



(12)发明专利

(10)授权公告号 CN 111242723 B

(45)授权公告日 2020.09.15

(21)申请号 202010008820.0

(22)申请日 2020.01.02

(65)同一申请的已公布的文献号  
申请公布号 CN 111242723 A

(43)申请公布日 2020.06.05

(73)专利权人 平安科技(深圳)有限公司  
地址 518000 广东省深圳市福田区福田街  
道福安社区益田路5033号平安金融中  
心23楼

专利权人 中国平安财产保险股份有限公司

(72)发明人 喻宁 史良洵 陈克炎 朱园丽  
朱艳乔 陈皓云

(74)专利代理机构 深圳市沃德知识产权代理事  
务所(普通合伙) 44347  
代理人 高杰 于志光

(51)Int.Cl.

G06Q 30/06(2012.01)

G06Q 50/00(2012.01)

H04W 4/029(2018.01)

(56)对比文件

CN 109189935 A,2019.01.11

CN 102047697 A,2011.05.04

CN 110598769 A,2019.12.20

CN 110442746 A,2019.11.12

CN 105007171 A,2015.10.28

CN 109359812 A,2019.02.19

CN 106570722 A,2017.04.19

CN 110611689 A,2019.12.24

CN 110490632 A,2019.11.22

审查员 刘畅

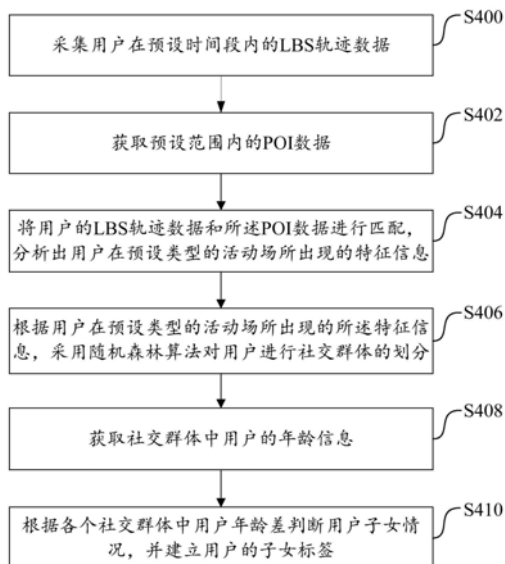
权利要求书1页 说明书11页 附图5页

(54)发明名称

用户子女情况判断方法、服务器及计算机可  
读存储介质

(57)摘要

本发明涉及一种数据分析技术,揭露了一种  
用户子女情况判断方法,该方法包括:采集用户  
在预设时间段内的LBS轨迹数据;获取预设范围  
内的POI数据;将用户的LBS轨迹数据和POI数据  
进行匹配,分析出用户在预设类型的活动场所出  
现的特征信息;根据所述特征信息,采用随机森  
林算法对用户进行社交群体的划分;获取所述社  
交群体中用户的年龄信息;根据各个社交群体中  
用户的年龄差判断用户子女情况,并建立用户的  
子女标签。本发明还提供一种服务器及计算机可  
读存储介质。本发明能够通过LBS轨迹数据对目  
标人群进行智能化信息采集,减轻人力劳动及节  
约成本,并根据所采集的数据和用户的年龄信息  
判断用户是否有子女,从而更容易地对用户进行  
有针对性的消费推荐。



1. 一种用户子女情况判断方法,其特征在于,所述方法包括步骤:  
采集用户在预设时间段内的基于位置的服务LBS轨迹数据;  
获取预设范围内的兴趣点POI数据;  
将用户的LBS轨迹数据和所述POI数据进行匹配,分析出用户在预设类型的活动场所出现的特征信息;  
根据用户在预设类型的活动场所出现的所述特征信息,采用随机森林算法对用户进行社交群体的划分,包括:根据每个用户的所述特征信息,针对用户在所述活动场所出现的时间、位置和出现次数作为分区特征,将同时出现在一个活动场所且出现次数不同的不同用户划分到同一社交群体;  
获取所述社交群体中用户的年龄信息;及  
根据各个所述社交群体中用户的年龄差判断用户子女情况,并建立用户的子女标签;  
所述根据各个社交群体中用户年龄差判断用户子女情况的步骤包括:根据所获取的用户的年龄信息计算各个社交群体中每两个用户之间的年龄差,若所述社交群体中有两个用户的年龄差大于18岁,则判断所述两个用户之间存在亲子关系,表示所述两个用户中年长用户有子女。
2. 如权利要求1所述的用户子女情况判断方法,其特征在于,该方法在采集到所述用户的LBS轨迹数据后还包括步骤:  
使用数据清洗技术对所采集的用户的LBS轨迹数据进行预处理,包括检查数据一致性,处理无效值和缺失值。
3. 如权利要求1或2所述的用户子女情况判断方法,其特征在于,该方法在分析出用户在预设类型的活动场所出现的特征信息之后还包括步骤:  
基于Fisher准则和 $\chi^2$ 检验的筛选变量算法,剔除与用户子女情况判断无关的特征。
4. 如权利要求1所述的用户子女情况判断方法,其特征在于,所述POI数据和所述预设类型的活动场所主要包括亲子类活动场所。
5. 如权利要求1所述的用户子女情况判断方法,其特征在于,所述特征信息包括用户的设备号、出现的每个所述预设类型的活动场所、在各活动场所出现的次数、出现的时间点、花费的时长。
6. 如权利要求5所述的用户子女情况判断方法,其特征在于,所述根据各个所述社交群体中用户的年龄差判断用户子女情况的步骤还包括:  
获取大数据中用户的性别信息,根据所述性别信息进一步确认所述年长用户的子女性别。
7. 一种服务器,其特征在于,所述服务器包括存储器、处理器,所述存储器上存储有可在所述处理器上运行的用户子女情况判断系统,所述用户子女情况判断系统被所述处理器执行时实现如权利要求1-6中任一项所述的用户子女情况判断方法的步骤。
8. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有用户子女情况判断系统,所述用户子女情况判断系统可被至少一个处理器执行,以使所述至少一个处理器执行如权利要求1-6中任一项所述的用户子女情况判断方法的步骤。

## 用户子女情况判断方法、服务器及计算机可读存储介质

### 技术领域

[0001] 本发明涉及数据分析技术领域,尤其涉及一种用户子女情况判断方法、服务器及计算机可读存储介质。

### 背景技术

[0002] 目前,业内对于如何确定用户是否有子女的情况,都是通过单独的个人信息采集,如问卷调查、保险等级等方式来询问用户是否有子女。这种方法非常复杂,需要经过被采集人的同意且愿意去填写子女情况,采集到的信息不全面,且真实性不高。并且,在采集后需要有专门的负责人对采集的数据进行整理和归档。在这个过程中,存在着简单且机械重复的人力劳动,效率非常低,且可能出现人为失误,容易造成人力和时间的浪费。以上情况将影响最后判断结果的准确性和及时性,导致对用户进行错误的消费推荐。

