



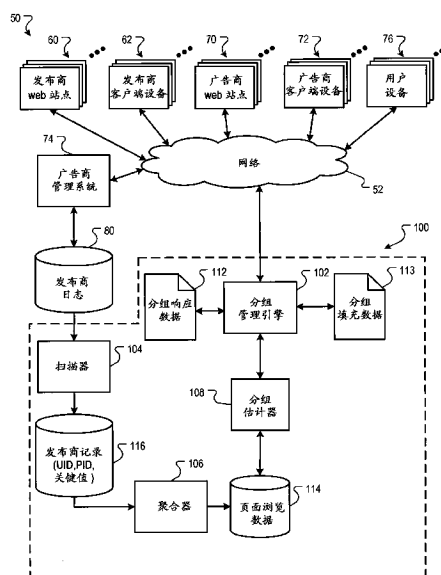
(45)授权公告日 2018.03.16

审查员 毛冉

权利要求书2页 说明书12页 附图5页

用于受众分组估计的方法和系统

本发明提供了一种计算机实现的方法和系统,该方法包括:在数据处理装置处接收用于发布商站点的分组定义,该分组定义由该发布商的一个或多个发布商键值来定义分组;由该数据处理装置向该分组指派具有满足由该分组定义所定义的该分组的关联页面浏览的用户标识符;对于该用户标识符确定页面浏览的第一数目,其等于与满足该分组定义的发布商键值关联的页面浏览的数目;对于该用户标识符确定页面浏览的第二数目,其等于与向该分组指派的该用户标识符关联的页面浏览的总数目,其中该第二数目大于该第一数目;确定和该第二数目与该第一数目之比成比例的该分组的分组质量测量;以及响应于该分组定义,提供描述该分组质量测量的响应数据。



1. 一种用于受众分组估计的计算机实现的方法,包括:

在数据处理装置处接收用于发布商站点的分组定义,所述分组定义由所述发布商的一个或多个发布商键值来定义分组;

由所述数据处理装置向所述分组指派具有满足由所述分组定义所定义的所述分组的关联页面浏览的用户标识符;

对于向所述分组指派的所述用户标识符确定页面浏览的第一数目,所述页面浏览的第一数目等于与满足所述分组定义的发布商键值关联的页面浏览的数目;

对于向所述分组指派的所述用户标识符确定页面浏览的第二数目,所述页面浏览的第二数目等于与向所述分组指派的用户标识符关联的页面浏览的总数目,其中所述第二数目大于所述第一数目;

确定和所述页面浏览的第二数目与所述页面浏览的第一数目之比成比例的所述分组的分组质量测量;以及

响应于所述分组定义,提供描述所述分组的所述分组质量测量的响应数据。

2. 根据权利要求1所述的计算机实现的方法,其中确定所述页面浏览的第一数目包括确定指示满足所述分组定义的发布商键值的页面浏览/键值映射的数目。

3. 根据权利要求1所述的计算机实现的方法,还包括:

生成分组填充数据,所述分组填充数据是向由所述分组定义数据所定义的所述分组指派的用户标识符和定义所述分组的数据的关联;以及

响应于所述分组定义,向发布商提供所述分组填充数据,以促进广告将与向所述分组指派的所述用户标识符关联的未来页面浏览作为目标。

4. 一种用于受众分组估计的系统,包括:

数据处理装置;以及

存储器储存系统,与所述数据处理装置以数据通信并且存储由所述数据处理装置可执行的指令,并且在这样执行时使所述数据处理装置执行如下操作,所述操作包括:

在数据处理装置处接收用于发布商站点的分组定义,所述分组定义由所述发布商的一个或多个发布商键值来定义分组;

由所述数据处理装置向所述分组指派具有满足由所述分组定义所定义的所述分组的关联页面浏览的用户标识符;

对于向所述分组指派的所述用户标识符确定页面浏览的第一数目,所述页面浏览的第一数目等于与满足所述分组定义的发布商键值关联的页面浏览的数目;

对于向所述分组指派的所述用户标识符确定页面浏览的第二数目,所述页面浏览的第二数目等于与向所述分组指派的用户标识符关联的页面浏览的总数目,其中所述第二数目大于所述第一数目;

确定和所述页面浏览的第二数目与所述页面浏览的第一数目之比成比例的所述分组的分组质量测量;以及

响应于所述分组定义,提供描述所述分组的所述分组质量测量的响应数据。

5. 根据权利要求4所述的系统,其中确定所述页面浏览的第一数目的所述操作包括确定指示满足所述分组定义的发布商键值的页面浏览/键值映射的数目。

6. 根据权利要求4所述的系统,其中所述操作还包括:

生成分组填充数据,所述分组填充数据是向由所述分组定义数据所定义的所述分组指派的用户标识符和定义所述分组的数据的关联;以及

响应于所述分组定义,向发布者提供所述分组填充数据,以促进广告将与向所述分组指派的所述用户标识符关联的未来页面浏览作为目标。

用于受众分组估计的方法和系统

[0001] 本申请是2012年4月11日提交的申请号为201080045960.X的发明名称为“受众分组估计”的发明专利申请的分案申请。

背景技术

[0002] 本说明书涉及在线广告。

[0003] 互联网提供了对各类资源的访问,例如视频和/或音频文件以及特定主题或特定新闻文章的web页面。访问这些资源提供了随这些资源提供广告的机会。例如,web页面可包括在其中呈现广告的广告位置(slot)。广告位置可以定义在web页面上,或者定义成与web页面一同呈现。

[0004] 通常,发布商和/或广告商希望得到描述广告或广告活动的受众分组(segment)的信息。获得该信息可有助于将分组作为目标和/或将行为作为目标。例如,假定到访某发布商站点的用户分组定义成对i) 烹饪和ii) 编织这两者感兴趣的用户。如果可以获得描述分组的准确信息,那么发布商和/或广告商可确定广告是仅以符合该分组的页面浏览(pageview)为目标,还是以属于该分组的所有用户为目标,或者完全不以该分组为目标。

