



(12) 发明专利

(10) 授权公告号 CN 114840552 B

(45) 授权公告日 2025. 05. 02

(21) 申请号 202210452350.6

G06F 16/215 (2019.01)

(22) 申请日 2022.04.27

G06Q 30/0201 (2023.01)

(65) 同一申请的已公布的文献号  
申请公布号 CN 114840552 A

(56) 对比文件

CN 104899205 A, 2015.09.09

CN 110990400 A, 2020.04.10

(43) 申请公布日 2022.08.02

审查员 王琪

(73) 专利权人 杭州凡声优行科技有限公司  
地址 310000 浙江省杭州市拱墅区祥符街  
道余杭塘路478号8幢3层368室

(72) 发明人 章垚鹏

(74) 专利代理机构 杭州裕阳联合专利代理有限  
公司 33289  
专利代理师 张解翠

(51) Int. Cl.

G06F 16/2452 (2019.01)

G06F 16/2455 (2019.01)

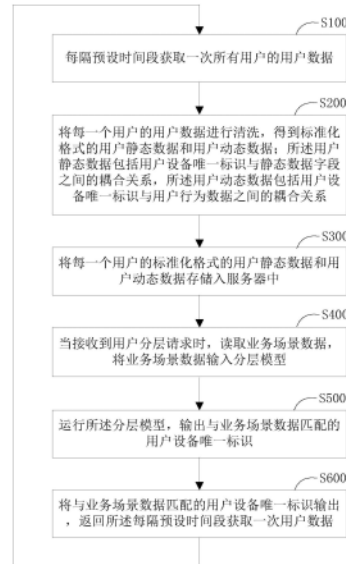
权利要求书3页 说明书12页 附图2页

(54) 发明名称

用户分层方法及系统

(57) 摘要

本申请涉及一种用户分层方法,业务人员只需要向业务终端输入业务场景数据,业务终端即可以自动进行用户分层操作,得到用户分层结果,不需要专业的数据处理人员介入,免除了与数据处理人员沟通的过程。此外,本方法将用户静态数据和用户动态数据进行了统一化、标准化处理,降低了由于数据底层处理逻辑不一致所带来的结果误差。



1. 一种用户分层方法,其特征在于,应用于业务终端,所述用户分层方法包括:  
每隔预设时间段获取一次所有用户的用户数据;  
将每一个用户的用户数据进行清洗,得到标准化格式的用户静态数据和用户动态数据,选取一个用户的用户数据,并读取用户数据中的用户设备唯一标识;  
提取用户数据中的至少一条用户行为数据;  
选取一条用户行为数据;  
将该条用户行为数据转化为行为事件ID和与行为事件ID关联的多个行为事件参数;  
返回所述选取一条用户行为数据,直至每一条用户行为数据均转化完毕;  
返回所述选取一个用户的用户数据,直至每一个用户的用户数据中的所有用户行为数据均转化完毕;

建立用户动态数据表,将与每一个用户设备唯一标识对应的行为事件ID、以及与行为事件ID关联的多个行为事件参数和用户设备唯一标识相对应的置入所述用户动态数据表中;所述用户静态数据包括用户设备唯一标识与静态数据字段之间的耦合关系,所述用户动态数据包括用户设备唯一标识与用户行为数据之间的耦合关系;

将每一个用户的标准化格式的用户静态数据和用户动态数据存储入服务器中;

当接收到用户分层请求时,读取业务场景数据,将业务场景数据输入分层模型;

运行所述分层模型,输出与业务场景数据匹配的用户设备唯一标识,运行分层模型,控制分层模型中的条件提取模块提取业务场景数据中的至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词配置项;

控制分层模型中的语法解析模块将至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词配置项合并并且转化为一条SQL查询语句;

将SQL查询语句与服务器中所有用户的标准化格式的用户静态数据和用户动态数据分别进行匹配,得到命中所述SQL查询语句的用户设备唯一标识;

将命中所述SQL查询语句的用户设备唯一标识作为与业务场景数据匹配的用户设备唯一标识;

将与业务场景数据匹配的用户设备唯一标识输出,返回所述每隔预设时间段获取一次用户数据。

2. 根据权利要求1所述的用户分层方法,其特征在于,所述每隔预设时间段向服务器获取一次所有用户的用户数据,包括:

每隔预设时间段向服务器获取所有用户的业务结果数据;所述业务结果数据包括用户商品购买记录、用户登记表填写记录和用户投诉记录中的一种或多种;

在本地的存储器中提取所有用户的用户行为数据;

通过第三方通信接口获取所有用户的第三方业务数据,所述第三方业务数据包括用户在第三方平台下的账号数据、用户在第三方平台下的客服沟通文字数据、以及用户在第三方平台下的客服沟通语音数据中的一种或多种。

3. 根据权利要求2所述的用户分层方法,其特征在于,所述将每一个用户的用户数据进行清洗,得到标准化格式的用户静态数据和用户动态数据,包括:

