



(12) 发明专利

(10) 授权公告号 CN 110619090 B

(45) 授权公告日 2022. 01. 21

(21) 申请号 201910715730.2

G06Q 10/06 (2012.01)

(22) 申请日 2019.08.05

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 108804551 A, 2018.11.13

申请公布号 CN 110619090 A

US 2018/0352373 A1, 2018.12.06

CN 108170712 A, 2018.06.15

(43) 申请公布日 2019.12.27

审查员 冯雅

(73) 专利权人 香港理工大学深圳研究院

地址 518057 广东省深圳市南山区高新园

南区粤兴一道18号香港理工大学产学

研大楼205室

(72) 发明人 史文中 刘哲维

(74) 专利代理机构 深圳中一专利商标事务所

44237

代理人 曹小翠

(51) Int. Cl.

G06F 16/9537 (2019.01)

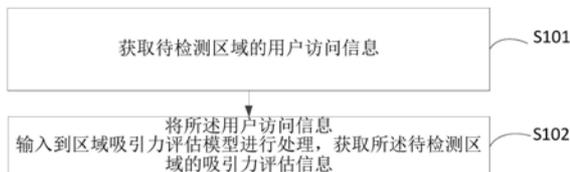
权利要求书2页 说明书10页 附图4页

(54) 发明名称

一种区域吸引力评估方法及设备

(57) 摘要

本发明适用于计算机技术领域,提供了一种区域吸引力评估方法及设备,包括:获取待检测区域的用户访问信息;将所述用户访问信息输入到区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息。上述方法,将待检测区域的用户访问信息输入到已训练的区域吸引力评估模型进行处理,可以准确的得到区域吸引力评估信息,基于区域吸引力评估信息可以对城市热点区域进行推荐,该模型不同于现有的经验模型,不过多的依赖历史数据,在不同的场景中都可以得到区域吸引力评估信息。



1. 一种区域吸引力评估方法,其特征在于,所述方法用以对热点区域进行推荐,包括:
获取待检测区域的用户访问信息;所述用户访问信息包括多个访问数量元素;所述访问数量元素为所述待检测区域中每个用户对所述待检测区域的访问数量;

将所述用户访问信息输入到预先训练的区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息;其中,在训练过程中,所述区域吸引力评估模型的输入为样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签,所述区域吸引力评估模型的输出为所述样本区域的样本吸引力评估信息;所述样本区域为聚集样本签到数据的区域;所述吸引力评估信息标签根据多个用户对所述样本区域的历史吸引力打分得到。

2. 如权利要求1所述的区域吸引力评估方法,其特征在于,在所述将所述用户访问信息输入到区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息之前,还包括:

获取样本训练集;所述样本训练集包括样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签;

对所述样本训练集进行训练,得到用于输出所述样本区域的样本吸引力评估信息的区域吸引力评估模型。

3. 如权利要求2所述的区域吸引力评估方法,其特征在于,所述获取样本训练集,包括:
获取所述样本用户在通过目标社交应用进行签到时产生的样本签到数据;所述样本签到数据包括所述样本用户的用户标识、签到地点的地点标识及其位置信息;

基于所述样本签到数据得到所述样本区域的样本用户访问信息;

获取所述样本区域的吸引力评估信息标签;

将每个所述样本区域的样本用户访问信息和每个所述样本区域的吸引力评估信息标签关联存储,得到样本训练集。

4. 如权利要求3所述的区域吸引力评估方法,其特征在于,所述基于所述样本签到数据得到所述样本区域的样本用户访问信息,包括:

对所述样本签到数据聚类,得到样本区域信息;其中,所述样本区域信息包括至少一个样本区域;所述样本区域信息包含在所述样本区域签到的样本用户的用户标识、位于所述样本区域中的签到地点的地点标识及其位置信息;

基于所述样本区域信息获取所述样本区域的样本用户访问信息。

5. 如权利要求3所述的区域吸引力评估方法,其特征在于,所述获取所述样本用户在通过目标社交应用进行签到时产生的样本签到数据,包括:

获取所述目标社交应用的初始签到数据;所述初始签到数据包括每个样本用户的签到数量;

当所述样本用户的签到数量满足预设条件时,获取所述样本用户的签到数据。

6. 一种区域吸引力评估装置,所述装置用以对热点区域进行推荐,其特征在于,包括:
第一获取单元,用于获取待检测区域的用户访问信息;所述用户访问信息包括多个访问数量元素;所述访问数量元素为所述待检测区域中每个用户对所述待检测区域的访问数量;

第一处理单元,用于将所述用户访问信息输入到预先训练的区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息;其中,在训练过程中,所述区域吸引力评估

模型的输入为样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签,所述区域吸引力评估模型的输出为所述样本区域的样本吸引力评估信息;所述样本区域为聚集样本签到数据的区域;所述吸引力评估信息标签根据多个用户对所述样本区域的历史吸引力打分得到。

7. 如权利要求6所述的区域吸引力评估装置,其特征在于,还包括:

第二获取单元,用于获取样本训练集;所述样本训练集包括样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签;

训练单元,用于对所述样本训练集进行训练,得到用于输出所述样本区域的样本吸引力评估信息的区域吸引力评估模型。

8. 如权利要求7所述的区域吸引力评估装置,其特征在于,所述第二获取单元,包括:

第三获取单元,用于获取所述样本用户在通过目标社交应用进行签到时产生的样本签到数据;所述样本签到数据包括所述样本用户的用户标识、签到地点的地点标识及其位置信息;

第二处理单元,用于基于所述样本签到数据得到所述样本区域的样本用户访问信息;

第四获取单元,用于获取所述样本区域的吸引力评估信息标签;

第三处理单元,用于将每个所述样本区域的样本用户访问数组和每个所述样本区域的吸引力评估信息标签关联存储,得到样本训练集。

9. 一种区域吸引力评估设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现如权利要求1至5任一项所述方法的步骤。