[0005] 然而,获得准确分组信息可能十分耗时,这是由于通常需要处理数百万条日志记录。

发明内容

[0006] 总的来说,本说明书所述主题的一方面可以体现在包括如下动作的方法中:由数据处理装置访问发布商日志,发布商日志包括描述针对发布商站点的页面浏览所递送的过往印象的数据,针对每个印象,该数据包括:用户标识符,识别发布商的发布商标识符,识别页面浏览的页面浏览值,以及为印象提供的发布商的多个发布商键值;由数据处理装置根据发布商日志生成每个发布商的发布商页面浏览数据,每个发布商的发布商页面浏览数据包括代表在第一时间段期间出现的页面浏览的页面浏览记录,每个页面浏览记录代表在第一时间段期间出现的第二时间段期间的和用户标识符及发布商标识符对相关的一个或多个页面浏览,每个页面浏览记录包括:用户标识符,发布商标识符以及提供给由页面浏览记录代表的针对一个或多个页面浏览所出现的印象的发布商键值;在数据处理装置接收发布商的分组定义,分组定义根据发布商的一个或多个发布商键值定义分组;由数据处理装置根据包括满足分组定义所定义的分组的发布商键值的记录,为分组指派用户标识符;由数据处理装置基于指派给分组的用户标识符和指派给分组的用户标识符的页面浏览记录,确定一个或多个分组特征;以及,响应于分组定义,提供描述分组的一个或多个特征的响应数据。该方面的其他实施例包括配置用于执行在计算机存储设备上编码的所述方法的动作的对应系统、装置和计算机程序。

[0007] 本说明书所述主题的另一创新方面可以体现在包括如下动作的方法中:由数据处理装置生成多个发布商的发布商页面浏览数据,每个发布商的发布商页面浏览数据包括代表在第一时间段期间出现的页面浏览的页面浏览记录,每个页面浏览记录代表与用户标识

符及发布商标识符对相关的一个或多个页面浏览,每个页面浏览记录包括:用户标识符、识别发布商的发布商标识符、以及为由页面浏览记录代表的针对一个或多个页面浏览所出现的印象提供的发布商键值;在数据处理装置接收发布商的分组定义,分组定义根据发布商的一个或多个发布商键值定义分组;由数据处理装置确定满足分组定义所定义的分组的发布商键值被提供给印象的页面浏览;由数据处理装置基于被确定满足分组定义的页面浏览确定一个或多个分组特征;以及,响应于分组定义,提供描述分组的一个或多个特征的响应数据。本方面的其他实施例包括配置用于执行在计算机存储设备上编码的所述方法的动作的对应系统、装置和计算机程序。

[0008] 下文的附图和描述阐述了本说明书所述主题的一个或多个实施例的具体内容。根据描述、附图以及权利要求,主题的其他特征、方面和益处变得明显。

附图说明

[0009] 图1是可以使用分组估计系统的环境的框图。

[0010] 图2是示例页面浏览记录的框图。

[0011] 图3是确定分组的一个或多个特征的示例过程的流程图。

[0012] 图4是确定分组质量测量的示例过程的流程图。

[0013] 图5是确定满足分组定义的页面浏览的示例过程的流程图。

[0014] 图6是确定分组的一个或多个特征的另一示例过程的流程图。

[0015] 各附图中的类似参考标号和标记指示类似的元素。

具体实施方式

[0016] 总体而言,本说明书的主题涉及估计受众分组大小以及填充 (populate) 受众分组。受众分组代表具有人口统计学特征的预期用户的聚合组,人口统计学特征例如为行为模式特征。此处所述为估计web站点的受众分组大小以及受众分组质量的分组估计系统。该分组估计系统使用定义web站点分组的分组定义数据并根据页面浏览和印象数据确定属于分组的用户数。比较分组内的页面浏览数和分组内以及分组外的页面浏览数以确定分组质量。根据分组大小和质量,发布商可做出不同的目标决策。例如,如果分组大小大但是质量不高,则发布商可选择仅以分组内的用户为目标,即,以属于分组的且正在浏览属于分组的页面的用户为目标。如果分组大小大而且质量高,则发布商可选择以分组内和组成部分外的用户为目标,即,无论属于分组的用户是否正在浏览属于分组的页面都以该用户为目标。最后,如果分组大小小,则发布商可选择不以分组为目标。

[0017] §1.0 示例操作环境

[0018] 图1是可在其中使用分组估计系统100的环境50的框图。环境50包括:计算机网络52(如局域网(LAN)、广域网(WAN)、互联网或其组合)、连接的发布商web站点60、发布商客户端设备62、广告商web站点70、广告商客户端设备72、广告商管理系统74、用户设备76和分组估计系统100。

[0019] 每个web站点60是和域名关联的一个或多个web页面资源,而且,每个web站点由一个或多个服务器托管。示例web站点是可包括文本、图形图像、多媒体内容和如脚本的编程元素的超文本标记语言(HTML)格式的web页面的集合。每个web站点60由例如管理和/或拥

有web站点的实体的发布商维护。简洁起见,术语“发布商”也用于指代由发布商管理和/或拥有的web站点60。类似的web站点70由对应的广告商维护,术语“广告商”也用于指代由广告商管理和/或拥有的web站点70。

[0020] 发布商客户端设备62、广告商客户端设备72和用户客户端设备76是受用户控制并且可通过网络52请求并接收数据的电子设备。客户端设备通常包括如web浏览器的用户应用,以促进通过网络52发送并接收数据,例如从发布商60或广告商70请求资源(例如页面内容)。示例客户端设备包括个人计算机、移动通信设备和可通过网络52发送和接收数据的其他设备。

[0021] 广告管理系统74可为发布商60的web页面提供广告商70的广告。例如,发布商60可把对一个或多个广告的广告请求提供给广告管理系统74。广告管理系统74通过把广告发送给请求的发布商60以放在发布商web页面上做出响应,web页面连同广告在用户客户端设备76上呈递时产生广告的印象。广告可包括到登陆(landing)页面的嵌入的链接,登陆页面例如广告商70网站上的页面,用户点击发布商web页面上显示的广告时会被导向到该登陆页面。