选取一个用户的用户数据,并读取用户数据中的用户设备唯一标识;  
将用户数据中的string格式字段取出,作为枚举型静态数据;  
将用户数据中的int格式字段取出,作为数值型静态数据;  
将用户数据中的timestamp字段取出,作为时间型静态数据;  
创建用户静态数据表,将与每一个用户设备唯一标识对应的枚举型静态数据、数值型静态数据和时间型静态数据,以及用户设备唯一标识相对应的置入所述用户静态数据表中;

返回所述选取一个用户的用户数据,并读取用户数据中的用户设备唯一标识,直至每一个用户的用户数据均清洗完毕。

4. 根据权利要求3所述的用户分层方法,其特征在于,所述行为事件参数包括枚举型行为事件参数、数值型行为事件参数和时间型行为事件参数中的一种或多种。

5. 根据权利要求4所述的用户分层方法,其特征在于,所述静态介词配置项包括大于、小于和等于中的一种或多种;所述动态介词配置项包括大于、小于和等于中的一种或多种;所述逻辑介词配置项包括且、或中的一种。

6. 根据权利要求5所述的用户分层方法,其特征在于,所控制分层模型中的语法解析模块将至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词配置项合并并且转化为SQL查询语句,包括:

将至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词组合拼接并生成JSON语句;

读取JSON语句中的标识符,依据JSON语句中的标识符拆分父条件和子条件;

读取JSON语句中的条件关键词,将JSON语句中的条件关键词转化为SQL关键字;

将所有SQL关键字组合拼接为SQL查询语句。

7. 根据权利要求6所述的用户分层方法,其特征在于,所述将SQL查询语句与服务器中所有用户的标准化格式的用户静态数据和用户动态数据分别进行匹配,得到命中所述SQL查询语句的用户设备唯一标识,包括:

从服务器中调取用户静态数据表,将SQL查询语句与所述用户静态数据表进行匹配,得到用户静态数据表中与SQL查询语句匹配的用户设备唯一标识,将用户静态数据表中与SQL查询语句匹配的用户设备唯一标识纳入第一唯一标识集合;

从服务器中调取用户动态数据表,将SQL查询语句与所述用户动态数据表进行匹配,得到用户动态数据表中与SQL查询语句匹配的用户设备唯一标识,将用户动态数据表中与SQL查询语句匹配的用户设备唯一标识纳入第二唯一标识集合;

取第一唯一标识集合和第二唯一标识集合的交集,得到第三唯一标识集合;

将第三唯一标识集合中的所有用户设备唯一标识作为命中所述SQL查询语句的用户设备唯一标识。

8. 一种用户分层系统,其特征在于,包括:

业务终端,用于执行如权利要求1-7中任意一项所述的用户分层方法;所述业务终端包括存储器;

服务器,与所述业务终端通信连接。

## 用户分层方法及系统

### 技术领域

[0001] 本申请涉及互联网数据处理技术领域,特别是涉及一种用户分层方法及系统。

### 背景技术

[0002] 用户分层方法是指在具体业务场景下对用户进行筛选,过滤出符合业务场景需求的目标用户群体的过程。在此过程中需要使用到用户数据,所需使用的用户数据主要包括用户的静态数据以及动态数据。用户的静态数据通常指用户本身的属性特征,如性别、年龄、工作性质等。用户的动态数据通常为用户的行为特征,如浏览网页、购买商品等,属于一种动态生成的数据形式。

[0003] 当业务人员依据业务场景提出用户分层请求时,传统的用户分层方法,数据处理人员一般的处理方式是通过人工查看的方式过滤出所需要的用户群体。

[0004] 然而,这种用户分层方法会造成用户分层结果不准确的问题。这是由于业务信息不对称导致的,业务人员和数据处理人员的工作性质不同,他们之间的沟通成本较高,在沟通过程中业务场景需求往往被曲解,导致用户分层结果与业务场景需求不匹配,不能满足业务场景需求。此外,用户的静态、动态数据分布在诸多数据库表中,不同数据处理人员的技术水平不同,他们对数据库表的理解存在差异,得到的用户分层结果不同,造成准确度参差不齐。

### 发明内容

[0005] 基于此,有必要针对传统用户分层方法是通过人工查看的方式过滤出所需要的用户群体,从而导致用户分层结果不准确的问题,提供一种用户分层方法及系统。

[0006] 本申请提供一种用户分层方法,应用于业务终端,所述用户分层方法包括:

[0007] 每隔预设时间段获取一次所有用户的用户数据;

[0008] 将每一个用户的用户数据进行清洗,得到标准化格式的用户静态数据和用户动态数据;所述用户静态数据包括用户设备唯一标识与静态数据字段之间的耦合关系,所述用户动态数据包括用户设备唯一标识与用户行为数据之间的耦合关系;

[0009] 将每一个用户的标准化格式的用户静态数据和用户动态数据存储入服务器中;

[0010] 当接收到用户分层请求时,读取业务场景数据,将业务场景数据输入分层模型;

