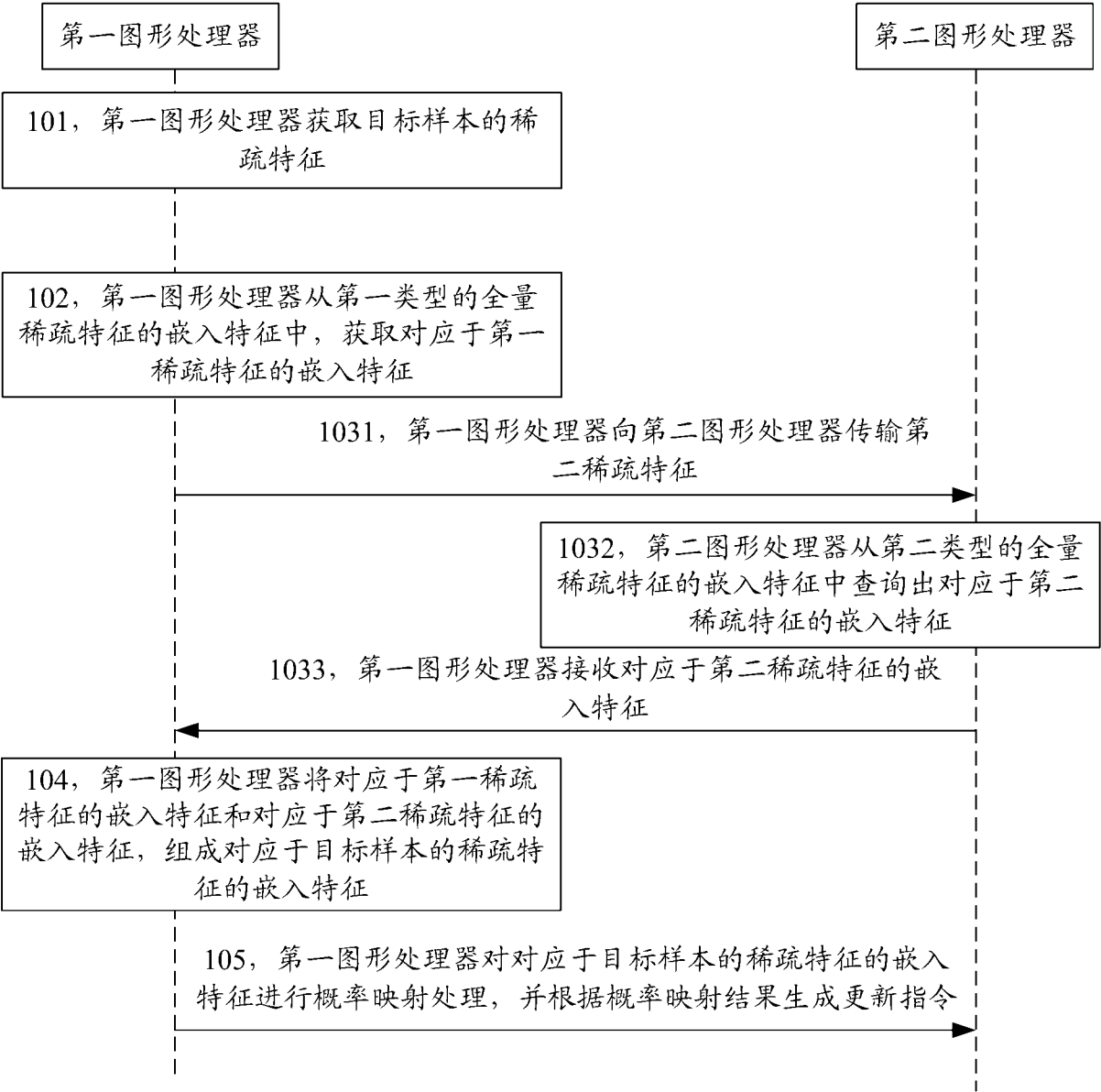


图 3B

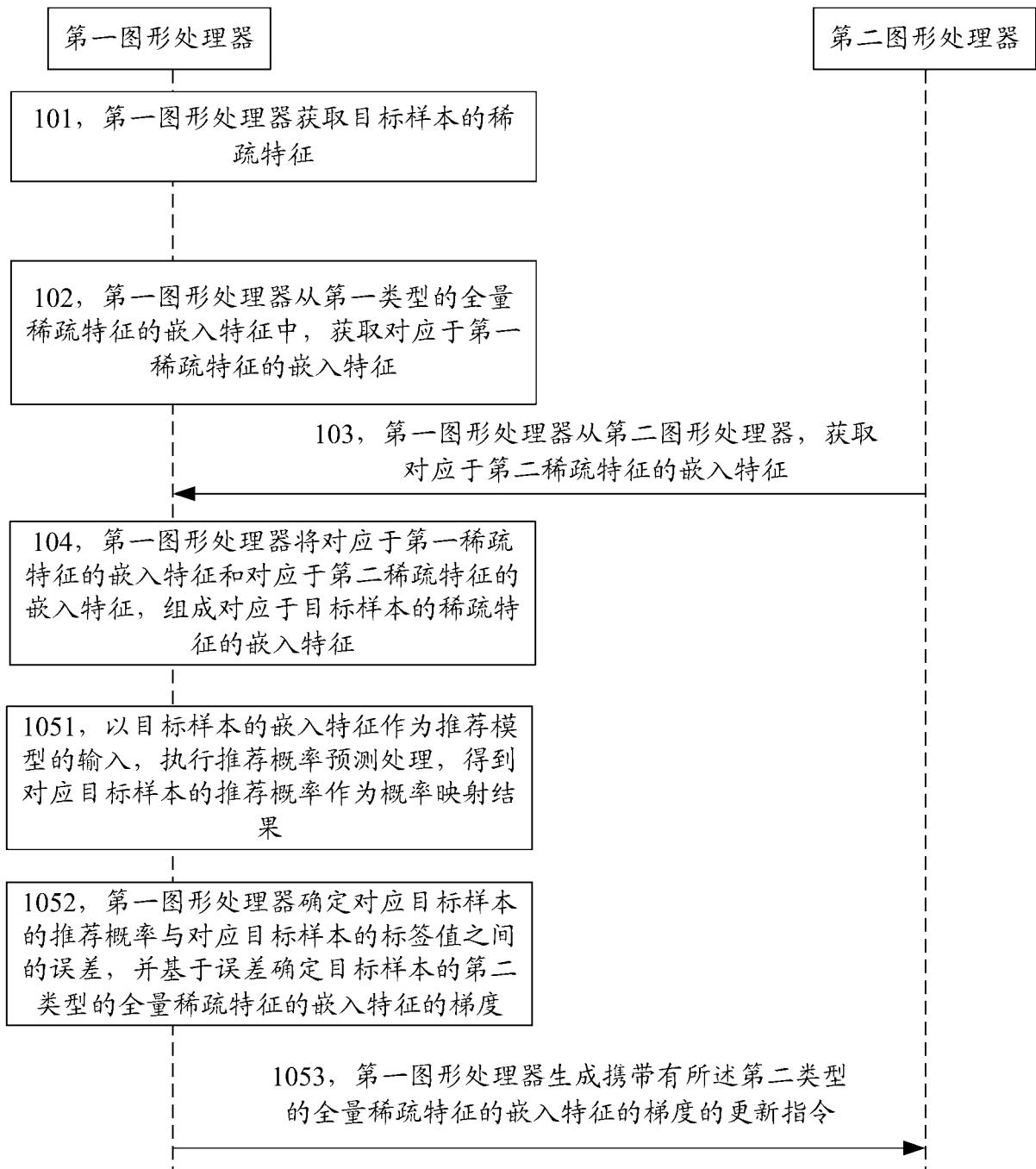


图 3C

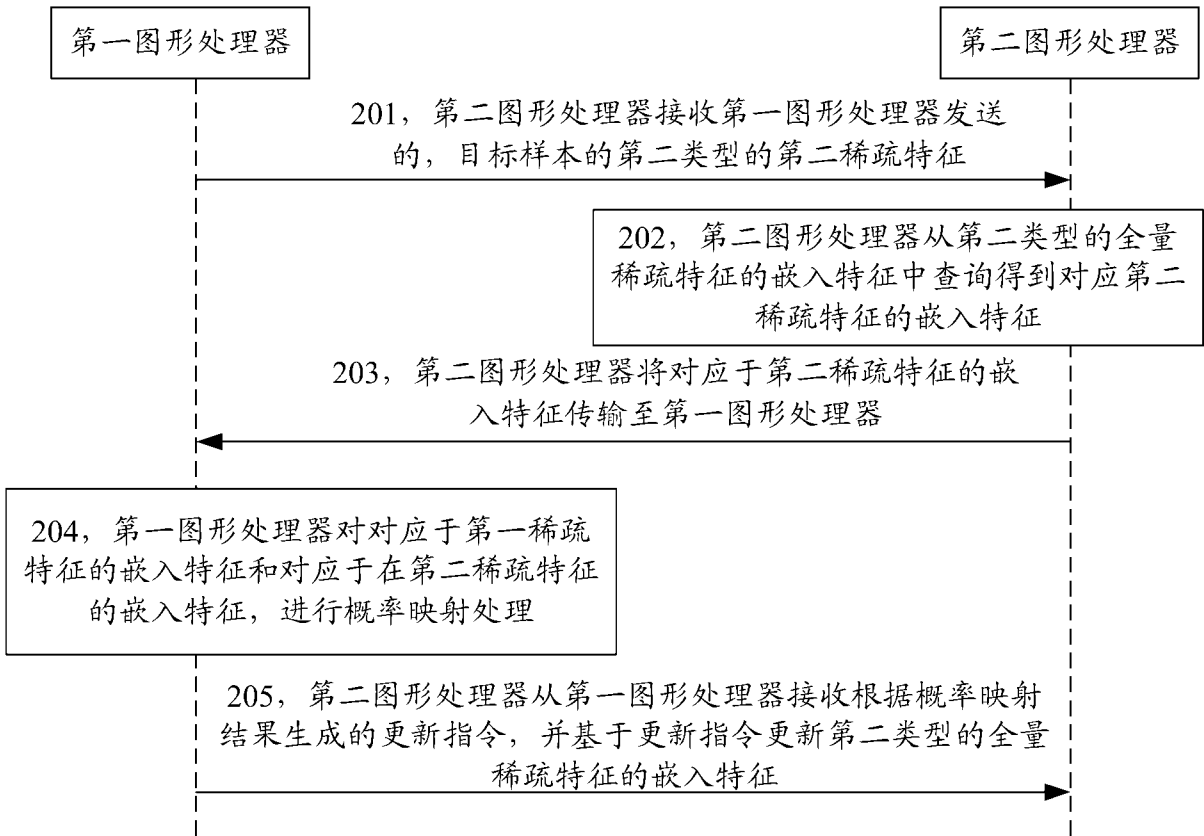


图 3D

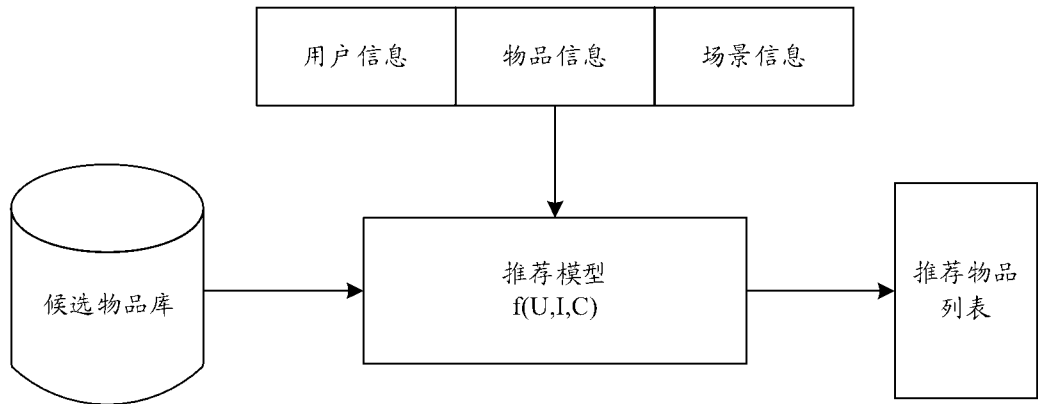


图 4

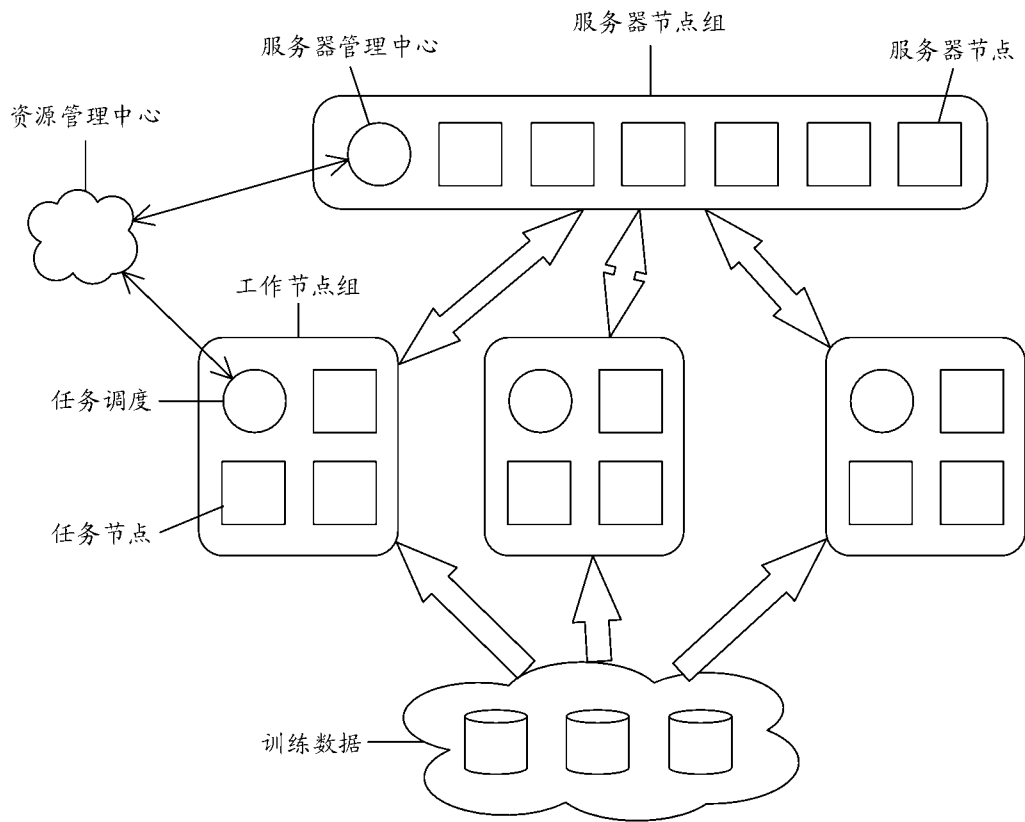


图 5

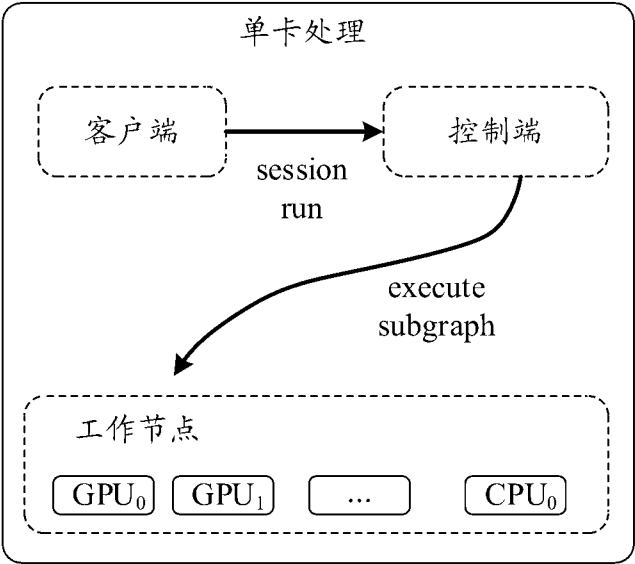


图 6

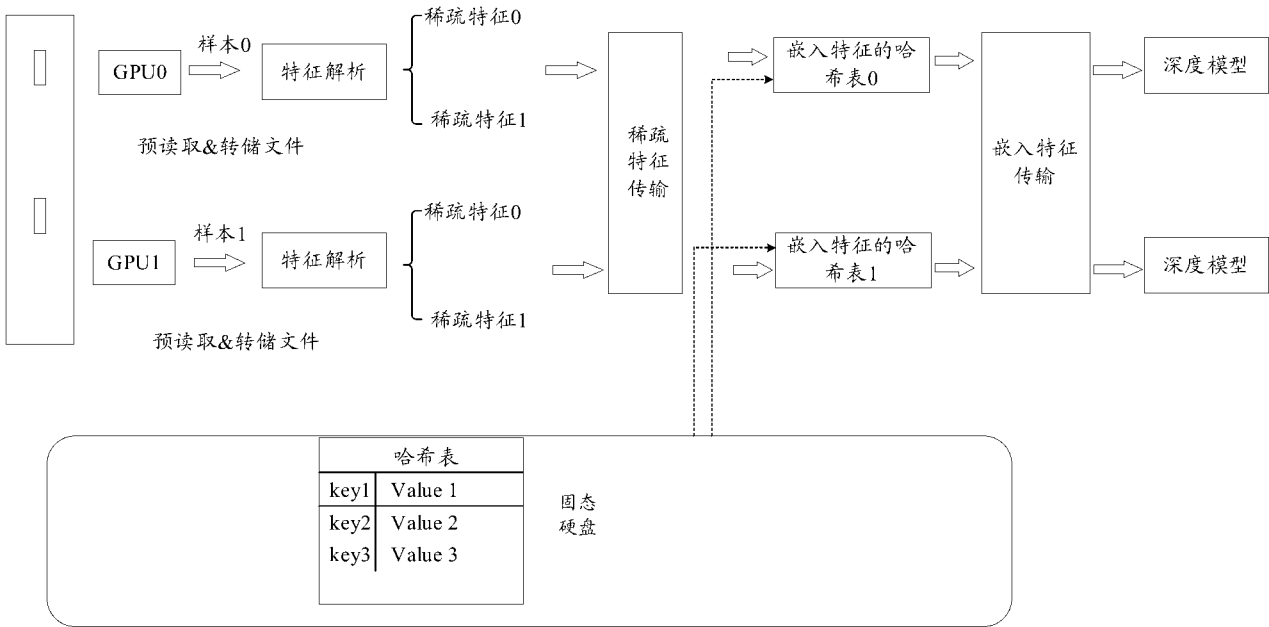


图 7

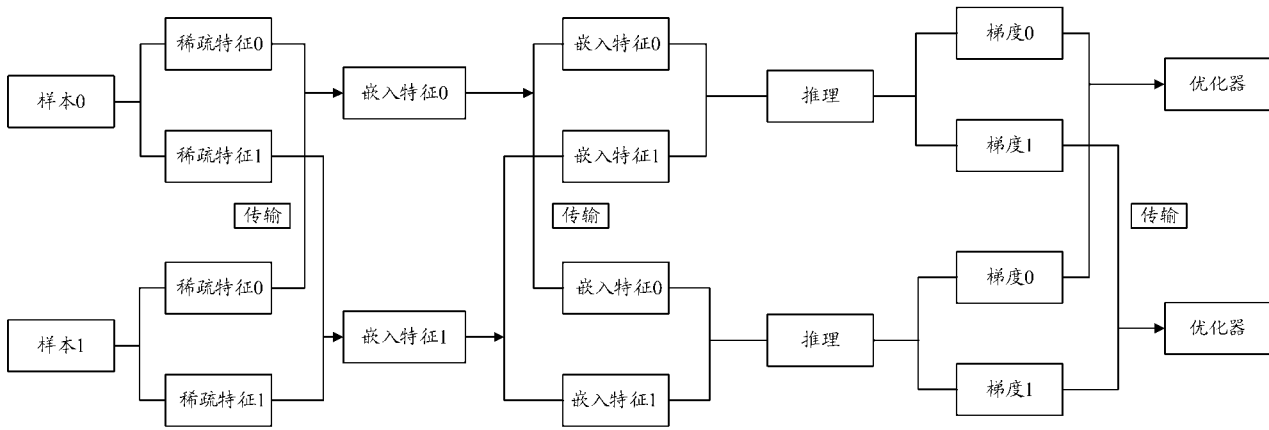


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/135891

A. CLASSIFICATION OF SUBJECT MATTER

G06F9/54(2006.01)i; G06F9/50(2006.01)i; G06N3/098(2023.01)i; G06N3/0985(2023.01)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:G06F