[0003] 因此,如何在克服以上不足的情况下,提供一种有效的判断用户是否有子女的处理方案,已经成为一个亟待解决的技术问题。

### 发明内容

[0004] 有鉴于此,本发明提出一种用户子女情况判断方法、服务器及计算机可读存储介质,以解决以上技术问题。

[0005] 首先,为实现上述目的,本发明提出一种用户子女情况判断方法,该方法包括步骤:

[0006] 采集用户在预设时间段内的基于位置的服务LBS轨迹数据;

[0007] 获取预设范围内的兴趣点POI数据;

[0008] 将用户的LBS轨迹数据和所述POI数据进行匹配,分析出用户在预设类型的活动场所出现的特征信息;

[0009] 根据用户在预设类型的活动场所出现的所述特征信息,采用随机森林算法对用户进行社交群体的划分;

[0010] 获取所述社交群体中用户的年龄信息;及

[0011] 根据各个所述社交群体中用户的年龄差判断用户子女情况,并建立用户的子女标签。

[0012] 可选地,该方法在采集到所述用户LBS轨迹数据后还包括步骤:

[0013] 使用数据清洗技术对所采集的用户LBS轨迹数据进行预处理,包括检查数据一致性,处理无效值和缺失值。

[0014] 可选地,该方法在分析出用户在预设类型的活动场所出现的特征信息之后还包括步骤:

[0015] 基于Fisher准则和 $\chi^2$ 检验的筛选变量算法,剔除与用户子女情况判断无关的特征。

[0016] 可选地,所述POI和所述预设类型的活动场所主要包括亲子类活动场所。

[0017] 可选地,所述特征信息包括用户的设备号、出现的每个所述预设类型的活动场所、在各活动场所出现的次数、出现的时间点、花费的时长。

[0018] 可选地,所述采用随机森林算法对用户进行社交群体的划分的步骤包括:根据每个用户的所述特征信息,针对用户在所述活动场所出现的时间、位置和出现次数作为分区特征,将同时出现在一个活动场所的不同用户进行社交群体的分类。

[0019] 可选地,所述根据各个社交群体中用户年龄差判断用户子女情况的步骤包括:

[0020] 根据所获取的用户的年龄信息计算各个社交群体中每两个用户之间的年龄差,若所述社交群体中有两个用户的年龄差大于18岁,则判断所述两个用户之间存在亲子关系,表示所述两个用户中年长用户有子女。

[0021] 可选地,所述根据各个所述社交群体中用户的年龄差判断用户子女情况的步骤还包括:

[0022] 获取大数据中用户的性别信息,根据所述性别信息进一步确认所述年长用户的子女性别。

[0023] 此外,为实现上述目的,本发明还提供一种服务器,包括存储器、处理器,所述存储器上存储有可在所述处理器上运行的用户子女情况判断系统,所述用户子女情况判断系统被所述处理器执行时实现如上述的用户子女情况判断方法的步骤。

[0024] 进一步地,为实现上述目的,本发明还提供一种计算机可读存储介质,所述计算机可读存储介质存储有用用户子女情况判断系统,所述用户子女情况判断系统可被至少一个处理器执行,以使所述至少一个处理器执行如上述的用户子女情况判断方法的步骤。

[0025] 相较于现有技术,本发明所提出的用户子女情况判断方法、服务器及计算机可读存储介质,可以通过LBS轨迹数据对目标人群进行智能化信息采集,减轻人力劳动及节约成本,并根据所采集的数据和随机森林算法对用户进行社交群体的划分,然后根据社交群体中用户的年龄信息判断用户是否有子女,从而更容易地对用户进行有针对性的消费推荐。一方面便于企业对消费者用户进行定向宣传推荐,较好地提升了推荐效果;在另一方面便于用户快速直接地获取所需的推荐消息,且避免被过多不需要的推荐所打扰,提升用户体验。

## 附图说明

[0026] 图1是本发明服务器一可选的硬件架构的示意图;

[0027] 图2是本发明用户子女情况判断系统第一实施例的程序模块示意图;

[0028] 图3是本发明用户子女情况判断系统第二实施例的程序模块示意图;

[0029] 图4是本发明用户子女情况判断方法第一实施例的流程示意图;

[0030] 图5是本发明用户子女情况判断方法第二实施例的流程示意图;

[0031] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

## 具体实施方式

[0032] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明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用以解释本发明,并不用于限定本发明。基于本发明中的实施例,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,都属于本发明保护的范围。

[0033] 需要说明的是,在本发明中涉及“第一”、“第二”等的描述仅用于描述目的,而不能理解为指示或暗示其相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。另外,各个实施例之间的技术方案可以相互结合,但是必须是以本领域普通技术人员能够实现为基础,当技术方案的结合出现相互矛盾或无法实现时应当认为这种技术方案的结合不存在,也不在本发明要求的保护范围之内。

[0034] 参阅图1所示,是本发明服务器2一可选的硬件架构的示意图。

[0035] 本实施例中,所述服务器2可包括,但不限于,可通过系统总线相互通信连接存储器11、处理器12、网络接口13。需要指出的是,图1仅示出了具有组件11-13的服务器2,但是应理解的是,并不要求实施所有示出的组件,可以替代的实施更多或者更少的组件。

[0036] 其中,所述服务器2可以是机架式服务器、刀片式服务器、塔式服务器或机柜式服务器等计算设备,该服务器2可以是独立的服务器,也可以是多个服务器所组成的服务器集群。

[0037] 所述存储器11至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器(例如,SD或DX存储器等)、随机访问存储器(RAM)、静态随机访问存储器(SRAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、可编程只读存储器(PROM)、磁性存储器、磁盘、光盘等。在一些实施例中,所述存储器11可以是所述服务器2的内部存储单元,例如该服务器2的硬盘或内存。在另一些实施例中,所述存储器11也可以是所述服务器2的外部存储设备,例如该服务器2上配备的插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)等。当然,所述存储器11还可以既包括所述服务器2的内部存储单元也包括其外部存储设备。本实施例中,所述存储器11通常用于存储安装于所述服务器2的操作系统和各类应用软件,例如用户子女情况判断系统200的程序代码等。此外,所述存储器11还可以用于暂时地存储已经输出或者将要输出的各类数据。