10. 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计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至5任一项所述方法的步骤。

一种区域吸引力评估方法及设备

技术领域

[0001] 本发明属于计算机技术领域,尤其涉及一种区域吸引力评估方法及设备。

背景技术

[0002] 社交媒体平台大、更新速度快、内容多样,所以社交媒体数据也是十分的庞大且具有价值的,当前,基于社交媒体数据可以进行空间热点的挖掘,从而对热点进行评估,向用户推荐热门信息,上述方法可以广泛的应用于城市规划、旅游推荐等领域。

[0003] 但是,目前对热点评估主要是采用一些经验模型进行评估,在应用场景中有着较大的局限性,没有办法准确的对城市热点区域进行推荐。

发明内容

[0004] 有鉴于此,本发明实施例提供了一种区域吸引力评估方法及设备,以解决现有技术中对热点评估主要是采用一些经验模型进行评估,在应用场景中有着较大的局限性,没有办法准确的对城市热点区域进行推荐的问题。

[0005] 本发明实施例的第一方面提供了一种区域吸引力评估方法,包括:

[0006] 获取待检测区域的用户访问信息;所述用户访问信息包括多个访问数量元素;所述访问数量元素为所述待检测区域中每个用户对所述待检测区域的访问数量;

[0007] 将所述用户访问信息输入到区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息;其中,在训练过程中,所述区域吸引力评估模型的输入为样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签,所述区域吸引力评估模型的输出为所述样本区域的样本吸引力评估信息。

[0008] 本发明实施例的第二方面提供了一种区域吸引力评估装置,包括:

[0009] 第一获取单元,用于获取待检测区域的用户访问信息;所述用户访问数组包括多个访问数量元素;所述访问数量元素为所述待检测区域中每个用户对所述待检测区域的访问数量;

[0010] 第一处理单元,用于将所述用户访问信息输入到区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息;其中,在训练过程中,所述区域吸引力评估模型的输入为样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签,所述区域吸引力评估模型的输出为所述样本区域的样本吸引力评估信息。

[0011] 进一步地,区域吸引力评估装置,还包括:

[0012] 第二获取单元,用于获取样本训练集;所述样本训练集包括样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签;

[0013] 训练单元,用于对所述样本训练集进行训练,得到用于输出所述样本区域的样本吸引力评估信息的区域吸引力评估模型。

[0014] 进一步地,所述第二获取单元,包括:

[0015] 第三获取单元,用于获取所述样本用户在通过目标社交应用进行签到时产生的样

本签到数据;所述样本签到数据包括所述样本用户的用户标识、签到地点的地点标识及其位置信息;

[0016] 第二处理单元,用于基于所述样本签到数据得到所述样本区域的样本用户访问信息;

[0017] 第四获取单元,用于获取所述样本区域的吸引力评估信息标签;

[0018] 第三处理单元,用于将每个所述样本区域的样本用户访问信息和每个所述样本区域的吸引力评估信息标签关联存储,得到样本训练集。

[0019] 进一步地,所述第二处理单元,具体用于:

[0020] 对所述样本签到数据聚类,得到样本区域信息;其中,所述样本区域信息包括至少一个样本区域;所述样本区域信息包含在所述样本区域签到的样本用户的用户标识、位于所述样本区域中的签到地点的地点标识及其位置信息;

[0021] 基于所述样本区域信息获取所述样本区域的样本用户访问数组。

[0022] 进一步地,所述第三获取单元,具体用于:

[0023] 获取所述目标社交应用的初始签到数据;所述初始签到数据包括每个样本用户的签到数量;

[0024] 当所述样本用户的签到数量满足预设条件时,获取所述样本用户的签到数据。

[0025] 本发明实施例的第三方面提供了一种区域吸引力评估设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如上述第一方面所述的区域吸引力评估方法的步骤。

[0026] 本发明实施例的第四方面提供了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计算机程序,所述计算机程序被处理器执行时实现如上述第一方面所述的区域吸引力评估方法的步骤。

[0027] 本发明实施例中,获取待检测区域的用户访问信息;将所述用户访问数组输入到区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息。上述方法,将待检测区域的用户访问信息输入到已训练的区域吸引力评估模型进行处理,可以准确的得到区域吸引力评估信息,基于区域吸引力评估信息可以对城市热点区域进行推荐,该模型不同于现有的经验模型,不过多的依赖历史数据,在不同的场景中都可以得到区域吸引力评估信息。

附图说明

[0028] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0029] 图1是本发明第一实施例提供的一种区域吸引力评估方法的示意流程图;

[0030] 图2是本发明第二实施例提供的另一种区域吸引力评估方法的示意流程图;

[0031] 图3是本发明第三实施例提供的另一种区域吸引力评估方法的示意流程图;

[0032] 图4是本发明第三实施例提供的另一种区域吸引力评估方法中S301细化的示意流程图;

[0033] 图5是本发明第三实施例提供的另一种区域吸引力评估方法中S302细化的示意流程图；

[0034] 图6是本发明第四实施例提供的一种区域吸引力评估装置的示意图；

[0035] 图7是本发明第五实施例提供的区域吸引力评估设备的示意图。

具体实施方式

[0036] 以下描述中,为了说明而不是为了限定,提出了诸如特定系统结构、技术之类的具体细节,以便透彻理解本发明实施例。然而,本领域的技术人员应当清楚,在没有这些具体细节的其它实施例中也可以实现本发明。在其它情况中,省略对众所周知的系统、装置、电路以及方法的详细说明,以免不必要的细节妨碍本发明的描述。