G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, DWPI, VEN, CNTXT, CNABS: 稀疏特征, 嵌入特征, 图形处理器, 概率映射, embedded feature, GPU, sparse feature, graphics processor, probability mapping

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 116974781 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 31 October 2023 (2023-10-31) description, paragraphs 73 and 125-131, and claims 1-20	1-20
A	CN 113971428 A (BEIJING DAJIA INTERCONNECTION INFORMATION TECHNOLOGY CO., LTD.) 25 January 2022 (2022-01-25) description, paragraphs 50-55, and claim 1	1-20
A	CN 112884157 A (BEIJING DAJIA INTERCONNECTION INFORMATION TECHNOLOGY CO., LTD.) 01 June 2021 (2021-06-01) entire document	1-20
A	CN 109919436 A (HUARONG RONGTONG (BEIJING) TECHNOLOGY CO., LTD.) 21 June 2019 (2019-06-21) entire document	1-20
A	US 2021319078 A1 (MICROSOFT TECHNOLOGY LICENSING, LLC) 14 October 2021 (2021-10-14) entire document	1-20



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

19 February 2024

Date of mailing of the international search report

23 February 2024

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/CN2023/135891

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116974781	A	31 October 2023	None			
CN	113971428	A	25 January 2022	None			
CN	112884157	A	01 June 2021	None			
CN	109919436	A	21 June 2019	None			
US	2021319078	A1	14 October 2021	US	2023028789	A1	26 January 2023
				WO	2021211199	A1	21 October 2021
				EP	4136551	A1	22 February 2023
				CN	115398423	A	25 November 2022
				IN	202217057399	A	04 August 2023

A. 主题的分类

G06F9/54(2006.01)i; G06F9/50(2006.01)i; G06N3/098(2023.01)i; G06N3/0985(2023.01)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:G06F G06N