[0011] 运行所述分层模型,输出与业务场景数据匹配的用户设备唯一标识;

[0012] 将与业务场景数据匹配的用户设备唯一标识输出,返回所述每隔预设时间段获取一次用户数据。

[0013] 进一步地,所述每隔预设时间段向服务器获取一次所有用户的用户数据,包括:

[0014] 每隔预设时间段向服务器获取所有用户的业务结果数据;所述业务结果数据包括用户商品购买记录、用户登记表填写记录和用户投诉记录中的一种或多种;

[0015] 在本地的存储器中提取所有用户的用户行为数据;

[0016] 通过第三方通信接口获取所有用户的第三方业务数据,所述第三方业务数据包括

用户在第三方平台下的账号数据、用户在第三方平台下的客服沟通文字数据、以及用户在第三方平台下的客服沟通语音数据中的一种或多种。

[0017] 进一步地,所述将每一个用户的用户数据进行清洗,得到标准化格式的用户静态数据和用户动态数据,包括:

[0018] 选取一个用户的用户数据,并读取用户数据中的用户设备唯一标识;

[0019] 将用户数据中的string格式字段取出,作为枚举型静态数据;

[0020] 将用户数据中的int格式字段取出,作为数值型静态数据;

[0021] 将用户数据中的timestamp字段取出,作为时间型静态数据;

[0022] 创建用户静态数据表,将与每一个用户设备唯一标识对应的枚举型静态数据、数值型静态数据和时间型静态数据,以及用户设备唯一标识相对应的置入所述用户静态数据表中;

[0023] 返回所述选取一个用户的用户数据,并读取用户数据中的用户设备唯一标识,直至每一个用户的用户数据均清洗完毕。

[0024] 进一步地,在返回所述选取一个用户的用户数据,并读取用户数据中的用户设备唯一标识,直至每一个用户的用户数据均清洗完毕之前,所述将每一个用户的用户数据进行清洗,得到标准化格式的用户静态数据和用户动态数据,还包括:

[0025] 选取一个用户的用户数据,并读取用户数据中的用户设备唯一标识;

[0026] 提取用户数据中的至少一条用户行为数据;

[0027] 选取一条用户行为数据;

[0028] 将该条用户行为数据转化为行为事件ID和与行为事件ID关联的多个行为事件参数;

[0029] 返回所述选取一条用户行为数据,直至每一条用户行为数据均转化完毕;

[0030] 返回所述选取一个用户的用户数据,直至每一个用户的用户数据中的所有用户行为数据均转化完毕;

[0031] 建立用户动态数据表,将与每一个用户设备唯一标识对应的行为事件ID、以及与行为事件ID关联的多个行为事件参数和用户设备唯一标识相对应的置入所述用户动态数据表中。

[0032] 进一步地,所述行为事件参数包括枚举型行为事件参数、数值型行为事件参数和时间型行为事件参数中的一种或多种。

[0033] 进一步地,所述运行分层模型,输出与业务场景数据匹配的用户设备唯一标识,包括:

[0034] 运行分层模型,控制分层模型中的条件提取模块提取业务场景数据中的至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词配置项;

[0035] 控制分层模型中的语法解析模块将至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词配置项合并且转化为一条SQL查询语句;

[0036] 将SQL查询语句与服务器中所有用户的标准化格式的用户静态数据和用户动态数据分别进行匹配,得到命中所述SQL查询语句的用户设备唯一标识;

[0037] 将命中所述SQL查询语句的用户设备唯一标识作为与业务场景数据匹配的用户设备唯一标识。

[0038] 进一步地,所述静态介词配置项包括大于、小于和等于中的一种或多种;所述动态介词配置项包括大于、小于和等于中的一种或多种;所述逻辑介词配置项包括且、或中的一种。

[0039] 进一步地,所控制分层模型中的语法解析模块将至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词配置项合并且转化为SQL查询语句,包括:

[0040] 将至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词组合拼接并生成JSON语句;

[0041] 读取JSON语句中的标识符,依据JSON语句中的标识符拆分父条件和子条件;

[0042] 读取JSON语句中的条件关键词,将JSON语句中的条件关键词转化为SQL关键字;

[0043] 将所有SQL关键字组合拼接为SQL查询语句。

[0044] 进一步地,所述将SQL查询语句与服务器中所有用户的标准化格式的用户静态数据和用户动态数据分别进行匹配,得到命中所述SQL查询语句的用户设备唯一标识,包括:

[0045] 从服务器中调取用户静态数据表,将SQL查询语句与所述用户静态数据表进行匹配,得到用户静态数据表中与SQL查询语句匹配的用户设备唯一标识,将用户静态数据表中与SQL查询语句匹配的用户设备唯一标识纳入第一唯一标识集合;

[0046] 从服务器中调取用户动态数据表,将SQL查询语句与所述用户动态数据表进行匹配,得到用户动态数据表中与SQL查询语句匹配的用户设备唯一标识,将用户动态数据表中与SQL查询语句匹配的用户设备唯一标识纳入第二唯一标识集合;