[0037] 为了说明本发明所述的技术方案,下面通过具体实施例来进行说明。

[0038] 请参见图1,图1是本发明第一实施例提供的一种区域吸引力评估方法的示意流程图。本实施例中区域吸引力评估方法的执行主体为具有区域吸引力评估功能的设备,例如,区域吸引力评估服务器。如图1所示的区域吸引力评估方法可包括:

[0039] S101:获取待检测区域的用户访问信息;所述用户访问信息包括多个访问数量元素;所述访问数量元素为所述待检测区域中每个用户对所述待检测区域的访问数量。

[0040] 设备获取待检测区域的用户访问信息,其中,待检测区域是根据实际情况划定的,例如,可以划定某一商圈为待检测区域,也可以划定某一商场中的某个区域为待检测区域。用户访问信息可以是一个数组,包括多个访问数量元素,访问数量元素为待检测区域中每个用户对所述待检测区域的访问数量,其中,待检测区域的用户访问数组可以记为 $visit(r_i)=\{uv_1^i, uv_2^i, \dots, uv_n^i\}$,其中 uv_n^i 为第n个用户对 r_i 区域的访问数量。

[0041] S102:将所述用户访问信息输入到区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息;其中,在训练过程中,所述区域吸引力评估模型的输入为样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签,所述区域吸引力评估模型的输出为所述样本区域的样本吸引力评估信息。

[0042] 设备中预先存储预先训练的区域吸引力评估模型,其中,区域吸引力评估模型可以由设备预先训练,也可以由其他设备预先训练。预先训练的区域吸引力评估模型是通过使用机器学习算法对样本训练集进行训练得到,区域吸引力评估模型的输入为样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签,区域吸引力评估模型的输出为样本区域的样本吸引力评估信息。

[0043] 其中,预先训练的区域吸引力评估模型可以包括输入层、隐藏层、损失函数层(输出层)。隐藏层包括两个以上的隐含层节点,输入层包括一个输入层节点,用于从外部接收输入的样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签。隐含层用于对样本用户访问信息进行处理。输出层用于输出样本区域的样本吸引力评估信息。

[0044] 设备将用户访问信息输入预先训练的区域吸引力评估模型,得到待检测区域的吸引力评估信息。

[0045] 本发明实施例中,获取待检测区域的用户访问信息;将所述用户访问信息输入到区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息。上述方法,将待检

测区域的用户访问信息输入到已训练的区域吸引力评估模型进行处理,可以准确的得到区域吸引力评估信息,基于区域吸引力评估信息可以对城市热点区域进行推荐,该模型不同于现有的经验模型,不过多的依赖历史数据,在不同的场景中都可以得到区域吸引力评估信息。

[0046] 请参见图2,图2是本发明第二实施例提供的另一种区域吸引力评估方法的示意流程图。本实施例中区域吸引力评估方法的执行主体为具有区域吸引力评估功能的设备,例如,区域吸引力评估服务器。为了获取更准确的区域吸引力评估信息,本实施例中限定了区域吸引力评估模型的训练过程,本实施例与第一实施例的区别在于S201~S202,其中,本实施例中的S203~S204与第一实施例中的 S101~S102相同,S201~S202在S203之前执行。如图2所示,S201~S202具体如下:

[0047] S201:获取样本训练集;所述样本训练集包括样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签。

[0048] 设备获取样本训练集,其中,样本训练集包括样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签,每个样本用户访问信息对应一个样本吸引力评估信息标签。样本训练集中的样本的数量,可以根据实际情况进行设置,在一定程度上来说,训练的样本数量越多,使用训练得到的神经网络模型进行识别时,识别出的结果越准确。

[0049] S202:对所述样本训练集进行训练,得到用于输出所述样本区域的样本吸引力评估信息的区域吸引力评估模型。

[0050] 本实施例中,区域吸引力评估模型由设备预先训练。设备对样本训练集进行训练,在训练过程中,将样本区域的样本用户访问数组及其对应的样本吸引力评估信息为训练数据,将其输入初始的神经网络模型,通过调整区域吸引力评估模型的损失函数不断的完善模型,从而得到最终的区域吸引力评估模型。

[0051] 请参见图3,图3是本发明第三实施例提供的另一种区域吸引力评估方法的示意流程图。本实施例中区域吸引力评估方法的执行主体为具有区域吸引力评估功能的设备,例如,区域吸引力评估服务器。为了获取到高质量的样本进行神经网络模型的训练,从而获取到准确的区域吸引力评估信息,本实施例与第二实施例的区别在于S301~S304,其中,本实施例中的S305~S307与第二实施例中的 S202~S204相同,S301~S304是第二实施例中的S201的进一步细化。如图3所示,S301~S304具体如下:

[0052] S301:获取所述样本用户在通过目标社交应用进行签到时产生的样本签到数据;所述样本签到数据包括所述样本用户的用户标识、签到地点的地点标识及其位置信息。

[0053] 社交媒体平台大、更新速度快、内容多样,所以社交媒体数据也是十分的庞大且具有价值的,当前很多社交应用都具有签到功能,本实施例中,社交应用中的签到是指社交应用中的位置签到,位置签到服务是位置服务在移动互联网的延伸和发展。位置服务指通过电信运营商的移动通信网络,采用全球定位系统 (Global Positioning System,GPS)、基站等相关定位技术,结合地理信息系统 (Geographic Information System,GIS) 地理信息系统,通过手机终端确定用户实际地理位置信息,以短信、彩信、语音、网页及客户端软件等方式为用户提供的地理位置信息服务,举例来说,用户可以通过定位功能获取当前所在地点,并且基于当前所在地点进行签到操作或者进行分享操作。当前常用的社交应用包括微信、微博、QQ空间等等,本实施例中的目标社交应用可以包括但是不限于上述常用的社交应