[0038] 所述处理器12在一些实施例中可以是中央处理器(Central Processing Unit, CPU)、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器12通常用于控制所述服务器2的总体操作。本实施例中,所述处理器12用于运行所述存储器11中存储的程序代码或者处理数据,例如运行所述的用户子女情况判断系统200等。

[0039] 所述网络接口13可包括无线网络接口或有线网络接口,该网络接口13通常用于在所述服务器2与其他电子设备之间建立通信连接。

[0040] 至此,已经详细介绍了本发明相关设备的硬件结构和功能。下面,将基于上述介绍提出本发明的各个实施例。

[0041] 首先,本发明提出一种用户子女情况判断系统200。

[0042] 参阅图2所示,是本发明用户子女情况判断系统200第一实施例的程序模块图。

[0043] 本实施例中,所述用户子女情况判断系统200包括一系列的存储于存储器11上的计算机程序指令,当该计算机程序指令被处理器12执行时,可以实现本发明各实施例的用户子女情况判断操作。在一些实施例中,基于该计算机程序指令各部分所实现的特定的操作,用户子女情况判断系统200可以被划分为一个或多个模块。例如,在图2中,所述用户子

女情况判断系统200可以被分割成采集模块201、获取模块202、匹配模块203、划分模块204、判断模块205。其中：

[0044] 所述采集模块201,用于采集用户在预设时间段内的基于位置的服务(Location Based Services,LBS) 轨迹数据。

[0045] 具体地,LBS是利用各类型的定位技术来获取定位设备当前的所在位置,通过移动互联网向定位设备提供信息资源和基础服务。其由用户对应的移动终端使用无线通信网络(或卫星定位系统),基于空间数据库,获取用户的地理位置坐标信息并与其他信息集成以向用户提供所需的与位置相关的增值服务。目前,LBS常用的定位技术包括:全球定位系统、WiFi定位、IP地址定位、三角测量法等。LBS的主要特点包括:覆盖范围广、定位精度高、操作简便。

[0046] 所述LBS轨迹数据具体包含的信息为用户对应的设备(移动终端,例如手机、平板电脑、智能穿戴式设备、车载设备、儿童电话手表等)在某个时间点出现在某个经纬度(地点),即包括时间和位置(经纬度)信息。

[0047] 所述预设时间段可以是一周、一个月等。在所述预设时间段内,采集每个用户的所有LBS轨迹数据。

[0048] 所述获取模块202,用于获取预设范围内的兴趣点(Point of Interest,POI) 数据。

[0049] 具体地,在地理信息系统中,一个POI可以是一栋房子、一个商铺、一个邮筒、一个公交站等。每个POI包含名称、类别、坐标等信息,其中,每个分类都有相应的行业的代码和名称对应。简单来说,所述POI数据可以显示每个位置(经纬度)对应什么场所。一般情况下,所述POI数据包括兴趣点位置信息、兴趣点类别和兴趣点图片等。

[0050] 在本实施例中,所述POI主要包括小区、商圈、学校、公园、游乐场等亲子类活动场所。

[0051] 所述预设范围可以是该用户所在的地区或者该用户的所述LBS轨迹数据所涵盖的地理范围。

[0052] 所述匹配模块203,用于将用户的LBS轨迹数据和所述POI数据进行匹配,分析出用户在预设类型的活动场所出现的特征信息。

[0053] 具体地,根据所述POI数据可以知道哪些位置属于预设类型的场所(在本实施例中主要是亲子类活动场所),再将该用户的LBS轨迹数据和所述POI数据进行匹配,可以分析出所述LBS轨迹数据对应的预设类型的活动场所,即该用户出现在哪些预设活动场所中。另外,还可以进一步统计出用户在各活动场所出现的次数、出现的时间点、花费的时长等。也就是说,通过对采集到的所述LBS轨迹数据进行加工,并与所述POI数据进行匹配,可以得到该用户的(移动终端)设备号、出现的每个预设类型的活动场所、在各活动场所出现的次数、出现的时间点、花费的时长等特征信息。

[0054] 所述划分模块204,用于根据用户在预设类型的活动场所出现的所述特征信息,采用随机森林算法对用户进行社交群体的划分。

[0055] 具体地,所述采用随机森林算法对用户进行社交群体的划分,即根据每个用户的所述特征信息,针对同时出现在某一活动场所的不同用户(不同设备号)进行社交关系及群体的分类。

[0056] 该步骤具体过程如下：

[0057] (1) 在训练集中,对观测对象(所述用户/设备号)进行有放回的抽样,构建出不同的子数据集。

[0058] (2) 将观测对象在所述活动场所出现的时间、位置和出现次数作为分区特征,在每个子数据集再对分区特征进行有放回的抽样,得到新的子数据集。

[0059] (3) 对每个新的子数据集都建立决策树,每个决策树的建立标准为:尽可能使分区结果中每个叶子区域包含了在同一时间和地点出现且出现次数相同的不同设备号。

[0060] (4) 将随机决策树分区结果中每个叶子区域中的不同设备号视为社交群体。所述社交群体对应的用户中有可能包括子女和父母。

[0061] 使用随机森林算法,有效地避免了传统单个决策树会出现的训练速度过慢和容易过拟合的问题。

[0062] 例如:用户A/设备号001、用户B/设备号002和用户C/设备号003均在周六12:35-13:47出现在餐厅D,在周六14:05-17:30出现在游乐场E,在周日09:27-11:36出现在早教中心F,在周日17:42-18:53出现在餐厅D,则有可能会划分为一个社交群体。

[0063] 所述获取模块202,还用于获取社交群体中用户的年龄信息。

[0064] 具体地,为了进一步判断各个社交群体中用户之间的关系,尤其是是否存在亲子关系,需要获取各个社交群体中每个用户的年龄信息。

[0065] 例如,针对保险行业,可以从企业用户大数据库中获取各个用户的年龄信息。

[0066] 所述判断模块205,用于根据各个社交群体中用户年龄差判断用户子女情况,并建立用户的子女标签。

[0067] 具体地,在一般情况下,当一个社交群体中的用户存在亲子关系时,则用户之间(父母和子女)的年龄差应该大于18岁。因此,根据所获取的用户的年龄信息计算各个社交群体中每两个用户之间的年龄差,若有用户的年龄差大于18岁,则判断这两个用户之间存在亲子关系,即年长者有子女(子女为年幼者)。也就是说,如果用户的年龄差超过18岁,且有规律地出现在同一个所述预设类型的活动场所(同一亲子类活动场所),则可以认为其中年长者有子女。如果没有规律地出现在这些预设类型的活动场所,或无大于18岁的年龄差,则不认为这类用户有子女。