[0022] 发布商日志80存储描述针对特定发布商站点的页面浏览所递送的过往印象的数据记录。页面浏览代表载入互联网站点单个页面的请求。对于每个过往印象,数据包括用户标识符、发布商标识符、页面浏览值,以及为特定印象提供的发布商的发布商键值(key value)。用户标识符和发布商标识符每一个分别识别用户和发布商。页面浏览值识别出现印象的特定发布商站点的web页面的特定页面浏览实例。键值通常是包括发布商标识符、键以及和特定印象相关的值的元组。键值可以由发布商创建或指定,而且可包括关键字、数字、词或者短语以及可由广告管理系统74和/或每个对应的发布商60跟踪的如人口统计学或目标数据的描述属性。这类属性数据的示例包括上下文数据,即由发布商的印象标签定义的数据,例如兴趣、关键字和爱好;用户性别、年龄、收入水平和教育水平;请求web页面的用户或客户端设备的位置(例如邮编、城市和/或国家);以及可由广告管理系统74和/或发布商60跟踪的其它信息。发布商日志80还可存储另外的信息,如印象的时间和日期。

[0023] 一些实现方式中,为了保护用户隐私,广告管理系统对用户的印象数据进行匿名处理,这样,日志80中存储的数据不能和用户关联起来。例如,用户身份可被模糊处理或者设置成不和用户关联的唯一数字;用户地址(如果知道的话)可以被模糊处理成仅仅是邮政服务区,例如邮编。日志80也可被加密,从而在出现未经授权的系统访问时进一步保护用户信息。也可实现保护用户隐私的其它技术。

[0024] §2.0分组估计系统

[0025] 分组估计系统100可使用日志80中采集的数据估计站点60的受众分组大小。例如,系统100生成的估计可通过允许站点60的发布商在实际创建受众分组之前查看分组大小和质量的效果来为发布商提供资讯以做出明智决定。

[0026] 运行时,分组估计系统100促进将对消费者的广告作为目标。例如,系统100可向广告管理系统74提供描述所选受众分组的一个或多个特征的数据。该数据可用来将针对与发布商web站点60的所选受众分组内的用户标识符关联的未来页面浏览的广告作为目标。

[0027] 分组估计系统100采用扫描器104来访问并扫描发布商日志80。具体而言,扫描器104取回包括用户标识符、发布商标识符、页面浏览值和每个广告印象的键值的记录并根据

用户标识符和发布商标识符组织发布商记录80。这些经排序的数据存储在发布商记录数据存储116中。一些实现方式中,扫描器每天在采集了特定天的所有发布商日志后扫描发布商日志80。

[0028] 聚合器106访问发布商记录116并根据用户标识符/发布商标识符对进行聚合记录。这些经聚合的记录作为发布商页面浏览数据存储114中。一些实现方式中,每个发布商的发布商页面浏览数据包括代表在第一时间段期间(如一个月)出现的页面浏览的页面浏览记录。每个页面浏览记录代表和在第一时间段期间出现的第二时间段(如一个月内的某一天)期间与用户标识符和发布商标识符对关联的一个或多个页面浏览。因而,如果一个特定用户在一个个月中的二十个不同日子期间到访发布商,那么,页面浏览数据存储114中会存在具有对应的用户标识符和发布商标识符对的二十条记录。此外,每条记录包括为由页面浏览记录代表的针对页面浏览所出现的印象提供的发布商键值。图2详细描述页面浏览记录。

[0029] 分组管理引擎102接收发布商的分组定义信息。分组定义根据发布商的一个或多个发布商键值定义分组。分组定义被提供给分组估计器108,分组估计器108使用页面浏览数据114和分组定义信息来确定一个或多个分组特征,如,分组大小、分组的页面浏览数和/或分组质量。

[0030] 一些实现方式中,分组估计器108根据包括满足分组定义所定义的分组键值的页面浏览数据114记录为分组指派用户标识符。根据这些指派的用户标识符确定一个或多个分组特征。图2详细描述对这些特征的确定。

[0031] 一些实现方式中,分组估计器108还确定分组质量测量。一些实现方式中,质量测量由下式给定:

$$[0032] \quad Q_s = f\left(\frac{N_{S+S^*}}{N_s}\right)$$

[0033] 其中

[0034] Q_s 是质量测量;

[0035] N_s 是分组内的页面浏览数;并且

[0036] N_{S+S^*} 是分组内和分组外的页面浏览总数。

[0037] 质量测量使得发布商能够判断特定受众分组的质量。总体而言,质量测量越高,可通过把分组外的页面浏览选择为目标而生成的印象数越大。例如,假定100个用户属于发布商web站点的第一分组。第一分组定义成对体育类感兴趣的用户。这些用户提供1000次分组内的页面浏览,即,包括体育类发布商键值的页面浏览总共有1000次。对于同样的这100个用户,发布商web站点上总共有5000次页面浏览,即,除了包括体育类发布商键值的总共1000次页面浏览外,这些用户在发布商web站点还有不包括体育类发布商键值的额外的4000次页面浏览。而且,假定另外100个用户属于发布商web站点的第二分组。第二分组定义成对钩针编织类感兴趣的用户,并且这些用户提供1000次分组内的页面浏览。然而,对于这100个用户,发布商web站点上总共只有1100次页面浏览。因而,第一分组的质量比第二分组质量高。

[0038] 可以把所确定的一个或多个分组特征以分组响应数据112的形式提供给发布商,这些特征例如为分组大小、分组的页面浏览数、属于分组的用户和/或分组质量。发布商可