用。

[0054] 设备获取样本用户通过目标社交应用进行签到产生的签到数据,其中,目标社交应用为用户群体比较大、数据比较丰富的、且具有位置签到功能的社交媒体应用。样本用户可以通过在目标社交应用上发布动态进行签到,也可以通过直接点击“位置签到”的虚拟按钮进行签到,样本用户通过目标社交应用进行签到产生的签到数据包括用户的用户标识、签到地点的地点标识及其位置信息。其中,样本用户的用户标识为用于识别用户身份的唯一标识,签到地点的地点标识及其位置信息为签到地点的唯一标识及其对应的位置信息,位置信息可以是坐标的方式,也可以是经纬度的方式,此处不做限制。

[0055] 进一步地,为了提升签到数据的质量,从而得到高质量的样本信息,S301 可以包括S3011~S3012,如图4所示,S3011~S3012具体如下:

[0056] S3011:获取所述目标社交应用的初始签到数据;所述初始签到数据包括每个样本用户的签到数量。

[0057] 设备获取目标社交应用的初始签到数据,初始签到数据为未经过任何筛选和处理过的签到数据,初始签到数据包括每个样本用户的签到数量,即每个样本用户在目标社交应用上的所有签到操作的次数。

[0058] S3012:当所述样本用户的签到数量满足预设条件时,获取所述样本用户的签到数据。

[0059] 为了提升签到数据的质量,需要筛选出有价值的签到数据,即签到数量多的样本用户的签到数据是更具有价值的,更能体现一个区域的吸引力。设备预设了条件用于筛选符合预设条件的样本用户,预设当样本用户的签到数量满足预设条件时,符合预设条件的样本用户即为目标社交应用中的活跃用户,获取活跃用户的签到数据,得到高质量的签到数据。举例来说,可以设每个样本用户签到数量为变量 x_i ,计算每个样本用户发的平均签到数量 μ ,以及标准差 σ ,然后利用三倍标准差原则,将样本用户发文数量在三倍标准差以外的用户去除,即所剩用户满足 $x_i \in [\mu - 3\sigma, \mu + 3\sigma]$,获取剩余样本用户的签到数据。

[0060] S302:基于所述样本签到数据得到所述样本区域的样本用户访问信息。

[0061] 设备基于样本签到数据得到样本区域的样本用户访问信息,样本用户访问数组包括多个样本访问数量元素,样本访问数量元素为待检测区域中每个样本用户对所述待检测区域的访问数量。一种实施方式中,预设了样本区域,然后对样本签到数据进行统计,获取到样本区域中每个样本用户对样本区域的访问数量;另一种实施方式中,可以对样本签到数据进行聚类,获取样本区域,再基于样本签到数据获取样本区域的样本用户访问信息。

[0062] 进一步地,为了获取高质量的样本训练集,从而训练出结果准确的区域吸引力评估模型,S302可以包括S3021~S3022,如图5所示,S3021~S3022具体如下:

[0063] S3021:对所述样本签到数据聚类,得到样本区域信息;其中,所述样本区域信息包括至少一个样本区域;所述样本区域信息包含在所述样本区域签到的样本用户的用户标识、位于所述样本区域中的签到地点的地点标识及其位置信息。

[0064] 将物理或抽象对象的集合分成由类似的对象组成的多个类的过程被称为聚类。由聚类所生成的簇是一组数据对象的集合,这些对象与同一个簇中的对象彼此相似,与其他簇中的对象相异。设备对样本签到数据聚类,聚类分析计算方法主要有以下几种:

[0065] 1、划分方法(partitioning methods),即给定一个有N个元组或者纪录的数据集,

分裂法将构造 K 个分组,每一个分组就代表一个聚类, $K < N$ 。大部分划分方法是基于距离的。给定要构建的分区数 k ,划分方法首先创建一个初始化划分。然后,它采用一种迭代的重新定位技术,通过把对象从一个组移动到另一个组来进行划分。一个好的划分的一般准备是:同一个簇中的对象尽可能相互接近或相关,而不同的簇中的对象尽可能远离或不同。还有许多评判划分质量的其他准则。传统的划分方法可以扩展到子空间聚类,而不是搜索整个数据空间。当存在很多属性并且数据稀疏时,这是有用的。为了达到全局最优,基于划分的聚类可能需要穷举所有可能的划分,计算量极大。

[0066] 2、层次方法(hierarchical methods),这种方法对给定的数据集进行层次似的分解,直到某种条件满足为止。具体又可分为“自底向上”和“自顶向下”两种方案。例如在“自底向上”方案中,初始时每一个数据纪录都组成一个单独的组,在接下来的迭代中,它把那些相互邻近的组合成一个组,直到所有的记录组成一个分组或者某个条件满足为止。代表算法有变色龙(CHAMELEON)算法等。层次聚类方法可以是基于距离的或基于密度或连通性的。层次聚类方法的一些扩展也考虑了子空间聚类。

[0067] 3、基于密度的方法(density-based methods),基于密度的方法与其它方法的一个根本区别是:它不是基于各种各样的距离的,而是基于密度的。这样就能克服基于距离的算法只能发现“类圆形”的聚类的缺点。这个方法的指导思想就是,只要一个区域中的点的密度大过某个阈值,就把它加到与之相近的聚类中去。

[0068] 4、基于网格的方法(grid-based methods),这种方法首先将数据空间划分成为有限个单元(cell)的网格结构,所有的处理都是以单个的单元为对象的。这么处理的一个突出的优点就是处理速度很快,通常这是与目标数据库中记录的个数无关的,它只与把数据空间分为多少个单元有关。

[0069] 5、基于模型的方法(model-based methods),基于模型的方法给每一个聚类假定一个模型,然后去寻找能够很好的满足这个模型的数据集。这样一个模型可能是数据点在空间中的密度分布函数或者其它。它的一个潜在的假定就是:目标数据集是由一系列的分布所决定的。通常有两种尝试方向:统计的方案和神经网络的方案。