[0047] 取第一唯一标识集合和第二唯一标识集合的交集,得到第三唯一标识集合;

[0048] 将第三唯一标识集合中的所有用户设备唯一标识作为命中所述SQL查询语句的用户设备唯一标识。

[0049] 本申请提供一种用户分层系统,包括:

[0050] 业务终端,用于执行如前述内容提及的用户分层方法;所述业务终端包括存储器;

[0051] 服务器,与所述业务终端通信连接。

[0052] 本申请涉及一种用户分层方法,业务人员只需要向业务终端输入业务场景数据,业务终端即可以自动进行用户分层操作,得到用户分层结果,不需要专业的数据处理人员介入,免除了与数据处理人员沟通的过程。此外,本方法将用户静态数据和用户动态数据进行了统一化、标准化处理,降低了由于数据底层处理逻辑不一致所带来的结果误差。

## 附图说明

[0053] 图1为本申请一实施例提供的用户分层方法的流程示意图。

[0054] 图2为本申请一实施例提供的用户分层系统的结构示意图。

[0055] 附图标记:

[0056] 100-业务终端;110-存储器;200-服务器。

## 具体实施方式

[0057] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0058] 本申请提供一种用户分层方法及系统。

[0059] 一方面,本申请提供一种用户分层方法。需要说明的是,本申请提供的用户分层方法应用于业务终端。即本申请提供的用户分层方法不需要专业的数据处理人员进行操作,只需要业务人员操作即可。

[0060] 此外,本申请提供的用户分层方法不限制其执行主体。可选地,本申请提供的用户分层方法的执行主体可以为一种业务终端。具体地,本申请提供的用户分层方法的执行主体可以为所述业务终端中的一个或多个处理器。

[0061] 如图1所示,在本申请的一实施例中,所述用户分层方法应用于业务终端。

[0062] 所述用户分层方法包括:

[0063] S100,每隔预设时间段获取一次所有用户的用户数据。

[0064] 具体地,用户在用户终端执行注册账号,在应用程序上浏览信息,购物下单时这几种操作是,都会产生用户数据,服务器会及时采集这些用户数据存储于服务器中。可选地,有些用户数存储于本地存储器中,有些用户数据存储于服务器中。业务终端会定时从服务器和本地存储器中一次性的获取所有用户的用户数据。

[0065] 可选地,预设时间段可以为12小时。可选地,预设时间段可以为24小时。

[0066] S200,将每一个用户的用户数据进行清洗,得到标准化格式的用户静态数据和用户动态数据。所述用户静态数据包括用户设备唯一标识与静态数据字段之间的耦合关系。所述用户动态数据包括用户设备唯一标识与用户行为数据之间的耦合关系。

[0067] 具体地,用户数据的来源不同,格式也不同,本步骤可以对不同来源,不同格式的用户数据进行整合,按照预先设定的格式统一对用户数据进行清洗,主要目的是形成标准化的用户数据格式。

[0068] 此外,本步骤还可以将用户静态数据和用户动态数据区分开来,便于后续处理。

[0069] S300,将每一个用户的标准化格式的用户静态数据和用户动态数据存储入服务器中。

[0070] 具体地,服务器可以设立两个不同的存储区域分别存储用户静态数据和用户动态数据。

[0071] S400,当接收到用户分层请求时,读取业务场景数据,将业务场景数据输入分层模型。

[0072] 具体地,S100至S300为定时清洗数据的流程,S400至S600为具体的分层流程。

[0073] 可选地,当业务终端接收到用户分层请求时,业务终端会抓取用户分层请求中附带的业务场景数据,将业务场景数据输入至业务终端中的分层模型。

[0074] S500,运行所述分层模型,输出与业务场景数据匹配的用户设备唯一标识。

[0075] 具体地,分层模型是一个已训练过的深度学习模型,分层模型可以自动根据业务场景数据去自动筛选与业务场景数据匹配的用户设备唯一标识输出,而不需要专业数据处理人员的介入。

[0076] S600,将与业务场景数据匹配的用户设备唯一标识输出,返回所述S100。

[0077] 具体地,本方法的最终目标是筛选出满足业务场景需求的用户设备唯一标识。

[0078] 本实施例中,业务人员只需要向业务终端输入业务场景数据,业务终端即可以自动进行用户分层操作,得到用户分层结果,不需要专业的数据处理人员介入,免除了与数据处理人员沟通的过程。此外,本方法将用户静态数据和用户动态数据进行了统一化、标准化处理,降低了由于数据底层处理逻辑不一致所带来的结果误差。

[0079] 在本申请的一实施例中,所述S100包括:

[0080] S110,每隔预设时间段向服务器获取所有用户的业务结果数据。所述业务结果数据包括用户商品购买记录、用户登记表单填写记录和用户投诉记录中的一种或多种。