[0068] 可选地,所述获取模块202还可以获取大数据中用户的性别信息,所述判断模块205根据所述性别信息进一步确认该年长用户的子女性别。

[0069] 当按照上述方案判断出用户的子女情况后,可以对用户根据判断结果建立子女标签(例如有无子女、子女个数、性别等)。从而,可以对用户进行有针对性的消费推荐(例如对有子女的用户推荐儿童险),提升推荐效果和用户体验。

[0070] 本实施例提供的用户子女情况判断系统,可以通过LBS轨迹数据对目标人群进行智能化信息采集,并根据所采集的数据和随机森林算法对用户进行社交群体的划分,然后根据社交群体中用户的年龄信息判断用户是否有子女,从而更容易地对用户进行有针对性的消费推荐。

[0071] 其中,LBS轨迹数据的采集可以从用户移动终端中自动完成,更精确地收集了目标用户在特定地点出现的时间和次数,减轻了繁琐的人力劳动(问卷调查),很大程度上节约了成本。

[0072] 并且,该系统可以快速识别出现有用户是否有子女,并预测出新的用户的子女标签情况,收集到更多样化的用户信息。之后可以根据用户的子女标签进行精准推荐,判断用户是否需要某种特定的与孩童相关的产品,以及消费的意愿。一方面便于企业(商家)对消费者用户进行定向宣传推荐,较好地提升了推荐效果;在另一方面便于用户快速直接地获取所需的推荐消息,且避免被过多不需要的推荐所打扰,提升用户体验。

[0073] 参阅图3所示,是本发明用户子女情况判断系统200第二实施例的程序模块图。本实施例中,所述的用户子女情况判断系统200除了包括第一实施例中的所述采集模块201、获取模块202、匹配模块203、划分模块204、判断模块205之外,还包括清洗模块206和筛选模块207。

[0074] 所述清洗模块206,用于在所述采集模块201采集到用户LBS轨迹数据后,使用数据清洗技术对所采集的用户LBS轨迹数据进行预处理。

[0075] 具体地,数据清洗是指利用数理统计、数据挖掘等技术或预定义的清理规则将脏数据转化为满足数据质量要求的数据,主要包括检查数据一致性,处理无效值和缺失值等。

[0076] 在本实施例中,所述预处理为使用数据清洗技术将采集的原始数据(用户LBS轨迹数据)中的缺失值、异常值、错误数据删去。本实施例采用现有常用的数据清洗技术即可,具体数据清洗过程在此不再赘述。

[0077] 所述筛选模块207,用于针对所述匹配模块203分析的用户特征信息进行特征筛选,即基于Fisher准则和 $\chi^2$ 检验的筛选变量算法,剔除与用户子女情况判断无关的特征。

[0078] 具体地,在特征工程中,对定类特征进行基于Fisher准则和 $\chi^2$ 检验的筛选变量算法,剔除无关、冗余的特征。在分类数据和分类数据之间进行 $\chi^2$ 检验,研究它们之间的差异关系。如果多个特征的假设检验结果为它们无显著性差异特征,则删去冗余的特征。

[0079] 在本实施例中,根据用户的所述LBS轨迹数据和所述POI数据,可以分析出该用户的多个特征,但这些特征可能与后续判断用户是否有子女有关,也可能关联不大。因此,需要基于Fisher准则和 $\chi^2$ 检验的筛选变量算法,剔除与用户子女情况判断无关的特征,保留关联性较大的特征,分析得到这些特征信息后,用于后续进行用户分类和子女情况判断。

[0080] 本实施例提供的用户子女情况判断系统,可以通过对所采集的用户LBS轨迹数据进行数据清洗和通过特征筛选剔除与用户子女情况判断无关的特征,来进一步提高对用户子女情况分析判断的结果的准确性和处理效率。

[0081] 此外,本发明还提出一种用户子女情况判断方法。

[0082] 参阅图4所示,是本发明用户子女情况判断方法第一实施例的流程示意图。在本实施例中,根据不同的需求,图4所示的流程图中的步骤的执行顺序可以改变,某些步骤可以省略。

[0083] 该方法包括以下步骤:

[0084] 步骤S400,采集用户在预设时间段内的LBS轨迹数据。

[0085] 具体地,LBS是利用各类型的定位技术来获取定位设备当前的所在位置,通过移动互联网向定位设备提供信息资源和基础服务。其由用户对应的移动终端使用无线网络(或卫星定位系统),基于空间数据库,获取用户的地理位置坐标信息并与其他信息集成以向用户提供所需的与位置相关的增值服务。目前,LBS常用的定位技术包括:全球定位系统、WiFi定位、IP地址定位、三角测量法等。LBS的主要特点包括:覆盖范围广、定位精度高、操作



简便。

[0086] 所述LBS轨迹数据具体包含的信息为用户对应的设备(移动终端,例如手机、平板电脑、智能穿戴式设备、车载设备、儿童电话手表等)在某个时间点出现在某个经纬度(地点),即包括时间和位置(经纬度)信息。

[0087] 所述预设时间段可以是一周、一个月等。在所述预设时间段内,采集每个用户的所有LBS轨迹数据。

[0088] 步骤S402,获取预设范围内的POI数据。

[0089] 具体地,在地理信息系统中,一个POI可以是一栋房子、一个商铺、一个邮筒、一个公交站等。每个POI包含名称、类别、坐标等信息,其中,每个分类都有相应的行业的代码和名称对应。简单来说,所述POI数据可以显示每个位置(经纬度)对应什么场所。一般情况下,所述POI数据包括兴趣点位置信息、兴趣点类别和兴趣点图片等。

[0090] 在本实施例中,所述POI主要包括小区、商圈、学校、公园、游乐场等亲子类活动场所。

[0091] 所述预设范围可以是该用户所在的地区或者该用户的所述LBS轨迹数据所涵盖的地理范围。

[0092] 步骤S404,将用户的LBS轨迹数据和所述POI数据进行匹配,分析出用户在预设类型的活动场所出现的特征信息。

[0093] 具体地,根据所述POI数据可以知道哪些位置属于预设类型的场所(在本实施例中主要是亲子类活动场所),再将该用户的LBS轨迹数据和所述POI数据进行匹配,可以分析出所述LBS轨迹数据对应的预设类型的活动场所,即该用户出现在哪些预设活动场所中。另外,还可以进一步统计出用户在各活动场所出现的次数、出现的时间点、花费的时长等。也就是说,通过对采集到的所述LBS轨迹数据进行加工,并与所述POI数据进行匹配,可以得到该用户的(移动终端)设备号、出现的每个预设类型的活动场所、在各活动场所出现的次数、出现的时间点、花费的时长等特征信息。

[0094] 步骤S406,根据用户在预设类型的活动场所出现的所述特征信息,采用随机森林算法对用户进行社交群体的划分。

[0095] 具体地,所述采用随机森林算法对用户进行社交群体的划分,即根据每个用户的所述特征信息,针对同时出现在某一活动场所的不同用户(不同设备号)进行社交关系及群体的分类。