[0070] 设备对签到数据聚类,得到样本区域信息。在本实施例中,使用预设的密度聚类算法对签到数据聚类。密度聚类算法(Density-Based Spatial Clustering of Applications with Noise,DBSCAN)是一个比较有代表性的基于密度的聚类算法。预设采用DBSCAN算法对签到数据聚类,与划分和层次聚类方法不同,它将簇定义为密度相连的点的最大集合,能够把具有足够高密度的区域划分为簇,并可在噪声的空间数据库中发现任意形状的聚类。

[0071] 对于DBSCAN算法,共有两个参数需要确定,一个参数为每个类的半径(Eps),即兴趣区域的半径;一个参数为每个类中最少的签到点数(MinPts),即兴趣区域包含的签到地点的数量。

[0072] 其中,可以通过 k -dist图确定参数,以签到地点 p 为例, k -dist即为计算 p 与距离 p 最近的第 k 个点之间的距离。 k -dist图的构造方法为,为每一个用户的签到数据计算出其 k -dist值,然后将其按照 k -dist从大到小降序排列,排名第一的为 k -dist值最大的签到点,排名最后的为 k -dist最小的点。然后按此顺序,将其画到二维坐标系内,其中横坐标为按照 k -dist的降序排列的签到地点,纵坐标为 k -dist值。即 k -dist满足这样的条件: p 周围最近的 k

个点,都在p的k-dist 距离范围内。对k-dist图进行可视化分析,按照降序列出各签到地点的k-dist值,找到k-dist图的明显的谷底,并将谷底对应的阈值作为DBSCAN的Eps;之后,再使用如下公式计算MinPts:

$$[0073] \quad MinPts = \frac{1}{n} \sum_{i=1}^n p_i$$

[0074] 其中, p_i 为对应签到地点周围Eps距离内的签到地点的数量。区域吸引力评估设备基于签到数据、兴趣区域的半径和兴趣区域包含的签到地点的数量,采用预设的密度聚类算法对签到数据聚类,可以任选一个未被访问(unvisited)的签到地点开始,找出与其距离在Eps之内(包括Eps)的所有附近签到地点。如果附近签到地点的数量 \geq MinPts,则当前点与其附近点形成一个簇,并且出发点被标记为已访问(visited)。然后递归,以相同的方法处理该簇内所有未被标记为已访问(visited)的签到地点,从而对簇进行扩展。如果附近点的数量 $<$ MinPts,则该签到地点暂时被标记作为噪声点。如果簇充分地扩展,即簇内的所有签到地点被标记为已访问,然后用同样的算法去处理未被访问的签到地点,最后得到样本区域信息。

[0075] 其中,样本区域信息包括至少一个样本区域,样本区域为基于签到数据划分得到的区域;样本区域信息包含样本区域签到的样本用户的用户标识、位于样本区域中的签到地点的地点标识及其位置信息。

[0076] S3022:基于所述样本区域信息获取所述样本区域的样本用户访问信息。

[0077] 样本用户访问数组包括多个样本访问数量元素,样本访问数量元素为待检测区域中每个样本用户对待检测区域的访问数量,由于样本区域信息包含样本区域签到的样本用户的用户标识、位于样本区域中的签到地点的地点标识及其位置信息。所以,基于样本区域信息进行统计,就可以得到每个样本用户对待检测区域的访问数量,从而获取到样本区域的样本用户访问数组。

[0078] S303:获取所述样本区域的吸引力评估信息标签。

[0079] 设备中可以预先设定样本区域的吸引力评估信息标签,其中,样本区域的吸引力评估信息标签可以根据用户对该区域的历史吸引力打分得到,获取多个用户对该区域的历史吸引力打分,对历史吸引力打分进行加权计算,得到该样本区域的吸引力评估信息标签。

[0080] S304:将每个所述样本区域的样本用户访问信息和每个所述样本区域的吸引力评估信息标签关联存储,得到样本训练集。

[0081] 设备将每个样本区域的样本用户访问信息和其对应的样本区域的吸引力评估信息标签关联存储,一组为一个样本,得到样本训练集。

[0082] 应理解,上述实施例中各步骤的序号的大小并不意味着执行顺序的先后,各过程的执行顺序应以其功能和内在逻辑确定,而不应对本发明实施例的实施过程构成任何限定。

[0083] 请参见图6,图6是本发明第四实施例提供的一种区域吸引力评估装置的示意图。包括的各单元用于执行图1~图5对应的实施例中的各步骤,具体请参阅图 1~图5各自对应的实施例中的相关描述。为了便于说明,仅示出了与本实施例相关的部分。参见图6,区域吸引力评估装置6包括:

[0084] 第一获取单元610,用于获取待检测区域的用户访问信息;所述用户访问信息包括多个访问数量元素;所述访问数量元素为所述待检测区域中每个用户对所述待检测区域的访问数量;

[0085] 第一处理单元620,用于将所述用户访问信息输入到区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息;其中,在训练过程中,所述区域吸引力评估模型的输入为样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签,所述区域吸引力评估模型的输出为所述样本区域的样本吸引力评估信息。

[0086] 进一步地,区域吸引力评估装置6,还包括:

[0087] 第二获取单元,用于获取样本训练集;所述样本训练集包括样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签;

[0088] 训练单元,用于对所述样本训练集进行训练,得到用于输出所述样本区域的样本吸引力评估信息的区域吸引力评估模型。

[0089] 进一步地,所述第二获取单元,包括:

[0090] 第三获取单元,用于获取所述样本用户在通过目标社交应用进行签到时产生的样本签到数据;所述样本签到数据包括所述样本用户的用户标识、签到地点的地点标识及其位置信息;

[0091] 第二处理单元,用于基于所述样本签到数据得到所述样本区域的样本用户访问信息;