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNKI, DWPI, VEN, CNTXT, CNABS: 稀疏特征, 嵌入特征, 图形处理器, 概率映射, embedded feature, GPU, sparse feature, graphics processor, probability mapping

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 116974781 A (腾讯科技(深圳)有限公司) 2023年10月31日 (2023 - 10 - 31) 说明书73,125-131段, 权利要求1-20	1-20
A	CN 113971428 A (北京达佳互联信息技术有限公司) 2022年1月25日 (2022 - 01 - 25) 说明书第50-55段, 权利要求1	1-20
A	CN 112884157 A (北京达佳互联信息技术有限公司) 2021年6月1日 (2021 - 06 - 01) 全文	1-20
A	CN 109919436 A (华融融通(北京)科技有限公司) 2019年6月21日 (2019 - 06 - 21) 全文	1-20
A	US 2021319078 A1 (MICROSOFT TECHNOLOGY LICENSING, LLC) 2021年10月14日 (2021 - 10 - 14) 全文	1-20

☐ 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

2024年2月19日

国际检索报告邮寄日期

2024年2月23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

董刚

电话号码 (+86) 010-53961348

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/CN2023/135891

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	116974781	A	2023年10月31日	无	
CN	113971428	A	2022年1月25日	无	
CN	112884157	A	2021年6月1日	无	
CN	109919436	A	2019年6月21日	无	
US	2021319078	A1	2021年10月14日	US	2023028789 A1 2023年1月26日
				WO	2021211199 A1 2021年10月21日
				EP	4136551 A1 2023年2月22日
				CN	115398423 A 2022年11月25日
				IN	202217057399 A 2023年8月4日