[0081] S120,在本地的存储器中提取所有用户的用户行为数据。

[0082] S130,通过第三方通信接口获取所有用户的第三方业务数据,所述第三方业务数据包括用户在第三方平台下的账号数据、用户在第三方平台下的客服沟通文字数据、以及用户在第三方平台下的客服沟通语音数据中的一种或多种。

[0083] 具体地,用户数据包括三种类型的数据:业务结果数据、用户行为数据和第三方业务数据。所述业务结果数据包括用户商品购买记录、用户登记表单填写记录和用户投诉记录中的一种或多种。

[0084] 用户行为数据一般是埋点数据,由业务终端采集,用户行为数据记录了用户打开页面、浏览页面等行为,例如:行为发生时间(2022-03-07 10:00:00),网页名称(XX网页),用户设备ID(可以是IMEI码,例如123456),停留时长(5秒)等。

[0085] 第三方业务数据是第三方平台通过第三方通信接口提供的数据。第三方平台指的是除了用户终端和业务终端之外的平台,例如A企业开发的聊天软件平台,与业务终端签署过信息交互协议,那么用户在A企业开发的聊天软件平台下注册账号的账号数据,用户在A企业开发的聊天软件平台下与客户沟通的文字数据和语音数据都会通过第三方通信接口被和业务终端抓取到。

[0086] 在本申请的一实施例中,所述S200包括:

[0087] S211,选取一个用户的用户数据,并读取用户数据中的用户设备唯一标识。

[0088] S212,将用户数据中的string格式字段取出,作为枚举型静态数据。

[0089] S213,将用户数据中的int格式字段取出,作为数值型静态数据。

[0090] S214,将用户数据中的timestamp字段取出,作为时间型静态数据。

[0091] S215,创建用户静态数据表,将与每一个用户设备唯一标识对应的枚举型静态数据、数值型静态数据和时间型静态数据,以及用户设备唯一标识相对应的置入所述用户静态数据表中。

[0092] S230,返回所述S211,直至每一个用户的用户数据均清洗完毕。

[0093] 具体地,本实施例主要介绍用户静态数据的格式统一化。用户数据中的string格式字段我们定义为枚举型静态数据,此类数据表明的是用户的个人特征,很少发生变化。例如:性别为男、归属地为北京、职业类型为公务员。

[0094] 用户数据中的int格式字段我们定义为数值型静态数据。此类数据表明的是用户的一些可记数的特征,会随时间发生变化。例如:年龄为26,累计消费金额为1000。

[0095] 用户数据中的timestamp字段我们定义为时间型静态数据。此类数据表明的是用

户在某一时间点的特征。例如生日为1979-02-04,最近一次消费时间为2020-06-07。

[0096] 为了使得最终形成的标准化格式的用户静态数据包含用户设备(用户)和用户静态数据的耦合关系,在S215中,我们建立了用户静态数据表,将用户静态数据表作为标准化格式的用户静态数据。

[0097] 表1-用户静态数据表(示例性的)

[0098]	用户唯一标识	性别	年龄	最近一次消费时间	累计消费金额
	10001	男	26	-	0
	10002	女	-	2020-07-08	100
	10003	男	33	2019-06-05	500

[0099] 表1即用户静态数据表的一种实施方式,每个用户拥有一个用户唯一标识。用户唯一标识可以由业务终端进行编号,也可以采用原生的用户终端的设备唯一标识IMEI码。性别是枚举型静态数据。最近一次消费时间是时间型静态数据。年龄,累计消费金额是数值型静态数据。

[0100] 每行中记录的结果是打在用户身上的静态标签。值得一提的是,由于一个用户的用户静态数据仅存在一行结果,所以用户静态数据也具备唯一性,无法做到静态数据变化过程的记录,所以枚举型静态数据只能记录最新的结果,时间型枚举型静态数据只能记录首次、最近一次的结果,数值型枚举型静态数据只能记录计数、加和结果,这在应用过程中具有一定的局限性,所以需要用户动态数据进行补充,以弥补这点不足。

[0101] 在本申请的一实施例中,在S230之前,所述S200还包括:

[0102] S221,选取一个用户的用户数据,并读取用户数据中的用户设备唯一标识。

[0103] 具体地,每个用户拥有一个用户唯一标识。用户唯一标识可以由业务终端进行编号,也可以采用原生的用户终端的设备唯一标识IMEI码。

[0104] S222,提取用户数据中的至少一条用户行为数据。

[0105] S223,选取一条用户行为数据。

[0106] S224,将该条用户行为数据转化为行为事件ID和与行为事件ID关联的多个行为事件参数。

[0107] S225,返回所述S223,直至每一条用户行为数据均转化完毕。

[0108] 具体地,反复执行S223至S224,直至每一条用户行为数据均转化完毕。

[0109] S226,返回所述S221,直至每一个用户的用户数据中的所有用户行为数据均转化完毕。

[0110] 具体地,反复执行S222至S225,,直至每一个用户的用户数据中的所有用户行为数据均转化完毕。

[0111] S227,建立用户动态数据表,将与每一个用户设备唯一标识对应的行为事件ID、以及与行为事件ID关联的多个行为事件参数和用户设备唯一标识相对应的置入所述用户动态数据表中。