使用用户界面审查分组响应数据112以确定是否要进一步分析分组以及/或者以分组为目标。一些实现方式中,如果发布商决定以某分组为目标,分组管理引擎102可生成分组填充数据113,并把分组填充数据提供给广告管理系统74以作为目标。一些实现方式中,分组填充数据113是用户标识符和定义分组的数据的关联。该关联被用作目标标准。广告管理系统74或进行请求的发布商60可使用分组填充数据113将和指派给分组的用户标识符关联的未来页面浏览的广告作为目标,或者,一些实现方式中,将在被确定属于分组的任何其他用户到访特定发布商站点时和该用户关联的未来页面浏览的广告作为目标。

[0039] 一些实现方式中,系统100接收或者取回不带有上下文值的数据。该数据可占据存储空间并增加系统100的处理时间。因而,一些实现方式中,分组管理引擎102提供识别、标记和处置无用数据的机制。例如,分组管理引擎102可删除一次性唯一标识符、广告描述符、顾客错误等。一次性唯一标识符是识别唯一发生的事情的任何标识符,例如一个月内特定用户标识符只有一次页面浏览。广告描述符是描述广告大小、风格和/或位置的描述符。客户不可能使用该数据来构建用户组的行为简档。顾客错误可能与错误地构建的广告统一资源定位符有关。

[0040] 分组管理引擎102可实现用于管理和处置特定键值的策略。例如,分组管理引擎102可根据如特定广告描述符的排除键值类型列表事先处置特定键值。一些实现方式中,引擎102可识别低值键值并滤除这些键值。例如,可用聚合器106识别三十天内仅提供给少数用户(例如少于三个)的页面浏览的键值并将其滤除。

[0041] §3.0示例页面浏览记录

[0042] 图2是示例页面浏览记录200的框图。页面浏览记录200代表指示哪个发布商键值被提供给对应的页面浏览印象的页面浏览/键值映射。所描绘示例中的页面浏览记录200的格式为包括每列一个单元的记录。单元代表单个用户一天的聚合发布商日志数据。提供下面记录200用作示例,也可使用关联页面浏览和键值或发布商跟踪的其他属性的其他数据结构。

[0043] 示例页面浏览记录200被配置用于优化存储空间分配和处理效率。下文所述示例实现方式中,页面浏览记录200是存储一天的用户标识符/发布商标识符对的印象数据的简明记录。此外,记录200中仅存储了为该天的用户标识符的印象提供的键值。这些键值用记录中存储的位图索引。记录中不包括为印象提供的其他键值的列。

[0044] 页面浏览记录200用聚合器106生成,所产生的聚合记录存储在页面浏览数据存储114中。一些实现方式中,每天针对特定web页面为每个用户生成一个记录,持续生成多达30天。例如,对于发布商站点的任何给定web页面,可以存在每个用户标识符/发布商标识符对的30天的页面浏览数据114的多达30条记录。另外,也可使用其他时间段。

[0045] 然而,其他实现方式中,不跟踪特定web页面,仅跟踪发布商键值。这些实现方式中,每天为每个用户生成一条记录,持续生成多达30天。例如,可以存在每个用户标识符/发布商标识符对的30天的页面浏览数据114的多达30条记录。

[0046] 所描绘的页面浏览记录200包括代表用户标识符/发布商标识符对键的列202。一些实现方式中,该键是用户标识符和发布商标识符的元组,例如是用户标识符或者从用户标识符导出的而且和用户标识符唯一关联的数据以及发布商标识符。列204、206、208、210、212、214、216和218对应于发布商跟踪的且为特定一天(或其他最小时间间隔)的用户印象

提供的键值。该示例中,对于和为该天的印象提供的键值对应的每个键值存在一列。

[0047] 例如,假定发布商标识符为“1000”的发布商为和列202所示的用户标识符关联的用户会话供应了三个页面浏览。每个页面浏览还供应了一个或多个广告,而且为每个广告提供了印象键值。假定所有页面浏览都位于特定发布商站点和特定区域,站点例如为由值“S10”识别的站点,区域例如为由标识符“Z11”识别的web页面。发布商记录中一天的特定发布商和特定用户的示例键值为:

[0048] 爱好=编织;页面=烹饪;页面浏览=1

[0049] 兴趣=烹饪;页面浏览=1

[0050] 上下文=金融;关键字=再融资;c=n;页面浏览=4

[0051] 上下文=金融;页面浏览=4

[0052] 上下文=金融;q=wsj_prime;页面浏览=4

[0053] 兴趣=电影;爱好=编织;页面浏览=16

[0054] 上述示例中,本示例中使用键值“页面浏览=<x>”来把印象聚合到页面浏览中。例如,对于发布商“1000”,六个印象聚合到了三个页面浏览中(例如标号1、4和16)。页面浏览记录200代表使用上述数据为页面浏览1、4和16和六个印象所写的的数据。

[0055] 列204包括关键类型“兴趣”的键值“编织”。索引值指示位图中的索引,下文将对位图进行描述。类似的,列206、208、210、212和214包括发布商的不同键值,即,列206包括关键类型“兴趣”的键值“电影”;列208包括关键类型“兴趣”的键值“烹饪”;列210包括关键类型“关键字”的键值“再融资”;列212包括关键类型“上下文”的键值“金融”;以及,列214包括关键类型“q”(查询参数)的键值“wsj_prime”。列216包括指示站点S10的键值“S10”。该站点标识符识别发布商的特定web站点。列218包括指示如特定web页面的广告区域的键值“Z11”。

[0056] 对于每个键值,为其提供了键值的页面浏览的数目各不相同。例如,键值“编织”出现在两个页面浏览中(即页面浏览1和16,页面浏览总数为2);键值“电影”出现在一个页面浏览中(即,页面浏览16,页面浏览总数为1),等等。一些实现方式中,页面浏览数目也可作为元组值的一部分存储在单元中。然而,如下所述,使用位图也可确定页面浏览数量。