[0092] 第四获取单元,用于获取所述样本区域的吸引力评估信息标签;

[0093] 第三处理单元,用于将每个所述样本区域的样本用户访问信息和每个所述样本区域的吸引力评估信息标签关联存储,得到样本训练集。

[0094] 进一步地,所述第二处理单元,具体用于:

[0095] 对所述样本签到数据聚类,得到样本区域信息;其中,所述样本区域信息包括至少一个样本区域;所述样本区域信息包含在所述样本区域签到的样本用户的用户标识、位于所述样本区域中的签到地点的地点标识及其位置信息;

[0096] 基于所述样本区域信息获取所述样本区域的样本用户访问信息。

[0097] 进一步地,所述第三获取单元,具体用于:

[0098] 获取所述目标社交应用的初始签到数据;所述初始签到数据包括每个样本用户的签到数量;

[0099] 当所述样本用户的签到数量满足预设条件时,获取所述样本用户的签到数据。

[0100] 请参见图7,图7是本发明第五实施例提供的区域吸引力评估设备的示意图。如图7所示,该实施例的区域吸引力评估设备7包括:处理器70、存储器71以及存储在所述存储器71中并可在所述处理器70上运行的计算机程序72,例如区域吸引力评估程序。所述处理器70执行所述计算机程序72时实现上述各个区域吸引力评估方法实施例中的步骤,例如图1所示的步骤101至102。或者,所述处理器70执行所述计算机程序72时实现上述各装置实施例中各模块/单元的功能,例如图所示模块610至620的功能。

[0101] 示例性的,所述计算机程序72可以被分割成一个或多个模块/单元,所述一个或者多个模块/单元被存储在所述存储器71中,并由所述处理器70执行,以完成本发明。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机程序指令段,该指令段用于

描述所述计算机程序72在所述区域吸引力评估设备7中的执行过程。例如,所述计算机程序72可以被分割成第一获取单元、第一处理单元,各单元具体功能如下:

[0102] 第一获取单元,用于获取待检测区域的用户访问信息;所述用户访问信息包括多个访问数量元素;所述访问数量元素为所述待检测区域中每个用户对所述待检测区域的访问数量;

[0103] 第一处理单元,用于将所述用户访问信息输入到区域吸引力评估模型进行处理,获取所述待检测区域的吸引力评估信息;其中,在训练过程中,所述区域吸引力评估模型的输入为样本区域的样本用户访问信息及其对应的样本吸引力评估信息标签,所述区域吸引力评估模型的输出为所述样本区域的样本吸引力评估信息。

[0104] 所述区域吸引力评估设备可包括,但不仅限于,处理器70、存储器71。本领域技术人员可以理解,图7仅仅是区域吸引力评估设备10的示例,并不构成对区域吸引力评估设备7的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件,例如所述区域吸引力评估设备还可以包括输入输出设备、网络接入设备、总线等。

[0105] 所称处理器70可以是中央处理单元(Central Processing Unit,CPU),还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0106] 所述存储器71可以是所述区域吸引力评估设备7的内部存储单元,例如区域吸引力评估设备7的硬盘或内存。所述存储器71也可以是所述区域吸引力评估设备7的外部存储设备,例如所述区域吸引力评估设备7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。进一步地,所述存储器71还可以既包括所述区域吸引力评估设备7的内部存储单元也包括外部存储设备。所述存储器71用于存储所述计算机程序以及所述区域吸引力评估设备所需的其他程序和数据。所述存储器71还可以用于暂时地存储已经输出或者将要输出的数据。

[0107] 所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,仅以上述各功能单元、模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能单元、模块完成,即将所述装置的内部结构划分成不同的功能单元或模块,以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中,上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。另外,各功能单元、模块的具体名称也只是为了便于相互区分,并不用于限制本申请的保护范围。上述系统中单元、模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0108] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述或记载的部分,可以参见其它实施例的相关描述。

[0109] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员

可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0110] 在本发明所提供的实施例中,应该理解到,所揭露的装置/终端设备和方法,可以通过其它的方式实现。例如,以上所描述的装置/终端设备实施例仅仅是示意性的,例如,所述模块或单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通讯连接可以通过一些接口,装置或单元的间接耦合或通讯连接,可以是电性,机械或其它的形式。

[0111] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0112] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0113] 所述集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明实现上述实施例方法中的全部或部分流程,也可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一计算机可读存储介质中,该计算机程序在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机程序包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、电载波信号、电信信号以及软件分发介质等。需要说明的是,所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减,例如在某些司法管辖区,根据立法和专利实践,计算机可读介质不包括电载波信号和电信信号。

[0114] 以上所述实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围,均应包含在本发明的保护范围之内。