[0112] 具体地,为了使得最终形成的标准化格式的用户动态数据包含用户设备(用户)和

用户动态数据的耦合关系,我们建立了用户动态数据表,将用户动态数据表作为标准化格式的用户动态数据。

[0113] 表2-用户动态数据表(示例性的)

用户唯一标识	行为事件ID	枚举型行为事件参数	时间型行为事件参数	数值型行为事件参数
10001	K001(购买商品)	400	2020-12-30	2
[0114] 10001	K002(浏览页面)	-	2020-12-29	Page_1
10002	K002(浏览页面)	-	2020-12-25	Page_2
10003	K001(购买商品)	500	2020-12-30	1

[0115] 表2即用户动态数据表的一种实施方式。

[0116] 如表2所示,表格中的每一行为一个用户的一条用户行为数据。行为事件ID有K001和K002,K001代表购买商品这个行为事件,K002代表浏览页面这个行为事件。

[0117] 在本申请的一实施例中,所述行为事件参数包括枚举型行为事件参数、数值型行为事件参数和时间型行为事件参数中的一种或多种。

[0118] 具体地,用户行为数据也有不同类型,通过不同类型的行为事件参数表示。行为事件参数是对行为事件进一步的具体描述。

[0119] 如表2所示,表2中的枚举型行为事件参数代表的是购买商品的价格。表2中的时间型行为事件参数代表的是用户行为事件发生的时间。表2中的数值型行为事件参数在K001这个行为事件中代表的是购买商品的数量,在K002这个行为事件中代表的是用户浏览页面的ID(即Page\_1和Page\_2)。

[0120] 由于浏览页面并不会产生购买商品的价格,因此其枚举型行为事件参数用“-”表示。

[0121] 在本申请的一实施例中,S500包括:

[0122] S510,运行分层模型,控制分层模型中的条件提取模块提取业务场景数据中的至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词配置项。

[0123] S520,控制分层模型中的语法解析模块将至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词配置项合并并且转化为一条SQL查询语句。

[0124] S530,将SQL查询语句与服务器中所有用户的标准化格式的用户静态数据和用户动态数据分别进行匹配,得到命中所述SQL查询语句的用户设备唯一标识。

[0125] S540,将命中所述SQL查询语句的用户设备唯一标识作为与业务场景数据匹配的用户设备唯一标识。

[0126] 具体地,业务场景数据中包括业务场景需求信息,业务终端可以从中提取至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词配置项。

[0127] 例如业务场景数据中包括的业务场景需求信息为:在2021年12月20日至2022年12月30日之间购买商品大于1件,购买商品的价格大于100的女性。

[0128] 那么条件提取模块可以从业务场景需求信息中提取的静态筛选条件字段只有一个,为性别。静态介词配置项只有一个,为等于。静态内容字段只有一个,为女。动态筛选条件字段为3个,一个是2021年12月20日至2022年12月30日之间,一个是购买商品,还有一个是商品价格。动态介词配置项有两个,两个都是大于。动态内容字段有两个,一个为1,另一个为100。逻辑介词配置项为且。

[0129] 语法解析模块可以在不破坏业务场景需求信息原意的情况下将至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词配置项进行合并,转化为一条SQL查询语句。

[0130] 可选地,语法解析模块在S520中执行合并时,先合并至少一个静态筛选条件字段、至少一个静态介词配置项和至少一个静态内容字段,得到静态合并结果,承接上述例子,静态合并结果为“性别等于女”。进一步语法解析模块合并至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段,得到动态合并结果,承接上述例子,动态合并结果为“购买商品大于1,商品价格大于100”。语法解析模块最终将静态合并结果和动态合并结果通过逻辑介词配置项串联,得到合并最终结果,然后将合并最终结果转化为SQL查询语句。

[0131] 在本申请的一实施例中,所述静态介词配置项包括大于、小于和等于中的一种或多种;所述动态介词配置项包括大于、小于和等于中的一种或多种;所述逻辑介词配置项包括且、或中的一种。

[0132] 具体地,可以理解,静态筛选条件字段也有枚举型,时间型和数值型,这和标准化格式的用户静态数据中的枚举型静态数据,数值型静态数据和时间型静态数据是完全能够对应的上的,动态筛选条件字段也有枚举型,时间型和数值型,这和标准化格式的用户动态数据中的枚举型行为事件参数,数值型行为事件参数和时间型行为事件参数是完全能够对应的上的,这也是S530能够匹配的原因所在。

[0133] 在本申请的一实施例中,所述S520包括:

[0134] S521,将至少一个静态筛选条件字段、至少一个静态介词配置项、至少一个静态内容字段、至少一个动态筛选条件字段、至少一个动态介词配置项、至少一个动态内容字段和逻辑介词组合拼接并生成JSON语句。