[0057] 页面浏览记录200的每个相应字段包括相应的索引值。这些索引值用于索引存储在列220中的位图。这些索引值用于为页面浏览记录代表的一个或多个页面浏览的每一个生成页面浏览/键值映射。一些实现方式中,页面浏览/键值映射是在与提供给页面浏览印象的发布商键值的相应索引值对应的索引位置具有对应第一值(例如1)、并且在与未提供给页面浏览印象的发布商键值的相应索引值对应的索引位置具有对应的第二值(例如0)的页面浏览/键值位图。

[0058] 如图2所示,从左到右对页面浏览/键值位图进行索引,最左侧位于索引位置0,最右侧位于索引位置7。例如,第一位图242对应于页面浏览(1);第二位图244对应于页面浏览(4);以及第三位图246对应于页面浏览(16)。对于每个位图,列204对应于索引0,列206对应于索引1,列216对应于索引2,列208对应于索引3,列214对应于索引4,列210对应于索引5,列212对应于索引6,以及列218对应于索引7。如果某位低(即0),页面浏览不包括具有由该位索引的键值的印象。相反,如果某位高(即1),页面浏览包括具有由该位索引的键值的印象。虽然图2中的列未按索引值排序,但是,一些实现方式中,列可按位图索引递增顺序排列,即KV0、KV1、KV2...KV7。

[0059] 注意,所有三个印象都是针对站点标识符S10和区域标识符Z11的,以及注意站点标识符S10和区域标识符Z11分别由索引2和7进行索引。因而,每个位图中在索引2和7处的位等于1。

[0060] 对于索引0处的位图位置,由于存储在列204中且和索引0关联的键值“编织”被提供给具有第一页面浏览(页面浏览1)的印象,所以第一位图242包括值1。类似地,由于键值“编织”也被提供给具有第三页面浏览(页面浏览16)的印象,所以第三位图246在索引位置0处包括值1。相反,由于键值“编织”未被提供给具有第二页面浏览(页面浏览4)的印象,所以第二位图244在索引位置0处包括值0。

[0061] 以类似方式确定位图242、244和246的剩余位。一旦生成了位图,对所有位图特定索引位置处的所有相应位求和就生成等于为之提供键值的页面浏览数。

[0062] 页面浏览/键值位图有助于对指派给分组的用户标识符进行有效搜索。一些实现方式中,分组估计器108根据分组定义为每个页面浏览记录生成分组位掩码。该分组位掩码在与分组定义中的发布者键值相应的索引值对应的索引位置具有对应的第一值(例如1),并且在与不在分组定义中的发布者键值的相应的索引值对应的索引位置具有相应的第二值(例如0)。分组估计器108比较页面浏览记录的分组定义位掩码和页面浏览记录的每个页面浏览/键值位图,以确定哪个用户标识符属于分组。

[0063] 例如,假定分组被定义成从发布者1000接收到具有下述键值的印象的用户:

[0064] 原料=鹰嘴豆

[0065] 兴趣=电影

[0066] 兴趣=旅行

[0067] 关键字=再融资

[0068] 因而,分组定义由下面的布尔等式表示:

[0069] 原料=鹰嘴豆OR兴趣=电影OR兴趣=旅行OR关键字=再融资

[0070] 一些实现方式中,分组定义还可指定时间间隔中的最小页面浏览数。缺省定义是每任意给定天1个页面浏览;然而,分组定义也可指定其他页面浏览要求,例如任一一周的时间段内100个页面浏览;或者,任意给定天10个页面浏览,等等。

[0071] 对于记录200,分组估计器108会生成下面的分组定义位掩码:01000100。通过比较键值和记录200构建该位掩码。由于分组定义中既有键值“兴趣=电影”又有键值“关键字=再融资”,所以指派给索引位置1和5的值是1。

[0072] 分组估计器108执行逐位运算,以确定特定页面浏览是否满足分组定义。例如,匹配分组定义位掩码和页面浏览位图242、244和246,并且确定两个页面浏览满足分组定义。因而,用户标识符被指派给分组。此外,该天计数到两个页面浏览(即,位图244的第二页面浏览,这是由于存在“再融资”;和位图246的第三页面浏览,这是由于存在“电影”)。

[0073] 分组估计器108实现分组定义所需的必要布尔代数。例如,如果分组定义要求所有键值都存在(例如原料=鹰嘴豆AND兴趣=电影AND兴趣=旅行AND关键字=再融资),那么不计数任何页面浏览。

[0074] 分组定义位掩码以逐个记录为基础对分组定义中使用的位进行索引。由于用户标识符/发布商标识符对的每个记录仅对应于一天中(或者一些其他最小处理时间段)特定发布者页面的用户活动,所以每个用户标识符/发布商标识符对记录的对应的页面浏览/键值

位图可能不相同。因而,分组估计器108为每个记录生成分组定义位掩码。例如,假定记录200的同一用户第二天再次来到发布商站点并使得生成了20个页面浏览。有可能第二天的页面浏览记录的页面浏览/键值位图多于七位,这是由于会提供额外的或者不同的键值。

[0075] 例如,假定为第二天的所有页面浏览提供15个键值,这些键值包括“鹰嘴豆”、“电影”、“旅行”和“再融资”。因而,第二天的页面浏览/键值位图会有15位。现在,再假定各个键值的索引位置为1、3、14和15。考虑到该索引,同一分组定义的分组定义位掩码也有15位,而且包括值“010100000000011”。

[0076] 上述分组定义是相对简单的分组定义(“OR”运算),仅需要一个位掩码。然而,也可使用更复杂的分组定义。对于这些更复杂的定义(例如使用AND运算符、NAND运算符、NOR运算符等),每一记录可能要求多于一个位掩码。

[0077] 此外,如果要求某时间段的特定页面浏览计数,那么分组估计器108可响应于比较位掩码和位图确定用户是否属于所定义的分组来执行另外的决策操作(即,某时间段的页面浏览计数检查)。

[0078] §4.0示例过程

[0079] 图3是确定分组的一个或多个特征的示例过程300的流程图。过程300可在图1的分组估计系统100中实现。

[0080] 过程300访问发布商日志(302)。例如,扫描器104访问发布商日志80。发布商日志80存储描述针对特定发布商站点的页面浏览所递送的过往印象的数据。具体而言,发布商日志80可存储印象数据和页面浏览数据。对于日志80中的每个印象,数据可包括用户标识符、发布商标识符、页面浏览值以及为印象提供的发布商的任意数目的发布商键值。

