



(12) 发明专利

(10) 授权公告号 CN 111489245 B

(45) 授权公告日 2023.05.05

(21) 申请号 202010347492.7
 (22) 申请日 2020.04.28
 (65) 同一申请的已公布的文献号
 申请公布号 CN 111489245 A
 (43) 申请公布日 2020.08.04
 (73) 专利权人 上海风秩科技有限公司
 地址 200333 上海市普陀区真光路1473弄3号4层、5层、6层
 (72) 发明人 方依 陈羲 梁新敏
 (74) 专利代理机构 北京超成律师事务所 11646
 专利代理师 邓超
 (51) Int.Cl.
 G06Q 30/0601 (2023.01)

(56) 对比文件
 CN 108694672 A, 2018.10.23
 CN 109389461 A, 2019.02.26
 CN 111028060 A, 2020.04.17
 CN 108932640 A, 2018.12.04
 CN 109978470 A, 2019.07.05
 CN 110348868 A, 2019.10.18
 CN 109934618 A, 2019.06.25
 CN 107464177 A, 2017.12.12
 CN 110503498 A, 2019.11.26
 CN 106295951 A, 2017.01.04
 CN 110490691 A, 2019.11.22
 审查员 徐书芳

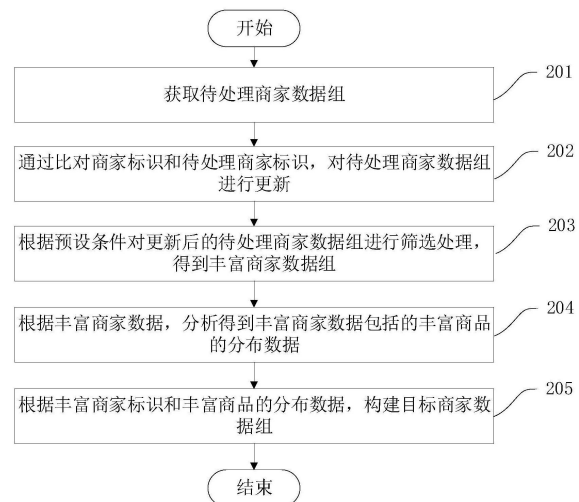
权利要求书3页 说明书9页 附图7页

(54) 发明名称

数据处理方法、装置、计算机设备和可读存储介质

(57) 摘要

本申请实施例提供了一种数据处理方法、装置、计算机设备和可读存储介质,涉及电商技术领域,数据处理方法应用于服务器,服务器存储有商家订单数据表,商家订单数据表包括商家标识,数据处理方法包括:获取待处理商家数据组;通过比对商家标识和待处理商家标识,对待处理商家数据组进行更新;根据预设条件对更新后的待处理商家数据组进行筛选处理,得到丰富商家数据组;根据丰富商家数据,分析得到丰富商家数据包括的丰富商品的分布数据;根据丰富商家标识和丰富商品的分布数据,构建目标商家数据组,能够将线上商家的数据处理为可参考的目标商家数据组。



1. 一种数据处理方法,其特征在于,应用于服务器,所述服务器存储有商家订单数据表,所述商家订单数据表包括商家标识、商品标识和订单标识,所述方法包括:

获取待处理商家数据组,所述待处理商家数据组包括待处理商家标识和所述待处理商家标识对应的待处理商家数据;

通过比对所述商家标识和所述待处理商家标识,对所述待处理商家数据组进行更新;

根据预设条件对更新后的待处理商家数据组进行筛选处理,得到丰富商家数据组,所述丰富商家数据组包括丰富商家标识和所述丰富商家标识对应的丰富商家数据;

根据所述丰富商家数据,分析得到所述丰富商家数据包括的丰富商品的分布数据;

根据所述丰富商家标识和所述丰富商品的分布数据,构建目标商家数据组;

其中,所述待处理商家数据包括待处理产品列表和待处理订单列表,所述通过比对所述商家标识和所述待处理商家标识,对所述待处理商家数据组进行更新的步骤,包括:

比对所述商家标识和所述待处理商家标识;

当所述商家标识中出现与所述待处理商家标识不相同的待添加商家标识时,从所述商家订单数据表中获取所述待添加商家标识对应的待添加商品标识和待添加订单标识;

根据所述待添加商品标识构建待添加产品列表;

根据所述待添加订单标识构建待添加订单列表;

根据所述待添加产品列表和所述待添加订单列表得到所述待添加商家标识对应的待添加商家数据;

当所述商家标识中出现与所述待处理商家标识相同的待更新商家标识时,从所述商家订单数据表中获取所述待更新商家标识对应的待更新商品标识和待更新订单标识;

将所述待更新商品标识加入所述待处理产品列表,并将所述待更新订单标识加入待处理订单列表得到更新后的待处理商家数据;

根据所述待添加商家标识和所述待添加商家数据和/或所述待更新商家标识和所述更新后的待处理商家数据,更新所述待处理商家数据组。

2. 根据权利要求1所述的方法,其特征在于,所述服务器还包括商品数据表,所述商品数据表包括商品详细数据;

所述将所述待更新商品标识加入所述待处理产品列表的步骤包括:

根据所述待更新商品标识,从所述商品数据表中获取所述待更新商品标识对应的商品详细数据;

将所述待更新商品标识和所述商品详细数据加入所述待处理产品列表。

3. 根据权利要求1所述的方法,其特征在于,所述预设条件为:

$$\text{len}(P) \times \alpha \leq \text{len}(O) \quad \text{where } \alpha > 1$$

其中, $\text{len}(P)$ 为待处理产品列表的待处理商品数, $\text{len}(O)$ 为待处理订单的待处理订单数, α 为预设调节系数。

4. 根据权利要求1所述的方法,其特征在于,所述丰富商家数据包括丰富商家产品列表;

所述根据所述丰富商家数据,分析得到所述丰富商家数据包括的丰富商品的分布数据的步骤,包括:

根据所述丰富商家产品列表,获取所述丰富商家产品列表中包括的丰富商品类别,并

获取所述丰富商品类别的数目；

将所述丰富商品类别和所述丰富商品类别的数目作为所述分布数据。

5. 根据权利要求4所述的方法,其特征在於,在所述根据所述丰富商家数据,分析得到所述丰富商家数据包括的丰富商品的分布数据的步骤之前,所述方法还包括:

获取所述丰富商家产品列表,并根据所述丰富商家产品列表得到所述丰富商家产品列表包括的丰富商家产品数量;

判断所述丰富商家产品数量是否超过预设数量阈值;

若是,则将所述丰富商家数据保留;

若否,则将所述丰富商家数据删除。

6. 一种数据处理装置,其特征在於,应用于服务器,所述服务器存储有商家订单数据表,所述商家订单数据表包括商家标识、商品标识和订单标识,所述装置包括:

获取模块,用于获取待处理商家数据组,所述待处理商家数据组包括待处理商家标识和所述待处理商家标识对应的待处理商家数据;

比对模块,用于通过比对所述商家标识和所述待处理商家标识,对所述待处理商家数据组进行更新;

筛选模块,用于根据预设条件对更新后的待处理商家数据组进行筛选处理,得到丰富商家数据组,所述丰富商家数据组包括丰富商家标识和所述丰富商家标识对应的丰富商家数据;