[0135] 例如,承接业务场景数据中包括的业务场景需求信息为:在2021年12月20日至2022年12月30日之间购买商品大于1件,购买商品的价格大于100的女性的这个例子,最终生成的JSON语句如下:

```
{
  "groupConditionJson": {
    "child": [{
      "labelId": "T01010101",
      "labelName": "性别",
      "fieldName": "gender",
      "labelType": 1,
      "value": "女"
    }]
  },
  "doConditionJson": {
    "child": [{
      "eventNo": "EN1120895879",
      "eventName": "购买商品",
```

[0136]

```

    "eventId": "buy",
    "value": "1",
    "child": [{
        "paramName": "商品价格",
        "paramId": "price",
        "paramType": 2,
        "value": "100"
    [0137]    }],
    "startTime": "2021-12-20",
    "endTime": "2021-12-30"
  ]
},
"logic": 1
}

```

[0138] S522, 读取JSON语句中的标识符, 依据JSON语句中的标识符拆分父条件和子条件。

[0139] 具体地, 上述JSON语句中有一些标识符, 例如“child”为父条件标识符, 代表从属于上一层级的父条件。本步骤依据JSON语句中的标识符拆分父条件和子条件。

[0140] S523, 读取JSON语句中的条件关键词, 将JSON语句中的条件关键词转化为SQL关键字。

[0141] 具体地, 例如性别等于女, 性别对应的SQL关键字为gender, 则转换的SQL语句即为gender='女'。

[0142] S524, 将所有SQL关键字组合拼接为SQL查询语句。

[0143] 具体地, 拼接后的SQL查询语句为:

[0144] WITH groupConditionJson AS

[0145] (

[0146] SELECT user\_id

[0147] FROM label\_table

[0148] WHERE gender='女'

[0149] ),

```
[0150] doConditionJson AS
[0151] (
[0152] SELECT user_id
[0153] ,count(*)
[0154] FROM event_table
[0155] WHERE price>100
[0156] AND      evnt_time BETWEEN'2021-12-20'
[0157] AND      '2021-12-30'
[0158] GROUP BY user_id
[0159] HAVING count(*)>1
[0160] )
[0161] SELECT user_id
[0162] FROM      groupConditionJson
[0163] JOIN      doConditionJson
[0164] ON        groupConditionJson.user_id=doConditionJson.user_id。
```

[0165] 在本申请的一实施例中,所述S530包括:

[0166] S531,从服务器中调取用户静态数据表,将SQL查询语句与所述用户静态数据表进行匹配,得到用户静态数据表中与SQL查询语句匹配的用户设备唯一标识,将用户静态数据表中与SQL查询语句匹配的用户设备唯一标识纳入第一唯一标识集合。

[0167] S532,从服务器中调取用户动态数据表,将SQL查询语句与所述用户动态数据表进行匹配,得到用户动态数据表中与SQL查询语句匹配的用户设备唯一标识,将用户动态数据表中与SQL查询语句匹配的用户设备唯一标识纳入第二唯一标识集合。

[0168] S533,取第一唯一标识集合和第二唯一标识集合的交集,得到第三唯一标识集合。

[0169] S534,将第三唯一标识集合中的所有用户设备唯一标识作为命中所述SQL查询语句的用户设备唯一标识。

[0170] 具体地,本步骤是查表过程,注意的是先分别查询用户静态数据表 and 用户动态数据表,分别获取查询结果,最后将查询结果取交集。最终将交集集中的用户设备唯一标识作为命中所述SQL查询语句的用户设备唯一标识,待业务人员后续使用。换言之,用户静态数据表中与SQL查询语句匹配的用户设备唯一标识可能为多个,也可以为一个。用户动态数据表中与SQL查询语句匹配的用户设备唯一标识可能为多个,也可以为一个。

[0171] 另一方面,本申请还提供一种用户分层系统。

[0172] 如图2所示,在本申请的一实施例中,所述用户分层系统包括业务终端100和服务器200。

[0173] 所述业务终端100用于执行如前述内容提及的用户分层方法。所述业务终端100包括存储器110。所述服务器200与所述业务终端100通信连接。

[0174] 具体地,所述业务终端100搭载有分层模型。分层模型是预先训练好的深度学习模型。

[0175] 需要说明的是,为了行文简洁,业务终端100,存储器110和服务器200只在本实施例进行标号,前述的用户分层方法的各个实施例中,业务终端100,存储器110和服务器200

均不进行标号。

[0176] 以上所述实施例的各技术特征可以进行任意的组合,各方法步骤也并不做执行顺序的限制,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0177] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对本申请专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请的保护范围应以所附权利要求为准。