[0081] 过程300根据发布商日志生成每个发布商的发布商页面浏览数据(302)。例如,扫描器根据发布商日志生成发布商记录116,聚合器106根据发布商记录116生成每个发布商的页面浏览数据。例如,发布商页面浏览数据可包括代表如一个月的时间段的在第一时间段期间出现的页面浏览的页面浏览记录。一个月时间段内的每个页面浏览记录可代表该月内出现的第二时间段(例如一天)期间和用户标识符及发布商标识符对关联的一个或多个页面浏览。这一天时间段内所代表的每个页面浏览记录包括具体用户标识符、发布商标识符以及为由于一个或多个页面浏览所出现的印象提供的发布商键值。虽然使用了一个月和一天的示例时间段,但是过程300可生成不同时间段的页面浏览记录。

[0082] 过程300接收发布商的分组定义(306)。例如,分组管理引擎102接收发布商的分组定义。分组定义根据发布商的一个或多个发布商键值定义分组。分组定义还可包括定义一些页面浏览的最小(或最大)页面浏览定义以及出现页面浏览的最小(或最大)时间段。分组估计器根据包括满足分组定义所定义的分组的发布商键值的记录,为分组指派用户标识符。

[0083] 过程300基于指派给分组的用户标识符和指派给分组的用户标识符的页面浏览记录确定一个或多个分组特征(308)。例如,分组估计器108基于指派给分组的当前用户确定分组特征。确定分组特征可包括根据可用的页面浏览记录确定页面浏览的第一数目。可用的页面浏览记录可包括指派给特定分组的用户标识符的记录。一些实现方式中,确定分组特征包括基于指派给分组的具体用户标识符的页面浏览记录确定目标标准。

[0084] 过程300响应于分组定义提供描述分组的一个或多个特征的响应数据(310)。例

如,分组管理引擎102使用分组定义确定并提供有关分组特征的响应数据。一个示例中,响应数据可被提供给广告管理系统,以促进广告将和指派给分组的具体用户标识符关联的未来页面浏览作为目标。

[0085] 图4是确定分组质量测量的示例过程400的流程图。例如,过程400可在分组估计器108内实现。

[0086] 过程400根据指派给分组的用户标识符的页面浏览记录确定页面浏览的第一数目(402)。例如,分组估计器108确定页面浏览数据114中的哪些页面浏览/键值映射指示满足分组定义的发布者键值。根据该信息,生成分组内的页面浏览的总数,分组内的页面浏览即包括满足分组定义的键值的页面浏览。

[0087] 过程400根据指派给分组的用户标识符的页面浏览记录确定页面浏览的第二数目(404)。例如,分组估计器108确定指派给分组的用户标识符的页面浏览记录代表的页面浏览总数,即,该用户的分组内的页面浏览及分组外的页面浏览这两者,一些实现方式中,这是和属于分组的用户标识符关联的所有页面浏览。

[0088] 过程400确定与页面浏览的第二数目和页面浏览的第一数目之比成比例的分组质量测量(406)。例如,分组估计器108使用页面浏览的第一数目和页面浏览的第二数目来计算分组质量测量。一些实现方式中,质量测量可以是分组内和分组外的页面浏览之和与分组内的页面浏览数目的比例的测量。

[0089] 图5是确定满足分组定义的页面浏览的示例过程500的流程图。例如,过程500可在聚合器106或分组估计器108中实现。

[0090] 过程500生成代表由页面浏览记录代表的一个或多个页面浏览中的每一个的页面浏览/键值位图(502)。例如,聚合器106生成位图来代表页面浏览记录中的页面浏览。运行时,对于每个页面浏览记录,发布者键值存储在页面浏览记录中的相应字段。每个字段和相应的索引值关联。

[0091] 如图2所详细描述,页面浏览/键值位图在和提供给页面浏览印象的发布者键值的相应索引值对应的索引位置包括第一值。此外,位图在和未提供给页面浏览印象的发布者键值的相应索引值对应的索引位置包括第二值。

[0092] 过程500根据分组定义为每个页面浏览记录生成位掩码(504)。例如,分组估计器108使用分组定义为每个页面浏览记录生成一个或多个位掩码。根据分组定义的复杂度,可能只需要为页面浏览记录生成一个位掩码,或者,对于复杂定义,需要为每个页面浏览记录生成多个位掩码。一个示例中,每个位掩码在与分组定义中的发布者键值的相应索引值对应的索引位置处包括第一值。此外,位掩码在与不在分组定义内的发布者键值的相应索引值对应的索引位置处包括第二键值。

[0093] 对于发布商的每个页面浏览记录,过程500比较页面浏览记录的分组定义位掩码和页面浏览记录的每个页面浏览/键值位图(506)。例如,分组估计器108执行比较分组定义位掩码和页面浏览/键值位图。如果页面浏览记录的分组定义位掩码和记录的页面浏览/键值位图之一匹配,而且位图和位掩码的对应比较生成布尔真值,那么记录中识别的用户可作为分组成员计数。

[0094] 图6是确定分组的一个或多个特征的另一示例过程600的流程图。例如,过程600可在分组管理引擎102或者分组估计器108或者二者的组合中实现。

[0095] 过程600接收发布商的分组定义(602)。例如,分组管理引擎102接收根据一个或多个键值定义分组的特定发布商的分组定义。键值涉及每个单独的发布商,因而可由每个发布商定义。

[0096] 过程600确定具有满足分组定义所定义的分组的发布商键值的页面浏览(604)。例如,分组估计器108确定具有和包括在分组定义所定义的分组中所需的键值匹配的键值的页面浏览。

[0097] 过程600基于被确定满足分组定义的页面浏览确定一个或多个分组特征(606)。例如,分组估计器108基于被预定满足分组定义的页面浏览确定页面浏览数据114的分组特征。