分析模块,用于根据所述丰富商家数据,分析得到所述丰富商家数据包括的丰富商品的分布数据;

构建模块,用于根据所述丰富商家标识和所述丰富商品的分布数据,构建目标商家数据组;

其中,所述待处理商家数据包括待处理产品列表和待处理订单列表,所述比对模块具体用于:

比对所述商家标识和所述待处理商家标识;当所述商家标识中出现与所述待处理商家标识不相同的待添加商家标识时,从所述商家订单数据表中获取所述待添加商家标识对应的待添加商品标识和待添加订单标识;根据所述待添加商品标识构建待添加产品列表;根据所述待添加订单标识构建待添加订单列表;根据所述待添加产品列表和所述待添加订单列表得到所述待添加商家标识对应的待添加商家数据;当所述商家标识中出现与所述待处理商家标识相同的待更新商家标识时,从所述商家订单数据表中获取所述待更新商家标识对应的待更新商品标识和待更新订单标识;将所述待更新商品标识加入所述待处理产品列表,并将所述待更新订单标识加入待处理订单列表得到更新后的待处理商家数据;根据所述待添加商家标识和所述待添加商家数据和/或所述待更新商家标识和所述更新后的待处理商家数据,更新所述待处理商家数据组。

7. 一种计算机设备,其特征在於,所述计算机设备包括处理器及存储有计算机指令的非易失性存储器,所述计算机指令被所述处理器执行时,所述计算机设备执行权利要求1-5中任意一项所述的数据处理方法。

8. 一种可读存储介质,其特征在於,所述可读存储介质包括计算机程序,所述计算机程序运行时控制所述可读存储介质所在计算机设备执行权利要求1-5中任意一项所述的数据

处理方法。

数据处理方法、装置、计算机设备和可读存储介质

技术领域

[0001] 本申请涉及电商技术领域,具体而言,涉及一种数据处理方法、装置、计算机设备和可读存储介质。

背景技术

[0002] 目前,随着电商行业的发展,越来越多的线上商店被开设。

[0003] 然而各个线上商店之间的数据一般是隔离的,并不能直接进行相互参考,因此想要通过参考其他线上商店进行商品推荐十分的困难。然而想要从现有的数据中获取能够参考的数据,如果直接从真实的线上商店通过人工进行逐个分析数据十分的不便,会占用大量人力物力;如果从现有的电子商务相关的开源数据库中获取的数据,获取的数据也不能直接使用,无法满足需求。

[0004] 有鉴于此,如何提供一种能够将线上商家的数据处理为可参考数据的方案,是本领域技术人员需要考虑的。

发明内容

[0005] 本申请提供了一种数据处理方法、装置、计算机设备和可读存储介质。

[0006] 本申请的实施例可以这样实现:

[0007] 第一方面,实施例提供一种数据处理方法,应用于服务器,所述服务器存储有商家订单数据表,所述商家订单数据表包括商家标识,所述方法包括:

[0008] 获取待处理商家数据组,所述待处理商家数据组包括待处理商家标识和所述待处理商家标识对应的待处理商家数据;

[0009] 通过比对所述商家标识和所述待处理商家标识,对所述待处理商家数据组进行更新;

[0010] 根据预设条件对更新后的待处理商家数据组进行筛选处理,得到丰富商家数据组,所述丰富商家数据组包括丰富商家标识和所述丰富商家标识对应的丰富商家数据;

[0011] 根据所述丰富商家数据,分析得到所述丰富商家数据包括的丰富商品的分布数据;

[0012] 根据所述丰富商家标识和所述丰富商品的分布数据,构建目标商家数据组。

[0013] 在可选的实施方式中,所述待处理商家数据包括待处理产品列表和待处理订单列表,所述商家订单数据表还包括商品标识和订单标识;

[0014] 所述通过比对所述商家标识和所述待处理商家标识,对所述待处理商家数据组进行更新的步骤,包括:

[0015] 比对所述商家标识和所述待处理商家标识;

[0016] 当所述商家标识中出现与所述待处理商家标识不相同的待添加商家标识时,从所述商家订单数据表中获取所述待添加商家标识对应的待添加商品标识和待添加订单标识;

[0017] 根据所述待添加商品标识构建待添加产品列表;

- [0018] 根据所述待添加订单标识构建待添加订单列表；
- [0019] 根据所述待添加产品列表和所述待添加订单列表得到所述待添加商家标识对应的待添加商家数据；
- [0020] 当所述商家标识中出现与所述待处理商家标识相同的待更新商家标识时，从所述商家订单数据表中获取所述待更新商家标识对应的待更新商品标识和待更新订单标识；
- [0021] 将所述待更新商品标识加入所述待处理产品列表，并将所述待更新订单标识加入待处理订单列表得到更新后的待处理商家数据；
- [0022] 根据所述待添加商家标识和所述待添加商家数据和/或所述待更新商家标识和所述更新后的待处理商家数据，更新所述待处理商家数据组。
- [0023] 在可选的实施方式中，所述服务器还包括商品数据表，所述商品数据表包括商品详细数据；
- [0024] 所述将所述待更新商品标识加入所述待处理产品列表的步骤包括：
- [0025] 根据所述待更新商品标识，从所述商品数据表中获取所述待更新商品标识对应的商品详细数据；
- [0026] 将所述待更新商品标识和所述商品详细数据加入所述待处理产品列表。
- [0027] 在可选的实施方式中，所述预设条件为：
- [0028] $\text{len}(P) \times \alpha \leq \text{len}(O)$ where $\alpha > 1$
- [0029] 其中， $\text{len}(P)$ 为待处理产品列表的待处理商品数， $\text{len}(O)$ 为待处理订单的待处理订单数， α 为预设调节系数。
- [0030] 在可选的实施方式中，所述丰富商家数据包括丰富商家产品列表；
- [0031] 所述根据所述丰富商家数据，分析得到所述丰富商家数据包括的丰富商品的分布数据的步骤，包括：
- [0032] 根据所述丰富商家产品列表，获取所述丰富商家产品列表中包括的丰富商品类别，并获取所述丰富商品类别的数目；
- [0033] 将所述丰富商品类别和所述丰富商品类别的数目作为所述分布数据。
- [0034] 在可选的实施方式中，在所述根据所述丰富商家数据，分析得到所述丰富商家数据包括的丰富商品的分布数据的步骤之前，所述方法还包括：
- [0035] 获取所述丰富商家产品列表，并根据所述丰富商家产品列表得到所述丰富商家产品列表包括的丰富商家产品数量；
- [0036] 判断所述丰富商家产品数量是否超过预设数量阈值；
- [0037] 若是，则将所述丰富商家数据保留；
- [0038] 若否，则将所述丰富商家数据删除。
- [0039] 第二方面，实施例提供一种数据处理装置，应用于服务器，所述服务器存储有商家订单数据表，所述商家订单数据表包括商家标识，所述装置包括：
- [0040] 获取模块，用于获取待处理商家数据组，所述待处理商家数据组包括待处理商家标识和所述待处理商家标识对应的待处理商家数据；
- [0041] 比对模块，用于通过比对所述商家标识和所述待处理商家标识，对所述待处理商家数据组进行更新；
- [0042] 筛选模块，用于根据预设条件对更新后的待处理商家数据组进行筛选处理，得到