[0096] 该步骤具体过程如下:

[0097] (1) 在训练集中,对观测对象(所述用户/设备号)进行有放回的抽样,构建出不同的子数据集。

[0098] (2) 将观测对象在所述活动场所出现的时间、位置和出现次数作为分区特征,在每个子数据集再对分区特征进行有放回的抽样,得到新的子数据集。

[0099] (3) 对每个新的子数据集都建立决策树,每个决策树的建立标准为:尽可能使分区结果中每个叶子区域包含了在同一时间和地点出现且出现次数相同的不同设备号。

[0100] (4) 将随机决策树分区结果中每个叶子区域中的不同设备号视为社交群体。所述社交群体对应的用户中有可能包括子女和父母。

[0101] 使用随机森林算法,有效地避免了传统单个决策树会出现的训练速度过慢和容易

过拟合的问题。

[0102] 例如：用户A/设备号001、用户B/设备号002和用户C/设备号003均在周六12:35-13:47出现在餐厅D,在周六14:05-17:30出现在游乐场E,在周日09:27-11:36出现在早教中心F,在周日17:42-18:53出现在餐厅D,则有可能会划分为一个社交群体。

[0103] 步骤S408,获取社交群体中用户的年龄信息。

[0104] 具体地,为了进一步判断各个社交群体中用户之间的关系,尤其是是否存在亲子关系,需要获取各个社交群体中每个用户的年龄信息。

[0105] 例如,针对保险行业,可以从企业用户大数据库中获取各个用户的年龄信息。

[0106] 步骤S410,根据各个社交群体中用户年龄差判断用户子女情况,并建立用户的子女标签。

[0107] 具体地,在一般情况下,当一个社交群体中的用户存在亲子关系时,则用户之间(父母和子女)的年龄差应该大于18岁。因此,根据所获取的用户的年龄信息计算各个社交群体中每两个用户之间的年龄差,若有用户的年龄差大于18岁,则判断这两个用户之间存在亲子关系,即年长者有子女(子女为年幼者)。也就是说,如果用户的年龄差超过18岁,且有规律地出现在同一个所述预设类型的活动场所(同一亲子类活动场所),则可以认为其中年长者有子女。如果没有规律地出现在这些预设类型的活动场所,或无大于18岁的年龄差,则不认为这类用户有子女。

[0108] 可选地,还可以获取大数据中用户的性别信息,根据所述性别信息进一步确认该年长用户的子女性别。

[0109] 当按照上述方案判断出用户的子女情况后,可以对用户根据判断结果建立子女标签(例如有无子女、子女个数、性别等)。从而,可以对用户进行有针对性的消费推荐(例如对有子女的用户推荐儿童险),提升推荐效果和用户体验。

[0110] 本实施例提供的用户子女情况判断方法,可以通过LBS轨迹数据对目标人群进行智能化信息采集,并根据所采集的数据和随机森林算法对用户进行社交群体的划分,然后根据社交群体中用户的年龄信息判断用户是否有子女,从而更容易地对用户进行有针对性的消费推荐。

[0111] 其中,LBS轨迹数据的采集可以从用户移动终端中自动完成,更精确地收集了目标用户在特定地点出现的时间和次数,减轻了繁琐的人力劳动(问卷调查),很大程度上节约了成本。

[0112] 并且,该方法可以快速识别出现有用户是否有子女,并预测出新的用户的子女标签情况,收集到更多样化的用户信息。之后可以根据用户的子女标签进行精准推荐,判断用户是否需要某种特定的与孩童相关的产品,以及消费的意愿。一方面便于企业(商家)对消费者用户进行定向宣传推荐,较好地提升了推荐效果;在另一方面便于用户快速直接地获取所需的推荐消息,且避免被过多不需要的推荐所打扰,提升用户体验。

[0113] 如图5所示,是本发明用户子女情况判断方法的第二实施例的流程示意图。本实施例中,所述用户子女情况判断方法的步骤S500、S504-S506、S510-S514与第一实施例的步骤S400-S410相类似,区别在于该方法还包括步骤S502和S508。值得注意的是,在其他实施例中,该方法也可以在所述第一实施例的基础上仅添加步骤S02或S08。

[0114] 该方法包括以下步骤:

[0115] 步骤S500,采集用户在预设时间段内的LBS轨迹数据。

[0116] 具体地,LBS是利用各类型的定位技术来获取定位设备当前的所在位置,通过移动互联网向定位设备提供信息资源和基础服务。其由用户对应的移动终端使用无线网络(或卫星定位系统),基于空间数据库,获取用户的地理位置坐标信息并与其他信息集成以向用户提供所需的与位置相关的增值服务。目前,LBS常用的定位技术包括:全球定位系统、WiFi定位、IP地址定位、三角测量法等。LBS的主要特点包括:覆盖范围广、定位精度高、操作简便。

[0117] 所述LBS轨迹数据具体包含的信息为用户对应的设备(移动终端,例如手机、平板电脑、智能穿戴式设备、车载设备、儿童电话手表等)在某个时间点出现在某个经纬度(地点),即包括时间和位置(经纬度)信息。

[0118] 所述预设时间段可以是一周、一个月等。在所述预设时间段内,采集每个用户的所有LBS轨迹数据。

[0119] 步骤S502,使用数据清洗技术对所采集的用户LBS轨迹数据进行预处理。

[0120] 具体地,数据清洗是指利用数理统计、数据挖掘等技术或预定义的清理规则将脏数据转化为满足数据质量要求的数据,主要包括检查数据一致性,处理无效值和缺失值等。

[0121] 在本实施例中,所述预处理为使用数据清洗技术将采集的原始数据(用户LBS轨迹数据)中的缺失值、异常值、错误数据删去。本实施例采用现有常用的数据清洗技术即可,具体数据清洗过程在此不再赘述。

[0122] 步骤S504,获取预设范围内的POI数据。

[0123] 具体地,在地理信息系统中,一个POI可以是一栋房子、一个商铺、一个邮筒、一个公交站等。每个POI包含名称、类别、坐标等信息,其中,每个分类都有相应的行业的代码和名称对应。简单来说,所述POI数据可以显示每个位置(经纬度)对应什么场所。一般情况下,所述POI数据包括兴趣点位置信息、兴趣点类别和兴趣点图片等。