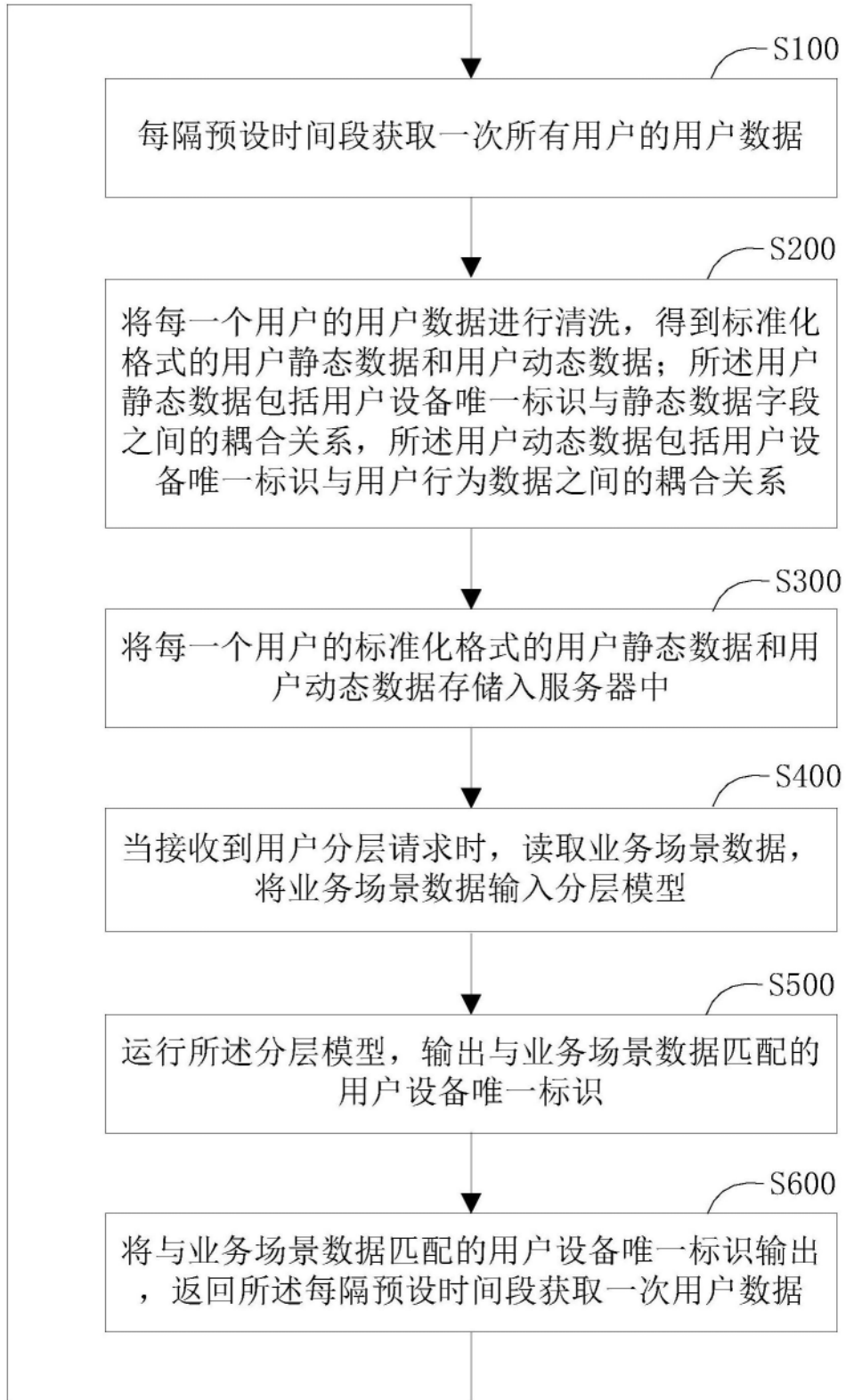


图1

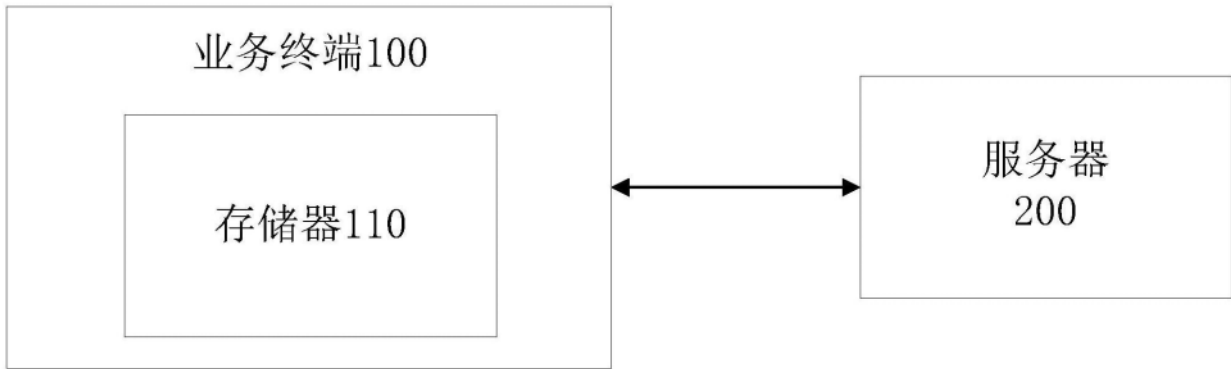


图2