丰富商家数据组,所述丰富商家数据组包括丰富商家标识和所述丰富商家标识对应的丰富商家数据;

[0043] 分析模块,用于根据所述丰富商家数据,分析得到所述丰富商家数据包括的丰富商品的分布数据;

[0044] 构建模块,用于根据所述丰富商家标识和所述丰富商品的分布数据,构建目标商家数据组。

[0045] 在可选的实施方式中,所述待处理商家数据包括待处理产品列表和待处理订单列表,所述商家订单数据表还包括商品标识和订单标识;

[0046] 所述比对模块具体用于:

[0047] 比对所述商家标识和所述待处理商家标识;当所述商家标识中出现与所述待处理商家标识不相同的待添加商家标识时,从所述商家订单数据表中获取所述待添加商家标识对应的待添加商品标识和待添加订单标识;根据所述待添加商品标识构建待添加产品列表;根据所述待添加订单标识构建待添加订单列表;根据所述待添加产品列表和所述待添加订单列表得到所述待添加商家标识对应的待添加商家数据;当所述商家标识中出现与所述待处理商家标识相同的待更新商家标识时,从所述商家订单数据表中获取所述待更新商家标识对应的待更新商品标识和待更新订单标识;将所述待更新商品标识加入所述待处理产品列表,并将所述待更新订单标识加入待处理订单列表得到更新后的待处理商家数据;根据所述待添加商家标识和所述待添加商家数据和/或所述待更新商家标识和所述更新后的待处理商家数据,更新所述待处理商家数据组。

[0048] 第三方面,实施例提供一种计算机设备,所述计算机设备包括处理器及存储有计算机指令的非易失性存储器,所述计算机指令被所述处理器执行时,所述计算机设备执行前述实施方式中任意一项所述的数据处理方法。

[0049] 第四方面,实施例提供一种可读存储介质,所述可读存储介质包括计算机程序,所述计算机程序运行时控制所述可读存储介质所在计算机设备执行前述实施方式中任意一项所述的数据处理方法。

[0050] 本申请实施例的有益效果包括,例如:

[0051] 采用本申请实施例提供的数据处理方法、装置、计算机设备和可读存储介质,通过获取待处理商家数据组,进而比对所述商家标识和所述待处理商家标识,对所述待处理商家数据组进行更新;然后根据预设条件对更新后的待处理商家数据组进行筛选处理,得到丰富商家数据组,再根据所述丰富商家数据,分析得到所述丰富商家数据包括的丰富商品的分布数据;最后根据所述丰富商家标识和所述丰富商品的分布数据,构建目标商家数据组,能够将线上商家的数据处理为可参考的目标商家数据组。

附图说明

[0052] 为了更清楚地说明本申请实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本申请的某些实施例,因此不应被看作是对范围的限定,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他相关的附图。

[0053] 图1为本申请实施例提供的一种线上商城的应用场景示意图;

- [0054] 图2为本申请实施例提供的一种数据处理方法的步骤流程示意图；
- [0055] 图3为本申请实施例提供的一种服务器的结构示意图；
- [0056] 图4为本申请实施例提供的另一种数据处理方法的步骤流程示意图；
- [0057] 图5为本申请实施例提供的另一种数据处理方法的步骤流程示意图；
- [0058] 图6为本申请实施例提供的另一种数据处理方法的步骤流程示意图；
- [0059] 图7为本申请实施例提供的另一种数据处理方法的步骤流程示意图；
- [0060] 图8为本申请实施例提供的一种数据处理装置的结构示意框图；
- [0061] 图9为本申请实施例提供的一种计算机设备的结构示意框图。
- [0062] 图标:100-计算机设备;110-数据处理装置;1101-获取模块;1102-比对模块;1103-筛选模块;1104-分析模块;1105-构建模块;111-存储器;112-处理器;113-通信单元。

具体实施方式

[0063] 为使本申请实施例的目的、技术方案和优点更加清楚,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本申请一部分实施例,而不是全部的实施例。通常在此处附图中描述和示出的本申请实施例的组件可以以各种不同的配置来布置和设计。

[0064] 因此,以下对在附图中提供的本申请的实施例的详细描述并非旨在限制要求保护的本申请的范围,而是仅仅表示本申请的选定实施例。基于本申请中的实施例,本领域普通技术人员在没有作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。

[0065] 应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步定义和解释。

[0066] 此外,若出现术语“第一”、“第二”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0067] 需要说明的是,在不冲突的情况下,本申请的实施例中的特征可以相互结合。

[0068] 随着电商行业的发展,各种微商城、线上商城也随之开设,而新开设的线上商城,一般是没有太多商品相关数据(用户购买行为数据等)的,又由于各个线上商城之间的数据并不是互通的,因此想要参考其他已经有较多商品相关数据的线上商城比较困难,如图1所示,同一用户在线上商城1和在线上商城2上的数据并不互通,相互之间难以关联。在现有技术中,若直接人工从其他线上商城中获取数据,整个过程会消耗大量的人力物力,若从电商相关的开源数据库中获取数据,获取的数据也并不能直接使用,不具备参考价值。基于此,本申请实施例提供了一种数据处理方法,应用于服务器,服务器存储有商家订单数据表,商家订单数据表包括商家标识,如图2所示,数据处理方法包括步骤201至步骤205。