[0098] 过程600响应于分组定义提供描述分组的一个或多个特征的响应数据(608)。例如,分组管理引擎102提供响应数据112。响应数据112描述每个分组的特征。

[0099] 过程600响应于发布商的输入,创建用于将由分组定义所定义的分组作为目标的分组填充数据(610)。例如,分组管理引擎102可创建用户标识符和定义分组的数据之间的关联。广告管理系统74或者进行请求的发布商60可使用分组填充数据113将和指派给分组的用户标识符关联的未来页面浏览的广告作为目标,或者,一些实现方式中,将当被确定属于分组的任何其他用户到访特定发布商站点时和该用户关联的未来页面浏览的广告作为目标。

[0100] 本说明书所述的主题和操作的实施例,包括本说明书所公开的结构及其结构等价物,可实现在数字电子电路、或计算机软件、固件或硬件或这些的其中一个或多个的组合中。本说明书所述的主题内容的实施例可实现为一种或多种计算机程序,即,在计算机存储介质上编码的一个或多个计算机程序指令模块,用于由数据处理装置执行或者控制数据处理装置的操作。作为替代或附加,程序指令可编码在人工生成的传播信号上,例如,机器生成的电信号、光信号或电磁信号,生成该信号以编码信息来传送到适当的接收器装置用于由数据处理装置执行。计算机存储介质可以是如下设备或者包括在如下设备上:计算机可读存储设备、计算机可读存储衬底、随机访问存储器阵列或设备或者串行访问存储器阵列或设备,或者是一个或多个这些设备的组合。此外,虽然计算机存储介质不是传播信号,但是,计算机存储介质可以是编码在人工生成的传播信号上的计算机程序指令的源或者目的地。计算机存储介质还可以是一个或多个独立的物理部件或介质(例如,多个CD、磁盘或者其他存储设备),或者包括在这些部件上。

[0101] 本说明书所述操作可以实现成数据处理装置对存储在一个或多个计算机可读存储设备或者从其他源接收的数据执行的操作。

[0102] 术语“数据处理装置”包括所有类型的用于处理数据的装置、设备和机器,举例来说包括可编程处理器、计算机、片上系统或者前述种类中的多个或其组合。装置可包括专用逻辑电路,例如FPGA(现场可编程门阵列)或ASIC(专用集成电路)。除硬件之外,装置还可包括为所涉及的计算机程序创建执行环境的代码,例如构成处理器固件、协议堆栈、数据库管理系统、操作系统、跨平台运行环境、虚拟机或一个或多个所述种类的组合的代码。装置和执行环境可实现各种不同的计算模型架构,如web服务、分布式计算和网格计算架构。

[0103] 计算机程序(也称为程序、软件、软件应用程序、脚本或代码)可用任何形式的编程语言编写,包括编译或解释语言、陈述式语言或过程式语言,而且,计算机程序可部署成任

何形式,包括作为独立程序或者作为模块、部件、子例程、对象或者适用于计算环境的其他单元。计算机程序可对应于文件系统的文件,但这不是必须的。程序可存储在保持其他程序或数据的文件的一部分中(例如存储在标记语言文档中的一个或多个脚本)、专用于所涉及程序的单个文件中或者在多个协调文件中(例如,存储一个或多个模块、子程序或代码的部分的文件)。计算机程序可部署为在一个计算机或位于一处或分布在多处并由通信网络互联的多个计算机上执行。

[0104] 本说明书所述过程和逻辑流程可以由执行一个或多个计算机程序以通过操作输入数据并生成输出而执行动作的一个或多个可编程处理器执行。过程和逻辑流程还可由专用逻辑电路执行,并且装置也可实现成专用逻辑电路,例如FPGA(现场可编程门阵列)或ASIC(专用集成电路)。

[0105] 举例而言,适用于执行计算机程序的处理器包括通用微处理器和专用微处理器,以及任何类型的数字计算机的任何一个或多个处理器。通常,处理器从只读存储或随机访问存储器或者二者接收指令和数据。计算机的主要元件是用于根据指令执行动作的处理器和用于存储指令和数据的一个或多个存储器设备。通常,计算机还包括用于存储数据的一个或多个大型存储设备,或者可操作地耦合到用于存储数据的一个或多个大型存储设备以从中接收数据或向其传输数据或两者兼有,大型存储设备例如为磁盘、磁光盘或光盘。然而,计算机不是必须具有这种设备。

[0106] 适用于存储计算机程序指令和数据的设备包括所有形式的非易失性存储器、介质和存储器设备,举例而言包括半导体存储器设备,例如EPROM、EEPROM和闪存设备;磁盘,例如内置磁盘或移动磁盘;磁光盘;以及CD-ROM和DVD-ROM磁盘。处理器和存储器可由专用逻辑电路进行补充,或者包括在专用逻辑电路中。

[0107] 为了提供和用户的交互,本说明书所述主题的实施例可在具有用于向用户显示信息的显示设备(例如CRT(阴极射线管)或LCD(液晶显示器)监视器)以及用户用来给计算机提供输入的键盘和指点设备(例如鼠标或轨迹球)的计算机上实现。也可用其他类型的设备来提供和用户的交互;例如,提供给用户的反馈可以是任何形式的感官反馈(例如视觉反馈、声音反馈或者触觉反馈);而且,可以任何形式接收来自用户的输入,包括声音、语音或者触觉输入。此外,计算机可通过向用户使用的设备发送文档并从该设备接收文档来和用户交互;例如,通过响应于从用户的客户端设备上的web浏览器接收到的请求发送web页面给该web浏览器。

[0108] 本说明书所述主题的实施例可用包括后端部件(例如数据服务器)或者包括中间件部件(例如应用服务器)或者包括前端部件(例如客户端计算机,其具有用户可用来与本说明书所述主题的实现方式交互的图形用户界面或者Web浏览器)或者这些后端部件、中间件部件或前端部件的任意组合的计算系统实现。系统的部件可用任何形式或介质的数字数据通信(例如通信网络)互连。通信网络的示例包括局域网(LAN)、广域网(WAN)、互联网(如因特网)以及对等网络(例如自组织对等网络)。