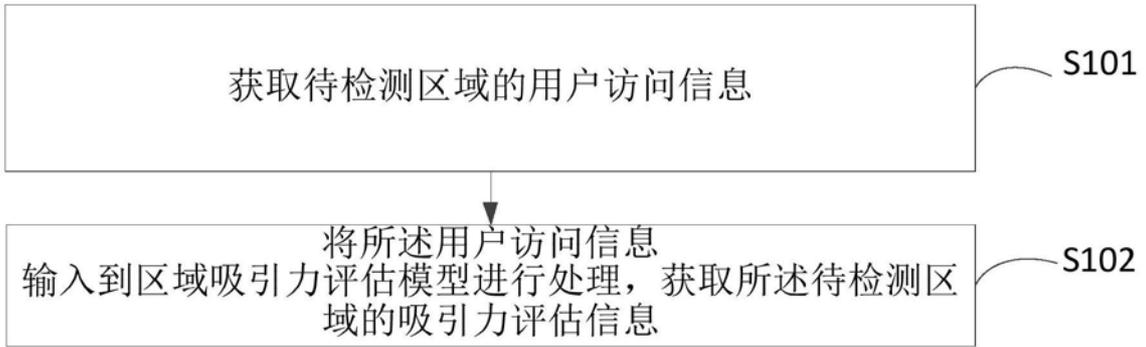


图1

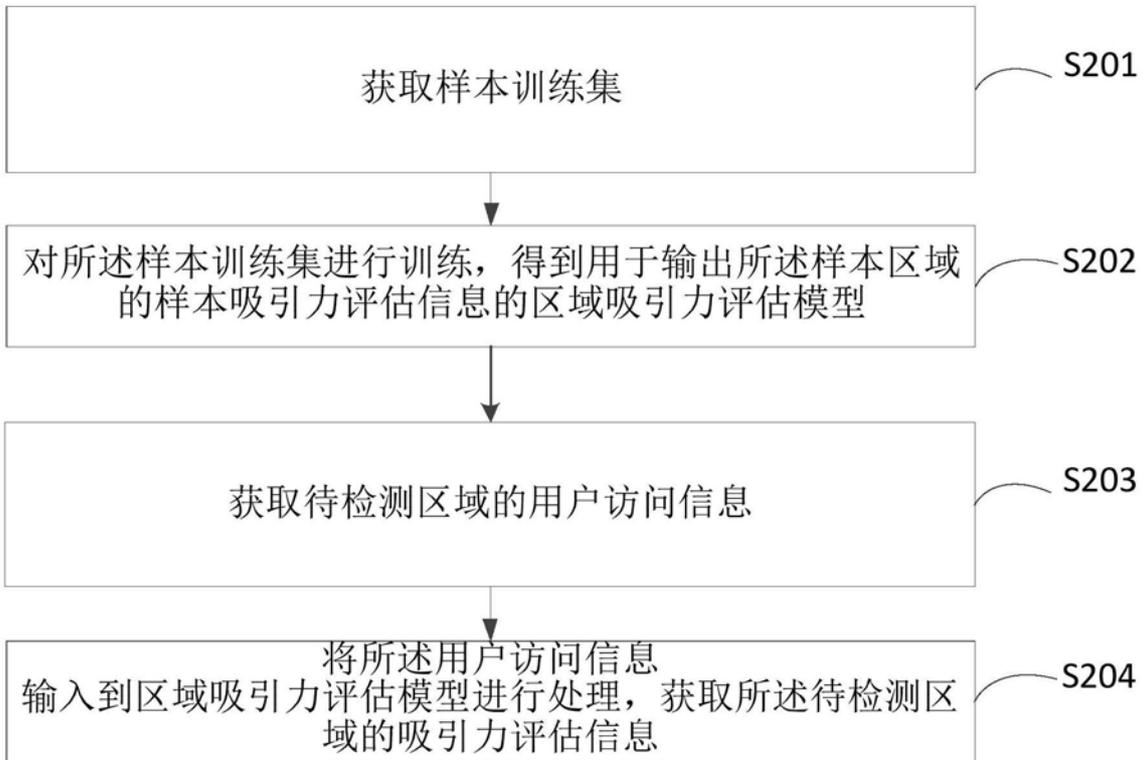


图2

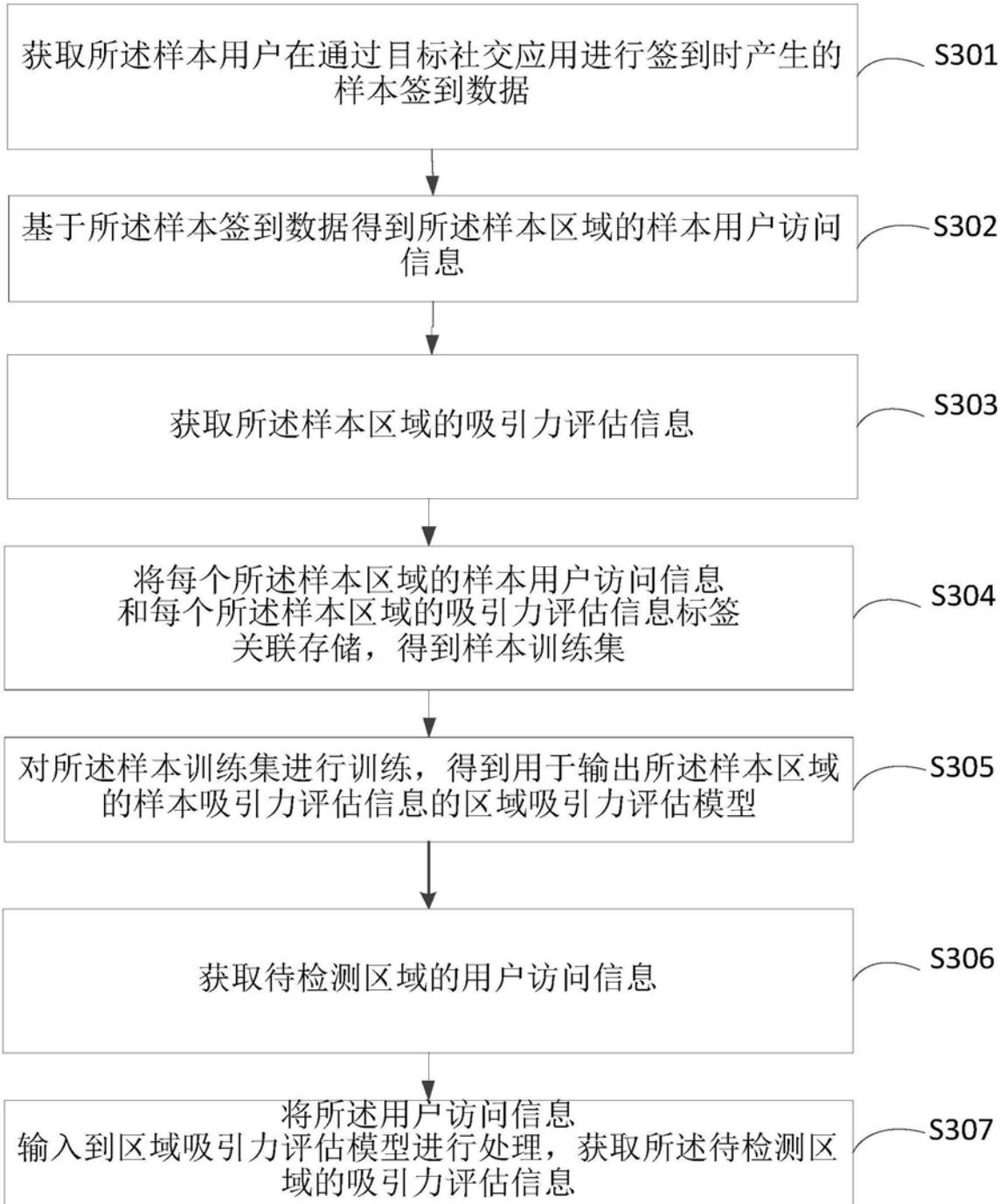


图3

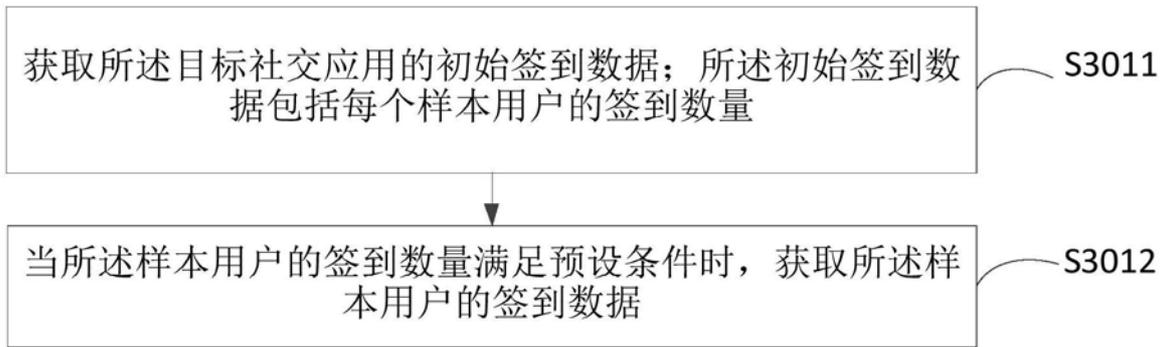