[0069] 步骤201,获取待处理商家数据组。

[0070] 其中,待处理商家数据组包括待处理商家标识和待处理商家标识对应的待处理商家数据。

[0071] 步骤202,通过比对商家标识和待处理商家标识,对待处理商家数据组进行更新。

[0072] 步骤203,根据预设条件对更新后的待处理商家数据组进行筛选处理,得到丰富商家数据组。

- [0073] 其中,丰富商家数据组包括丰富商家标识和丰富商家标识对应的丰富商家数据。
- [0074] 步骤204,根据丰富商家数据,分析得到丰富商家数据包括的丰富商品的分布数据。
- [0075] 步骤205,根据丰富商家标识和丰富商品的分布数据,构建目标商家数据组。
- [0076] 请参照图3,本申请实施例中服务器可以是Kaggle(数据发掘和预测竞赛)平台,在本申请实施例中可以将olist_order_items_dataset(商家订单相关数据表)作为order items(即商家订单数据表),在本申请实施例中,获取到处理商家数据组的方式可以是维护待处理商家dictionary(字典),维护待处理商家dictionary的key(键)可以是seller id(即待处理商家标识),而value(值)可以是一个seller对象(即待处理商家标识对应的待处理商家数据)。
- [0077] 在前述基础上,待处理商家数据包括待处理产品列表和待处理订单列表,商家订单数据表还包括商品标识和订单标识。本申请实施例还提供一种通过比对商家标识和待处理商家标识,对待处理商家数据组进行更新的示例,可如图4所示,以通过步骤202-1至步骤202-8实现。
- [0078] 步骤202-1,比对商家标识和待处理商家标识。
- [0079] 步骤202-2,当商家标识中出现与待处理商家标识不相同的待添加商家标识时,从商家订单数据表中获取待添加商家标识对应的待添加商品标识和待添加订单标识。
- [0080] 步骤202-3,根据待添加商品标识构建待添加产品列表。
- [0081] 步骤202-4,根据待添加订单标识构建待添加订单列表。
- [0082] 步骤202-5,根据待添加产品列表和待添加订单列表得到待添加商家标识对应的待添加商家数据。
- [0083] 步骤202-6,当商家标识中出现与待处理商家标识相同的待更新商家标识时,从商家订单数据表中获取待更新商家标识对应的待更新商品标识和待更新订单标识。
- [0084] 步骤202-7,将待更新商品标识加入待处理产品列表,并将待更新订单标识加入待处理订单列表得到更新后的待处理商家数据。
- [0085] 步骤202-8,根据待添加商家标识和待添加商家数据和/或待更新商家标识和更新后的待处理商家数据,更新待处理商家数据组。
- [0086] 请再次参考图3,order items还可以包括product id(即商品标识)和order id(即订单标识),其中,product id可以用于表示是哪一件商品,而order id可以用于表示是哪一个订单,在本申请实施例中,seller对象可以封装了seller id的products列表(待处理产品列表)以及seller id的order列表(待处理订单列表)。在本申请实施例中,products列表中包括的每个值可以是item对象(商品对象)的封装,而item对象内部可以封装商品的外观数据、商品名称等信息。在本申请实施例中,可以遍历order items表,而order items表中每个订单可以对应一个order id、product id和seller id(例如一个订单,可以对应该订单的订单号、订单中包括的商品名称以及该订单属于哪个商家),在遍历order items表时,可以将order items表中的seller id(即商家标识)和待处理商家dictionary中已有的key(即待处理商家标识)进行比对。
- [0087] 当order items表中出现待处理商家dictionary中的key不包括的seller id时(即商家标识中出现与待处理商家标识不相同的待添加商家标识时),可以在待处理商家

dictionary中新增加一个key,该key即为order items表中待处理商家dictionary中的key不包括的seller id,可以为其初始化新的seller对象(即根据待添加商品标识构建待添加产品列表和根据待添加订单标识构建待添加订单列表),构建完成后,有了两个空的列表,然后便可以将该seller id对应的order id加入待添加订单列表,将该seller id对应的product id加入待添加产品列表。例如,在遍历order items表时,有一seller id为商家A,该seller id在待处理商家dictionary中的key并未出现,那么可以将商家A作为新的key在待处理商家dictionary中构建,同时初始化行的seller对象,得到待添加产品列表和待添加订单列表,然后将商家A对应的商品A加入待添加产品列表,将商家A对应的订单A加入待添加订单列表。

[0088] 而当order items表中出现待处理商家dictionary中的key包括的seller id时(即当商家标识中出现与待处理商家标识相同的待更新商家标识时),可以从order items表中获取该seller id对应的product id和order id(即待更新商品标识和待更新订单标识),并将待更新商品标识和待更新订单标识分别加入该seller id对应的待处理产品列表和待处理订单列表中。例如,待处理商家dictionary中的一个key为商家B,在遍历order items表时有一seller id也是商家B,可以将order items表中商家B对应的商品B加入在待处理商家dictionary中key为商家B对应的待处理产品列表中,将order items表中商家B对应的订单B加入待处理商家dictionary中key为商家B对应的待处理订单列表中。

[0089] 通过上述流程,能够根据将order items表中的seller id是否已经在待处理商家dictionary中作为key出现,对待处理商家dictionary进行对应的更新。

[0090] 在此基础上,服务器还包括商品数据表,商品数据表包括商品详细数据。本申请实施例提供一种将待更新商品标识加入待处理产品列表的示例,可以由以下步骤实现,请参考图5。

[0091] 步骤202-7-1,根据待更新商品标识,从商品数据表中获取待更新商品标识对应的商品详细数据。

[0092] 步骤202-7-2,将待更新商品标识和商品详细数据加入待处理产品列表。

[0093] Kaggle平台中还包括olist_products_dataset,其包括了商品更为详细的信息,可以作为product表(即商品数据表)。应当理解的是,在对待处理商家dictionary进行更新时,当order items表中出现待处理商家dictionary中的key包括的seller id时,还可以从product表中获取该seller id对应的product id(即待更新商品标识)所对应的商品详细数据,然后将更新商品标识和其商品详细数据加入order items表中出现待处理商家dictionary中的key包括的seller id对应的seller对象中。例如,product表中有商品C的外观参数、历史价格商品详细数据,而待处理商家dictionary中的一个key也为商家C,在遍历order items表时,便可以将,product表中商家C对应的商品C的外观参数、历史价格商品详细数据也加入待处理商家dictionary中key为商家C对应的待处理产品列表中。通过上述的步骤,可以从order items表中获取更加详细的商品信息更新至待处理商家dictionary中。

[0094] 在此基础上,本申请实施例提供的预设条件为:

[0095] $\text{len}(P) \times \alpha \leq \text{len}(O)$ where $\alpha > 1$

[0096] 其中, $\text{len}(P)$ 为待处理产品列表的待处理商品数, $\text{len}(O)$ 为待处理订单的待处理

订单数, α 为预设调节系数。

[0097] 在进行筛选的过程中,可以通过上述条件进行筛选。例如,更新后的待处理商家dictionary中,有key为商家D,可以获得商家D对应的products列表和order列表,然后获取商家D对应的products列表中的待处理商品数(即商家D拥有的商品种类数量)和商家D对应的order列表中的待处理订单数,商家D对应的products列表中的待处理商品数可以是10种,商家D对应的order列表中的待处理订单数可以是152个,预设调节系数 α 可以设置为10,此时 $\text{len}(P) \times \alpha$ 即 $10 \times 10 = 100$,小于152,可以认为商家D为丰富商家,在将待处理商家dictionary中每个key都经过预设条件过滤后,便能够得到丰富商家数据组。在本申请实施例的其他实施方式中,预设调节系数 α 可以根据具体的商品种类进行调节,通过上述步骤,能够根据待处理商品数和待处理订单数来对待处理商家dictionary中的商家数据进行清洗,将行为数据较为丰富的商家数据保留下来,即丰富商家数据dictionary。

[0098] 除了上述方案,在根据丰富商家数据,分析得到丰富商家数据包括的丰富商品的分布数据的步骤之前,本申请实施例提供一种丰富商家筛选示例,可以通过以下步骤实现,请参考图6。

[0099] 步骤301,获取丰富商家产品列表,并根据丰富商家产品列表得到丰富商家产品列表包括的丰富商家产品数量。

[0100] 步骤302,判断丰富商家产品数量是否超过预设数量阈值。

[0101] 若是,则执行步骤303。

[0102] 若否,则执行步骤304。

[0103] 步骤303,将丰富商家数据保留。

[0104] 步骤304,将丰富商家数据删除。

[0105] 在获得了丰富商家数据dictionary后,还可以进行进一步地筛除工作,可以将丰富商家数据dictionary中各个key对应的products列表(丰富商家产品列表)获取,然后判断丰富商家产品数量是否超过预设数量阈值,即各个key对应的商品数量是否足够多,由于最终获取的数据可以被用于训练来解决现有技术中从开源数据平台中获取的数据无法直接使用的问题,太少的数据不利于训练,且少量数据测试具有偶然性,因此会将商品数量不超过预设数量阈值的seller id(即key)以及其对应的seller对象从丰富商家数据dictionary中删除,将商品数量超过预设数量阈值的seller id(即key)以及其对应的seller对象在丰富商家数据dictionary中保留。