[0124] 在本实施例中,所述POI主要包括小区、商圈、学校、公园、游乐场等亲子类活动场所。

[0125] 所述预设范围可以是该用户所在的地区或者该用户的所述LBS轨迹数据所涵盖的地理范围。

[0126] 步骤S506,将用户的LBS轨迹数据和所述POI数据进行匹配,分析出用户在预设类型的活动场所出现的特征信息。

[0127] 具体地,根据所述POI数据可以知道哪些位置属于预设类型的场所(在本实施例中主要是亲子类活动场所),再将该用户的LBS轨迹数据和所述POI数据进行匹配,可以分析出所述LBS轨迹数据对应的预设类型的活动场所,即该用户出现在哪些预设活动场所中。另外,还可以进一步统计出用户在各活动场所出现的次数、出现的时间点、花费的时长等。也就是说,通过对采集到的所述LBS轨迹数据进行加工,并与所述POI数据进行匹配,可以得到该用户的(移动终端)设备号、出现的每个预设类型的活动场所、在各活动场所出现的次数、出现的时间点、花费的时长等特征信息。

[0128] 步骤S508,基于Fisher准则和 $\chi^2$ 检验的筛选变量算法,剔除与用户子女情况判断无关的特征。

[0129] 具体地,在特征工程中,对定类特征进行基于Fisher准则和 $\chi^2$ 检验的筛选变量算

法,剔除无关、冗余的特征。在分类数据和分类数据之间进行 $x^2$ 检验,研究它们之间的差异关系。如果多个特征的假设检验结果为它们无显著性差异特征,则删去冗余的特征。

[0130] 在本实施例中,根据用户的所述LBS轨迹数据和所述POI数据,可以分析出该用户的多个特征,但这些特征可能与后续判断用户是否有子女有关,也可能关联不大。因此,需要基于Fisher准则和 $x^2$ 检验的筛选变量算法,剔除与用户子女情况判断无关的特征,保留关联性较大的特征,分析得到这些特征信息后,用于后续进行用户分类和子女情况判断。

[0131] 步骤S510,根据用户在预设类型的活动场所出现的所述特征信息,采用随机森林算法对用户进行社交群体的划分。

[0132] 具体地,所述采用随机森林算法对用户进行社交群体的划分,即根据每个用户的所述特征信息,针对同时出现在某一活动场所的不同用户(不同设备号)进行社交关系及群体的分类。

[0133] 该步骤具体过程如下:

[0134] (1) 在训练集中,对观测对象(所述用户/设备号)进行有放回的抽样,构建出不同的子数据集。

[0135] (2) 将观测对象在所述活动场所出现的时间、位置和出现次数作为分区特征,在每个子数据集再对分区特征进行有放回的抽样,得到新的子数据集。

[0136] (3) 对每个新的子数据集都建立决策树,每个决策树的建立标准为:尽可能使分区结果中每个叶子区域包含了在同一时间和地点出现且出现次数相同的不同设备号。

[0137] (4) 将随机决策树分区结果中每个叶子区域中的不同设备号视为社交群体。所述社交群体对应的用户中有可能包括子女和父母。

[0138] 使用随机森林算法,有效地避免了传统单个决策树会出现的训练速度过慢和容易过拟合的问题。

[0139] 例如:用户A/设备号001、用户B/设备号002和用户C/设备号003均在周六12:35-13:47出现在餐厅D,在周六14:05-17:30出现在游乐场E,在周日09:27-11:36出现在早教中心F,在周日17:42-18:53出现在餐厅D,则有可能会划分为一个社交群体。

[0140] 步骤S512,获取社交群体中用户的年龄信息。

[0141] 具体地,为了进一步判断各个社交群体中用户之间的关系,尤其是是否存在亲子关系,需要获取各个社交群体中每个用户的年龄信息。

[0142] 例如,针对保险行业,可以从企业用户大数据库中获取各个用户的年龄信息。

[0143] 步骤S514,根据各个社交群体中用户年龄差判断用户子女情况,并建立用户的子女标签。

[0144] 具体地,在一般情况下,当一个社交群体中的用户存在亲子关系时,则用户之间(父母和子女)的年龄差应该大于18岁。因此,根据所获取的用户的年龄信息计算各个社交群体中每两个用户之间的年龄差,若有用户的年龄差大于18岁,则判断这两个用户之间存在亲子关系,即年长者有子女(子女为年幼者)。也就是说,如果用户的年龄差超过18岁,且有规律地出现在同一个所述预设类型的活动场所(同一亲子类活动场所),则可以认为其中年长者有子女。如果没有规律地出现在这些预设类型的活动场所,或无大于18岁的年龄差,则不认为这类用户有子女。

[0145] 可选地,还可以获取大数据中用户的性别信息,根据所述性别信息进一步确认该

年长用户的子女性别。

[0146] 当按照上述方案判断出用户的子女情况后,可以对用户根据判断结果建立子女标签(例如有无子女、子女个数、性别等)。从而,可以对用户进行有针对性的消费推荐(例如对有子女的用户推荐儿童险),提升推荐效果和用户体验。

[0147] 本实施例提供的用户子女情况判断方法,可以通过LBS轨迹数据对目标人群进行智能化信息采集,减轻人力劳动及节约成本,并根据所采集的数据和随机森林算法对用户进行社交群体的划分,然后根据社交群体中用户的年龄信息判断用户是否有子女,从而更容易地对用户进行有针对性的消费推荐。一方面便于企业对消费者用户进行定向宣传推荐,较好地提升了推荐效果;在另一方面便于用户快速直接地获取所需的推荐消息,且避免被过多不需要的推荐所打扰,提升用户体验。并且,该方法可以通过对所采集的用户LBS轨迹数据进行数据清洗和通过特征筛选剔除与用户子女情况判断无关的特征,来进一步提高对用户子女情况分析判断的结果的准确性和处理效率。

[0148] 本发明还提供了另一种实施方式,即提供一种计算机可读存储介质,所述计算机可读存储介质存储有用户子女情况判断程序,所述用户子女情况判断程序可被至少一个处理器执行,以使所述至少一个处理器执行如上述的用户子女情况判断方法的步骤。

[0149] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0150] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本发明各个实施例所述的方法。

[0151] 以上仅为本发明的优选实施例,并非因此限制本发明的专利范围,凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本发明的专利保护范围内。