图4

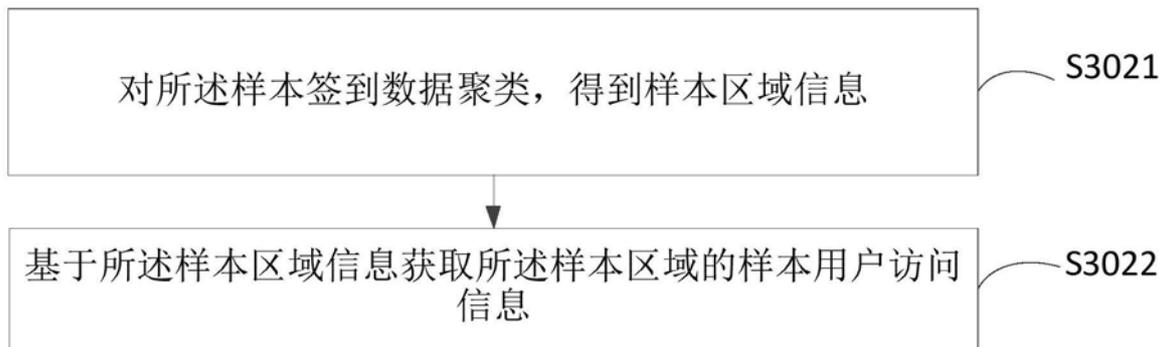


图5

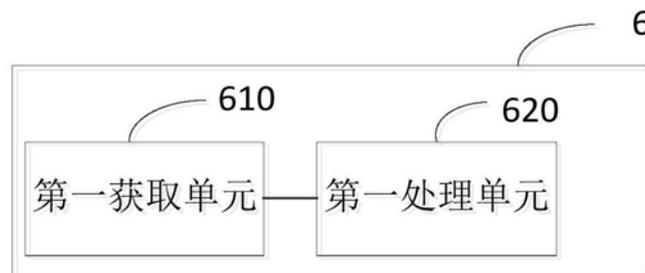


图6

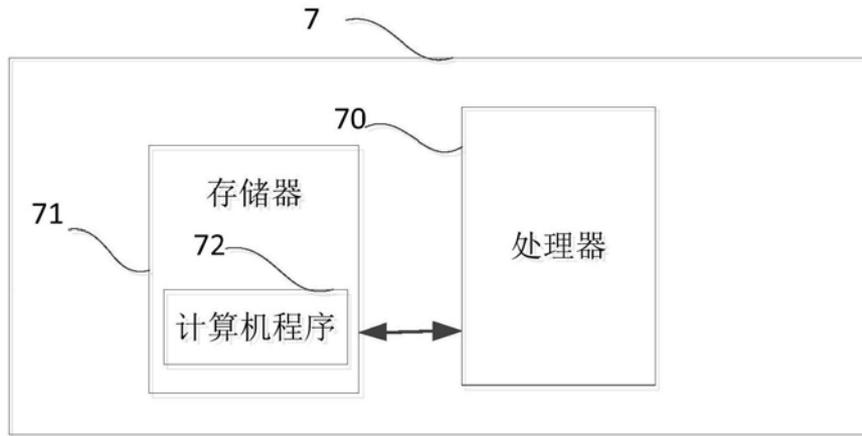


图7