[0106] 在上述基础上,丰富商家数据包括丰富商家产品列表。本申请实施例还提供一种根据丰富商家数据,分析得到丰富商家数据包括的丰富商品的分布数据的示例,可以通过以下步骤实现,如图7所示。

[0107] 步骤204-1,根据丰富商家产品列表,获取丰富商家产品列表中包括的丰富商品类别,并获取丰富商品类别的数目。

[0108] 步骤204-2,将丰富商品类别和丰富商品类别的数目作为分布数据。

[0109] 在经过前述的筛除步骤之后,可以得到目标商家数据组,目标商家数据组可以认为是新维护的一个字典,目标商家数据dictionary,其中,由经过筛选后的待处理商家dictionary中的key作为目标商家数据dictionary的key,而目标商家数据dictionary的value则可以是每个key(即seller id,商家)包含的category(即目标商家数据)的分布结

果,其中category的分布结果具体可以是指每个目标商品的富商品类别(set)和目标商品类别的数目(dict)。例如,目标商家数据dictionary的有key为商户E,商户E对应的products列表包括奶茶、蛋糕和咖啡,其中奶茶包括红枣奶茶和珍珠奶茶,蛋糕包括慕斯蛋糕、奶油蛋糕、抹茶蛋糕、原味蛋糕、绿豆蛋糕、冰淇淋蛋糕和草莓蛋糕,咖啡包括卡布奇诺和焦糖玛奇朵,那么对于商户E的category的分布结果,包括set为3,其中,奶茶对应的dict为2,蛋糕对应的dict为7,咖啡对应的dict为2。通过上述步骤,便能获取目标商家数据包括的目标商品的分布数据,能够将获取的数据用于给新开设的线上商城进行商品推荐等操作。

[0110] 本申请实施例还提供一种数据处理装置110,如图8所示,应用于服务器,服务器存储有商家订单数据表,商家订单数据表包括商家标识,装置包括:

[0111] 获取模块1101,用于获取待处理商家数据组,待处理商家数据组包括待处理商家标识和待处理商家标识对应的待处理商家数据。

[0112] 比对模块1102,用于通过比对商家标识和待处理商家标识,对待处理商家数据组进行更新。

[0113] 筛选模块1103,用于根据预设条件对更新后的待处理商家数据组进行筛选处理,得到丰富商家数据组,丰富商家数据组包括丰富商家标识和丰富商家标识对应的丰富商家数据。

[0114] 分析模块1104,用于根据丰富商家数据,分析得到丰富商家数据包括的丰富商品的分布数据。

[0115] 构建模块1105,用于根据丰富商家标识和丰富商品的分布数据,构建目标商家数据组。

[0116] 进一步地,待处理商家数据包括待处理产品列表和待处理订单列表,商家订单数据表还包括商品标识和订单标识;

[0117] 比对模块1102具体用于:

[0118] 比对商家标识和待处理商家标识;当商家标识中出现与待处理商家标识不相同的待添加商家标识时,从商家订单数据表中获取待添加商家标识对应的待添加商品标识和待添加订单标识;根据待添加商品标识构建待添加产品列表;根据待添加订单标识构建待添加订单列表;根据待添加产品列表和待添加订单列表得到待添加商家标识对应的待添加商家数据;当商家标识中出现与待处理商家标识相同的待更新商家标识时,从商家订单数据表中获取待更新商家标识对应的待更新商品标识和待更新订单标识;将待更新商品标识加入待处理产品列表,并将待更新订单标识加入待处理订单列表得到更新后的待处理商家数据;根据待添加商家标识和待添加商家数据和/或待更新商家标识和更新后的待处理商家数据,更新待处理商家数据组。

[0119] 进一步地,服务器还包括商品数据表,商品数据表包括商品详细数据;

[0120] 比对模块1102进一步具体用于:

[0121] 根据待更新商品标识,从商品数据表中获取待更新商品标识对应的待更新商品标识;将待更新商品标识和待更新商品标识加入待处理产品列表。

[0122] 进一步地,预设条件为:

[0123] $\text{len}(P) \times \alpha \leq \text{len}(O)$ where $\alpha > 1$

[0124] 其中, $len(P)$ 为待处理产品列表的待处理商品数, $len(O)$ 为待处理订单的待处理订单数, α 为预设调节系数。

[0125] 进一步地, 丰富商家数据包括丰富商家产品列表;

[0126] 分析模块1104具体用于:

[0127] 根据丰富商家产品列表, 获取丰富商家产品列表中包括的丰富商品类别, 并获取丰富商品类别的数目; 将丰富商品类别和丰富商品类别的数目作为分布数据。

[0128] 进一步地, 获取模块1101具体还用于:

[0129] 获取丰富商家产品列表, 并根据丰富商家产品列表得到丰富商家产品列表包括的丰富商家产品数量; 判断丰富商家产品数量是否超过预设数量阈值; 若是, 则将丰富商家数据保留; 若否, 则将丰富商家数据删除。

[0130] 本申请实施例提供一种计算机设备, 计算机设备100包括处理器及存储有计算机指令的非易失性存储器, 计算机指令被处理器执行时, 计算机设备100执行前述的数据处理方法。如图9所示, 图9为本申请实施例提供的计算机设备100的结构框图。计算机设备100包括数据处理装置110、存储器111、处理器112及通信单元113。

[0131] 为实现数据的传输或交互, 存储器111、处理器112以及通信单元113各元件相互之间直接或间接地电性连接。例如, 可通过一条或多条通讯总线或信号线实现这些元件相互之间电性连接。数据处理装置110包括至少一个可以软件或固件 (firmware) 的形式存储于存储器111中或固化在计算机设备100的操作系统 (operating system, OS) 中的软件功能模块。处理器112用于执行存储器111中存储的可执行模块, 例如数据处理装置110所包括的软件功能模块及计算机程序等。

[0132] 本申请实施例提供一种可读存储介质, 可读存储介质包括计算机程序, 计算机程序运行时控制可读存储介质所在计算机设备执行前述的数据处理方法。

[0133] 综上所述, 本申请实施例提供了一种数据处理方法、装置、计算机设备和可读存储介质, 通过获取待处理商家数据组, 进而比对所述商家标识和所述待处理商家标识, 对所述待处理商家数据组进行更新; 然后根据预设条件对更新后的待处理商家数据组进行筛选处理, 得到丰富商家数据组, 再根据所述丰富商家数据, 分析得到所述丰富商家数据包括的丰富商品的分布数据; 最后根据所述丰富商家标识和所述丰富商品的分布数据, 构建目标商家数据组, 能够将线上商家的数据处理为可参考的目标商家数据组。