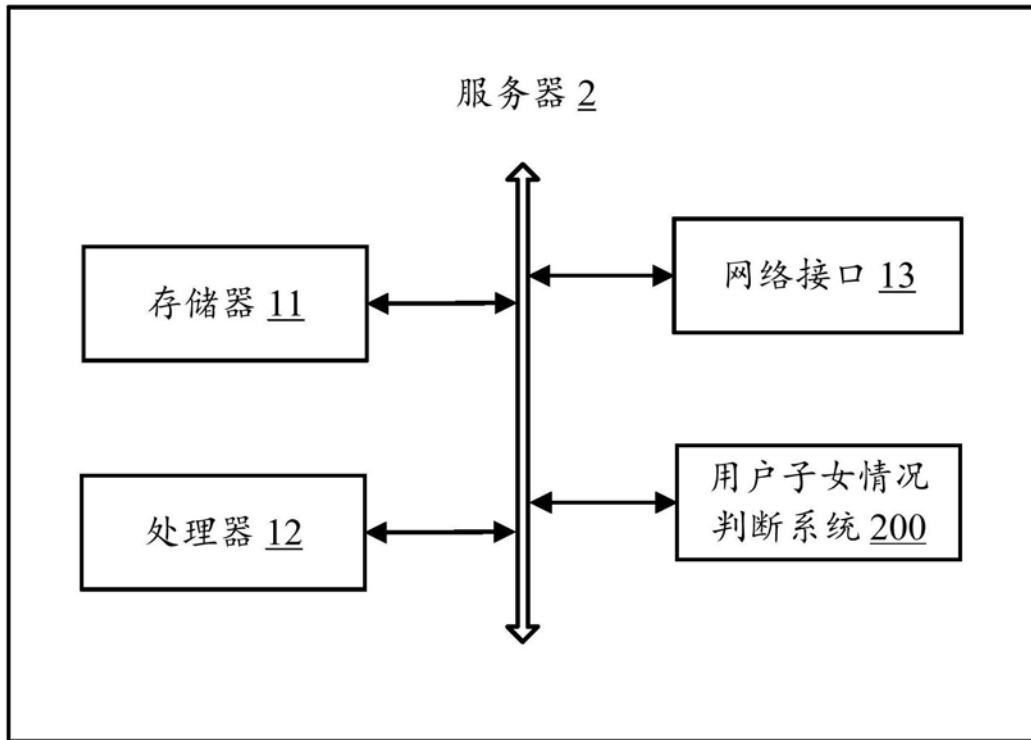


图1

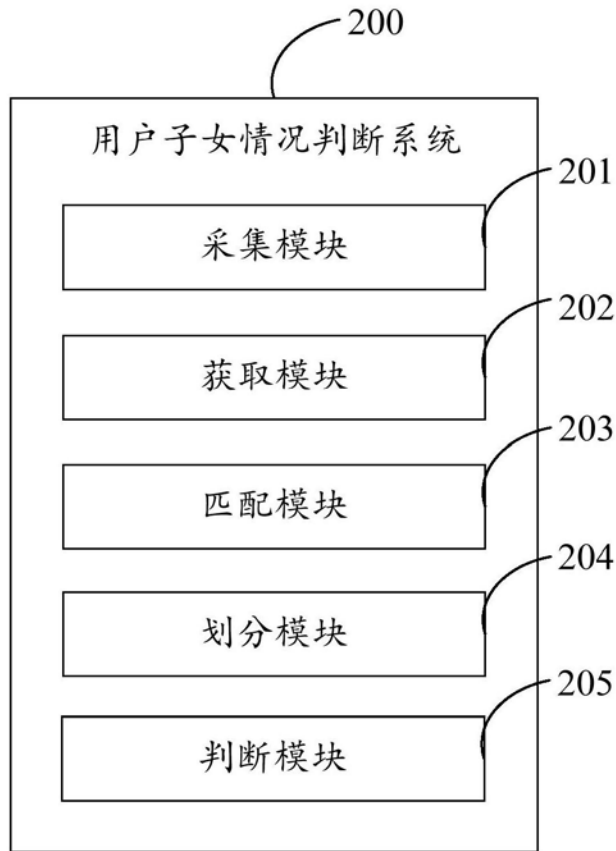


图2

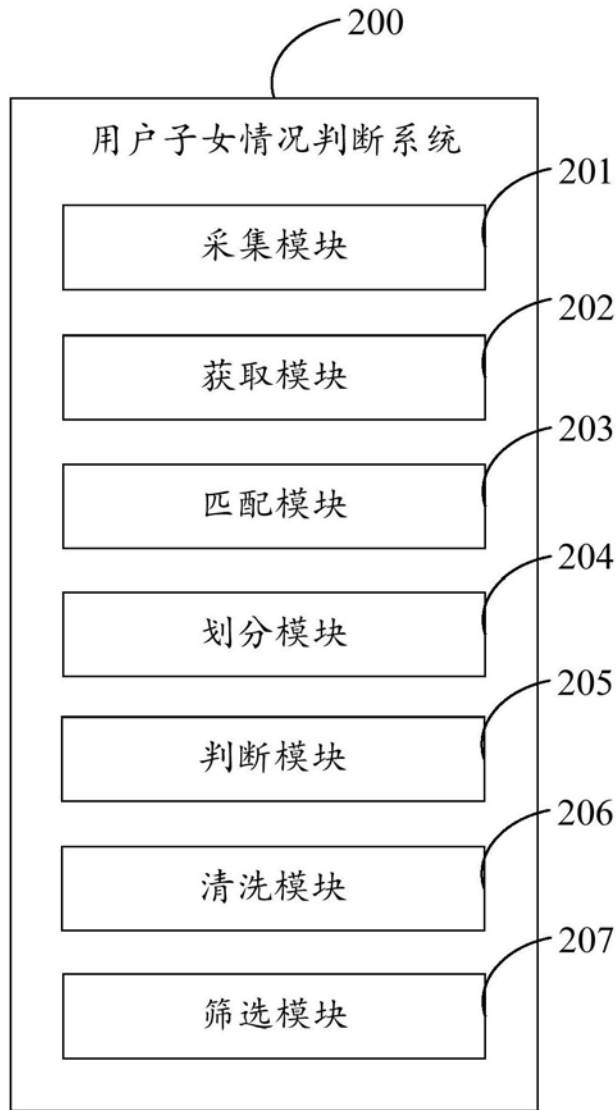