[0134] 以上所述, 仅为本申请的具体实施方式, 但本申请的保护范围并不局限于此, 任何熟悉本技术领域的技术人员在本申请揭露的技术范围内, 可轻易想到的变化或替换, 都应涵盖在本申请的保护范围之内。因此, 本申请的保护范围应以所述权利要求的保护范围为准。

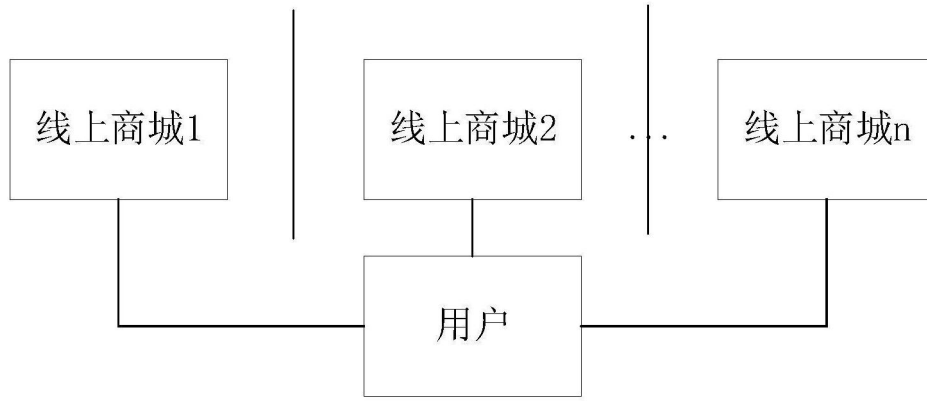


图1

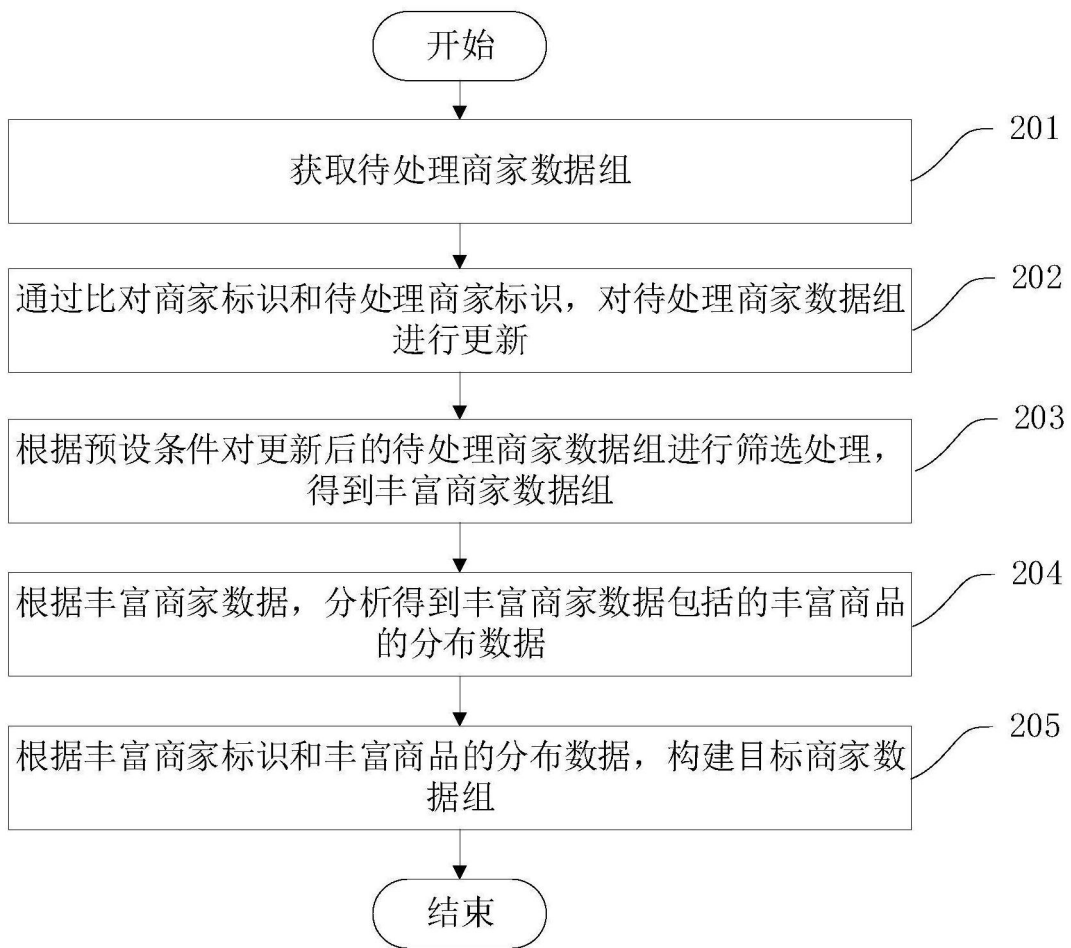


图2

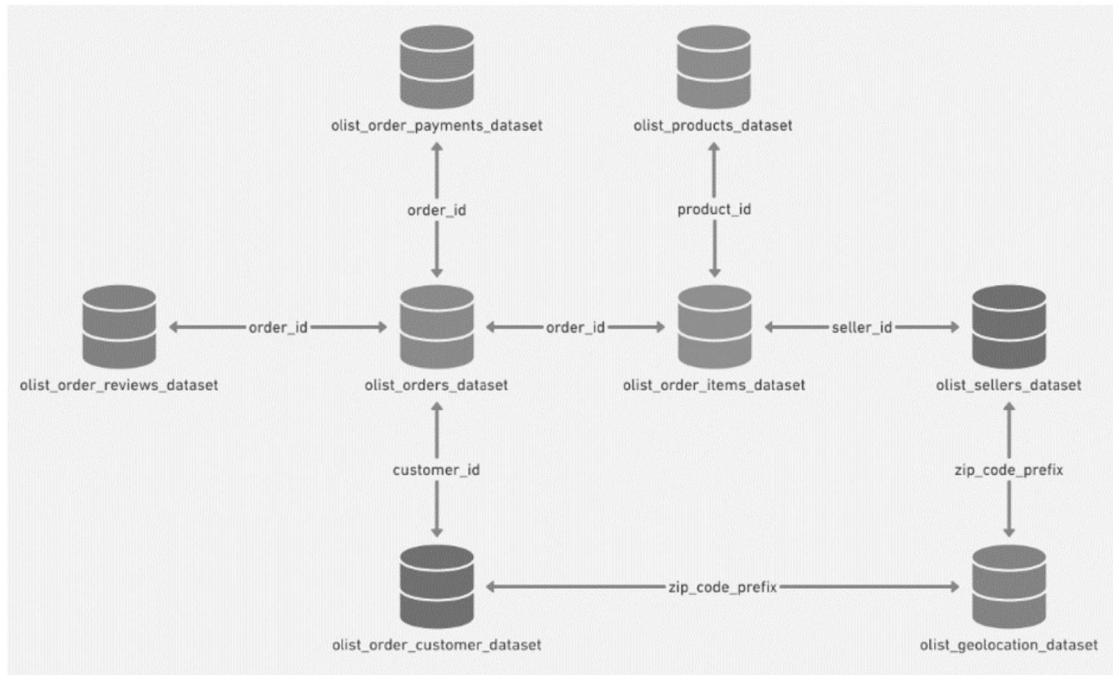


图3

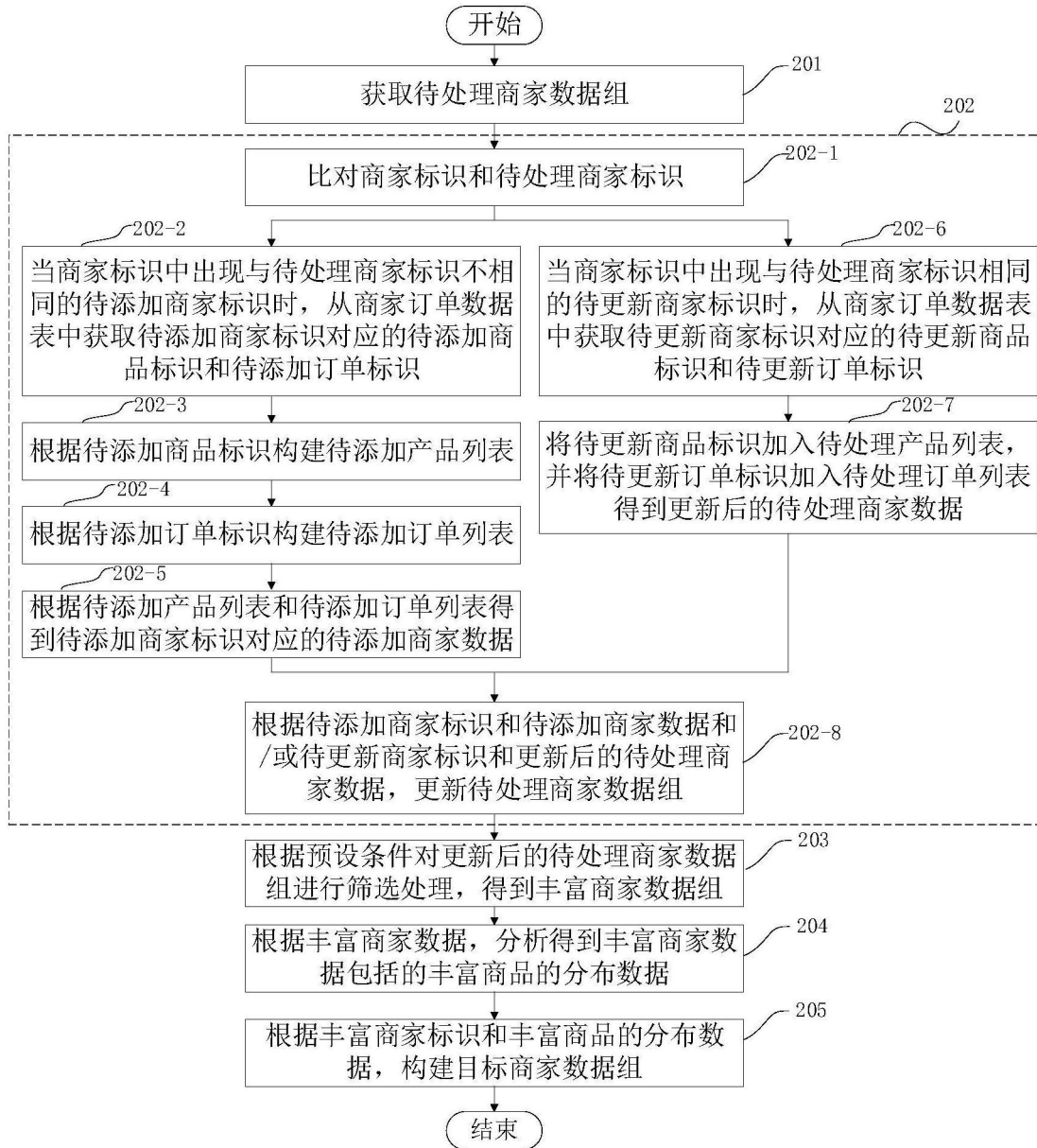


图4

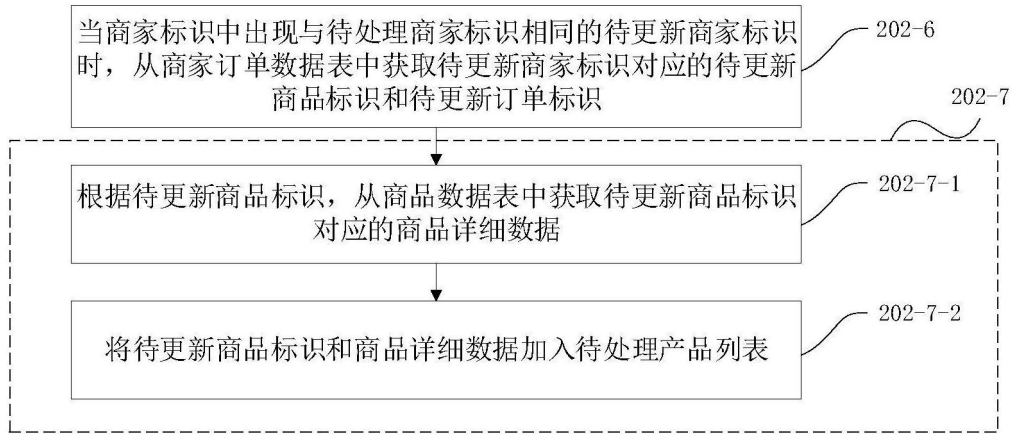


图5