图3



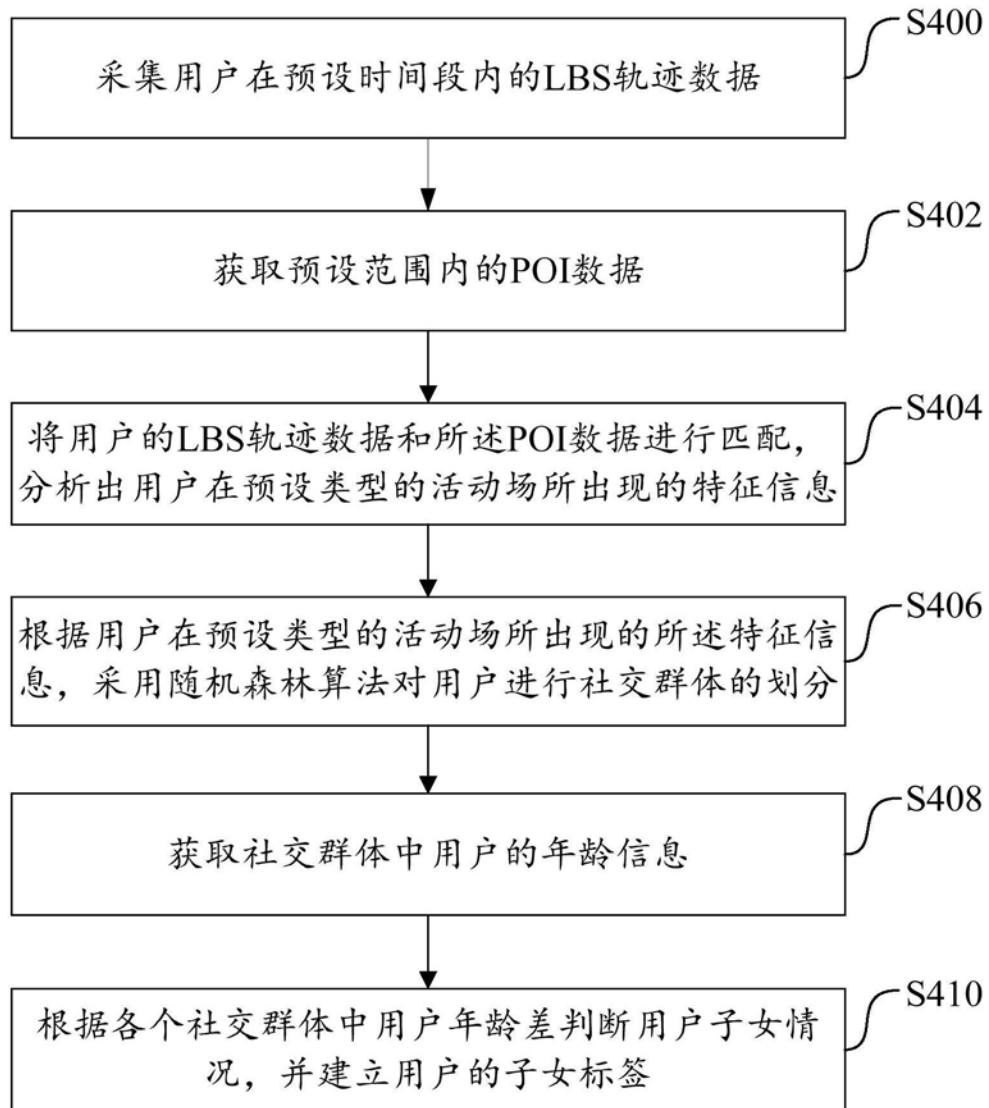


图4

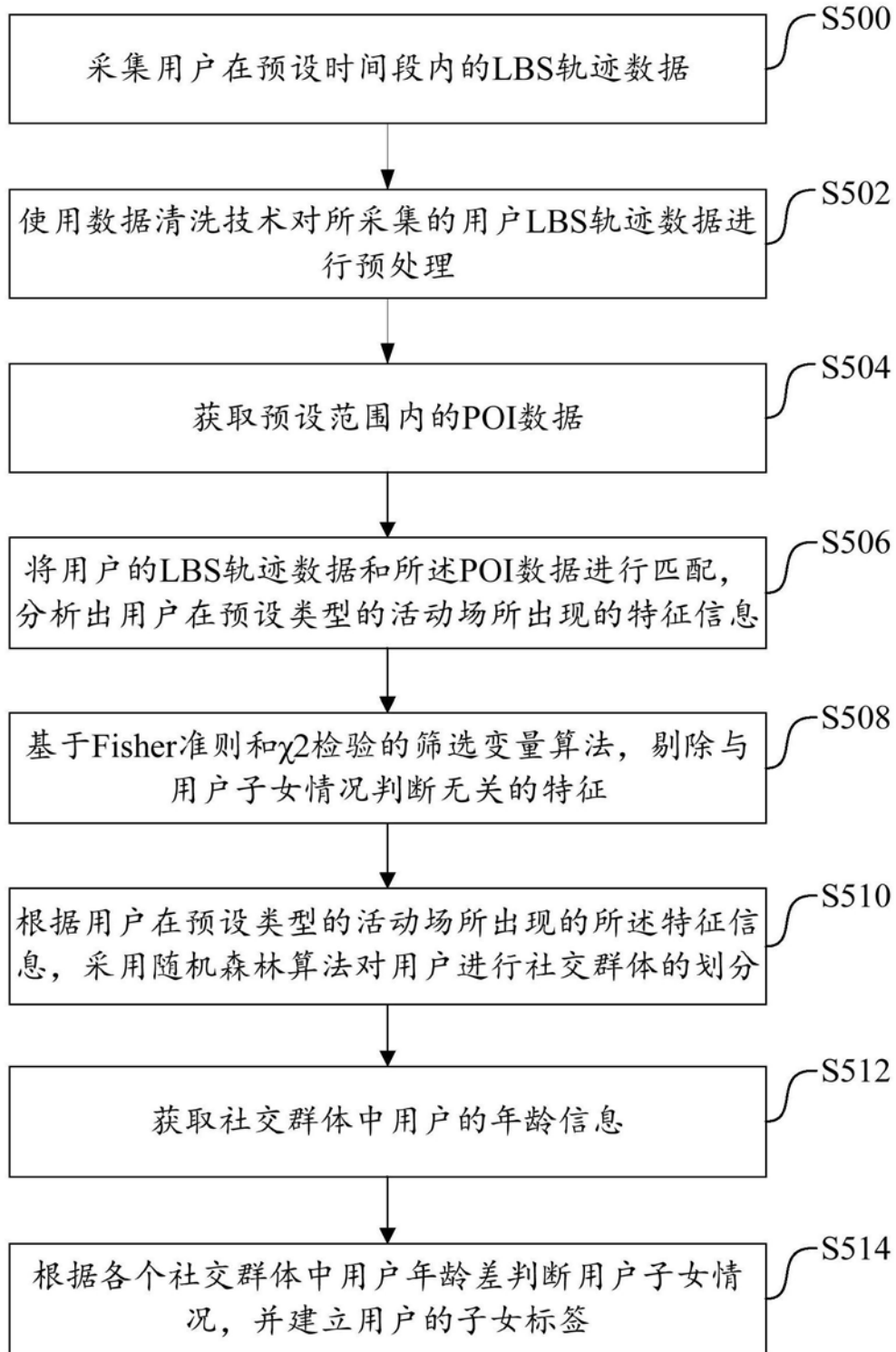


图5