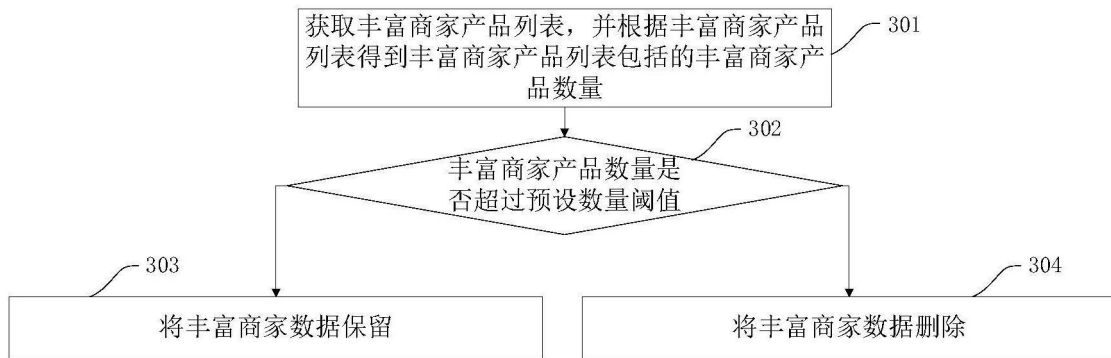


图6

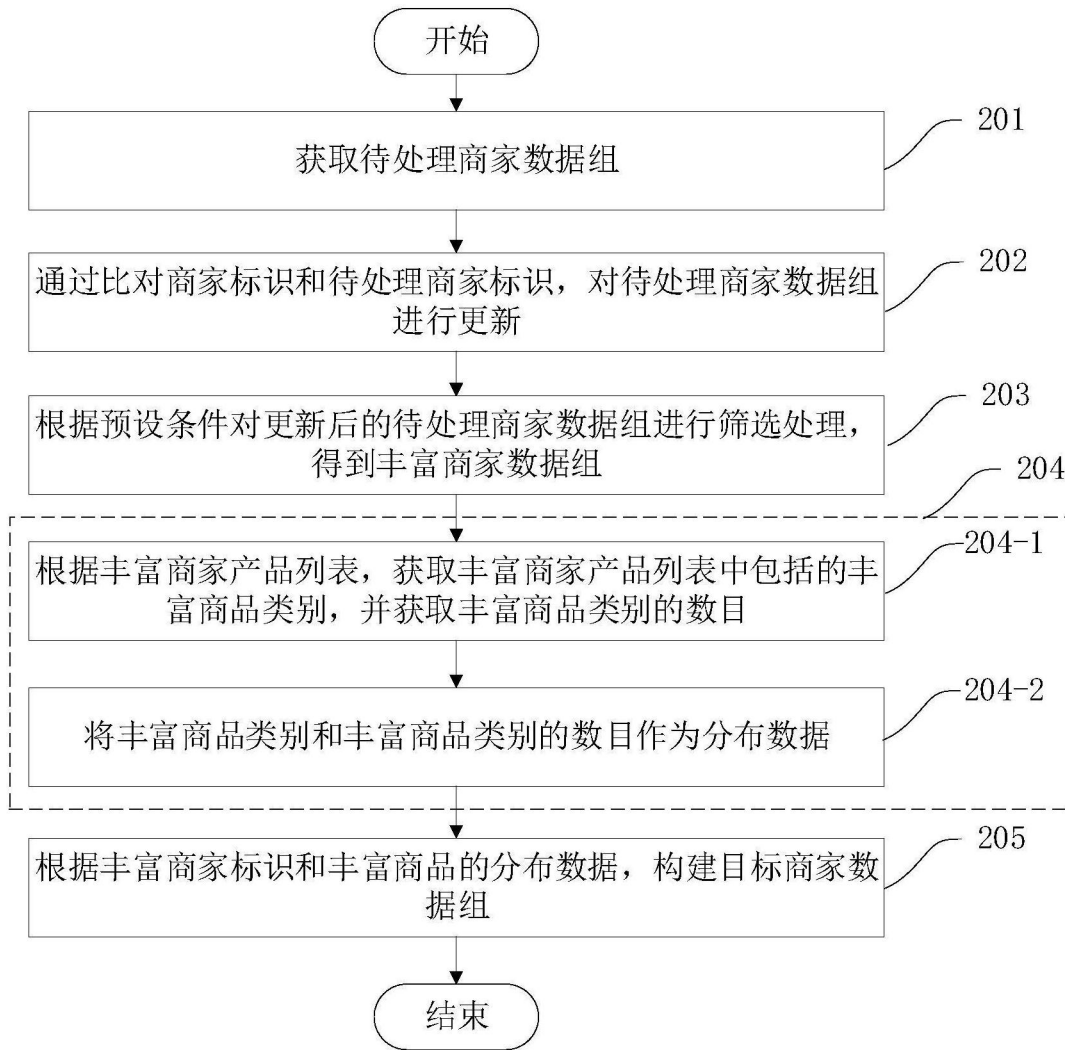


图7

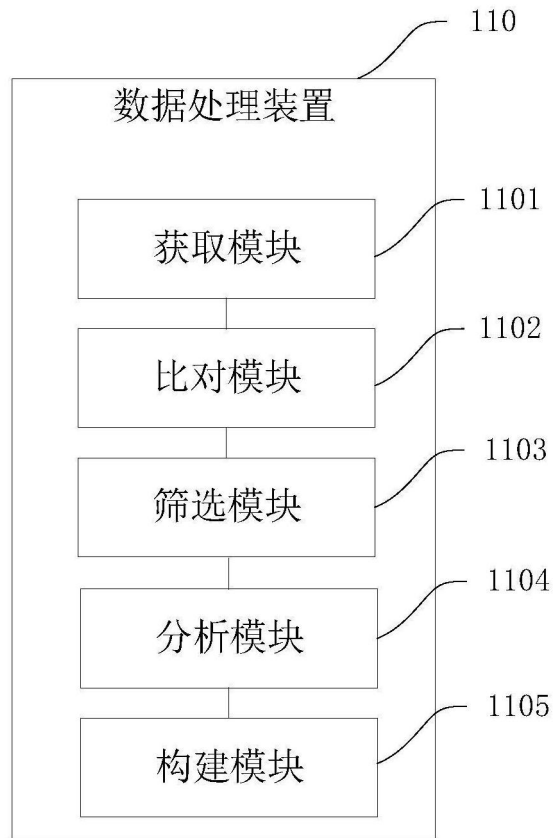


图8

100

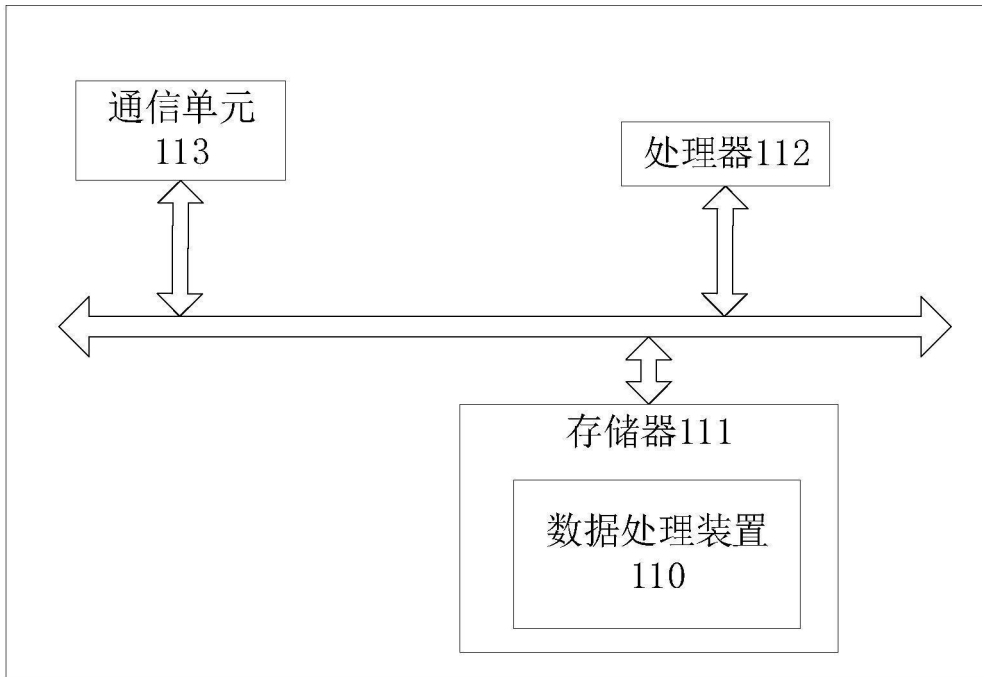


图9