[0109] 计算系统可包括客户端和服务端。客户端和服务端通常彼此远离而且典型通过通信网络交互。客户端和服务端的关系由于各个计算机上运行的而且彼此具有客户端-服务器关系的计算机程序而存在。一些实施例中,服务器传送数据(例如HTML页面)给客户端设备(例如为了显示数据并从和客户端设备交互的用户接收用户输入)。服务器可从客户端设

备接收客户端设备生成的数据(例如用户交互的结果)。

[0110] 虽然本说明书包括许多具体实现细节,但是这些实现细节不应理解成对任何发明或所要求保护范围的限制,相反,这是对特定发明的特定实施例特有的特征的描述。本说明书中在单独的实施例的上下文中描述的某些特征也可在单个实施例中结合实现。相反,单个实施例的上下文中描述的各个特征也可在多个实施例中分开实现或者实现成任何适当子组合。此外,虽然特征在上面被描述为按某个组合动作或者最初声称如此,但一些情况下,所声称组合的一个或多个特征可被排除在组合之外,而且,所声称组合可被导向到自组合或子组合变型。

[0111] 类似地,虽然操作在附图中描绘为具有特定次序,但是,这不应理解成要求为了实现期望效果要按所示的特定次序或者顺序执行这些操作,或者需要执行所有所示操作。某些情况下,多任务或并行处理是有益的。此外,上述实施例中各个系统部件分开不应理解成所有实施例都要求如此分开,并且应该理解,所述程序部件和系统通常可集成在单个软件产品中或者封装在多个软件产品中。

[0112] 因而,已经描述了主题的特定实施例。其它实施例在下述权利要求范围内。一些情况下,权利要求所记载的动作可按不同次序执行而且仍实现期望结果。此外,附图中所示过程不必要求所示的特定顺序或连续顺序来实现期望结果。某些情况下,多任务或并行处理可以是有益的。

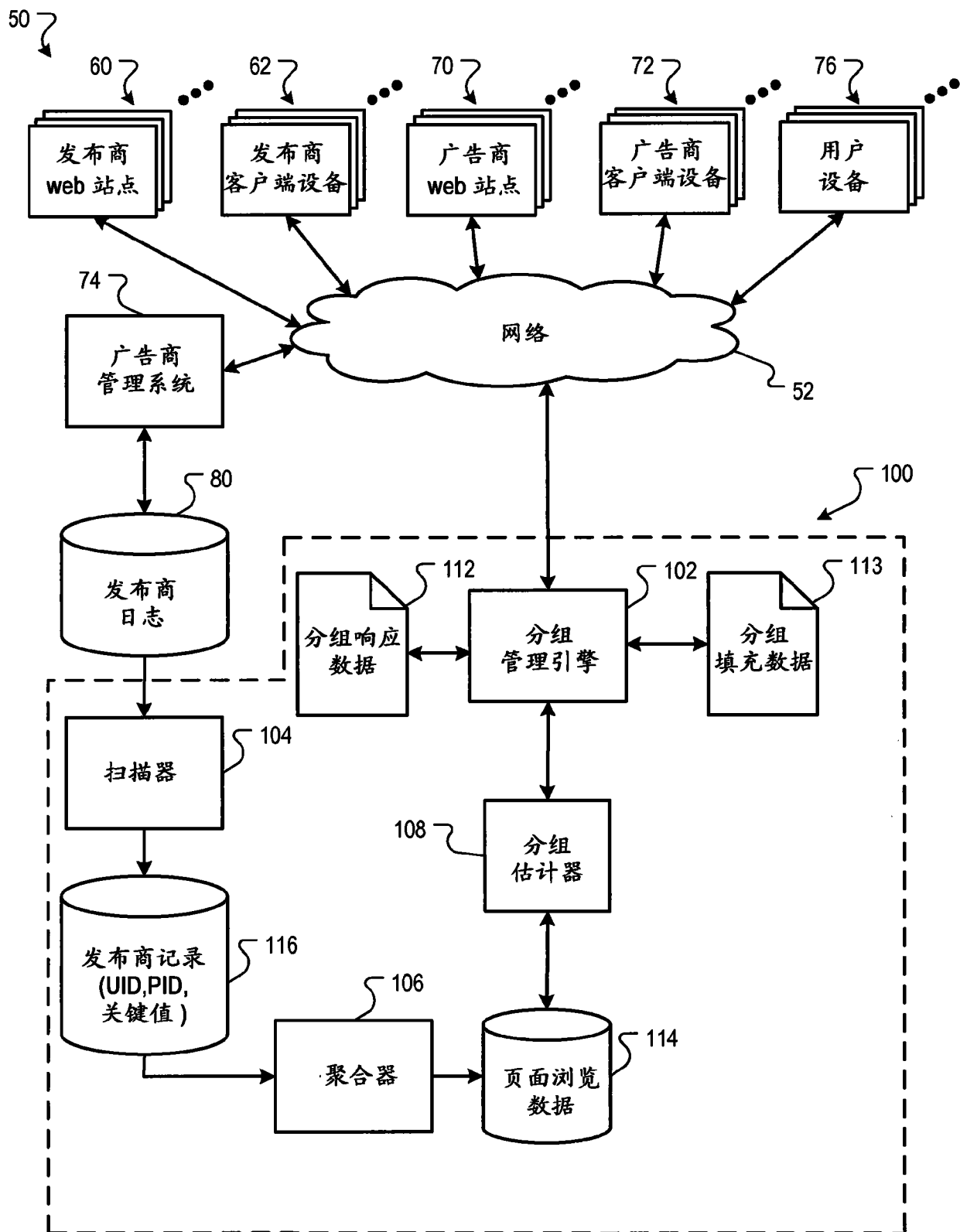


图1

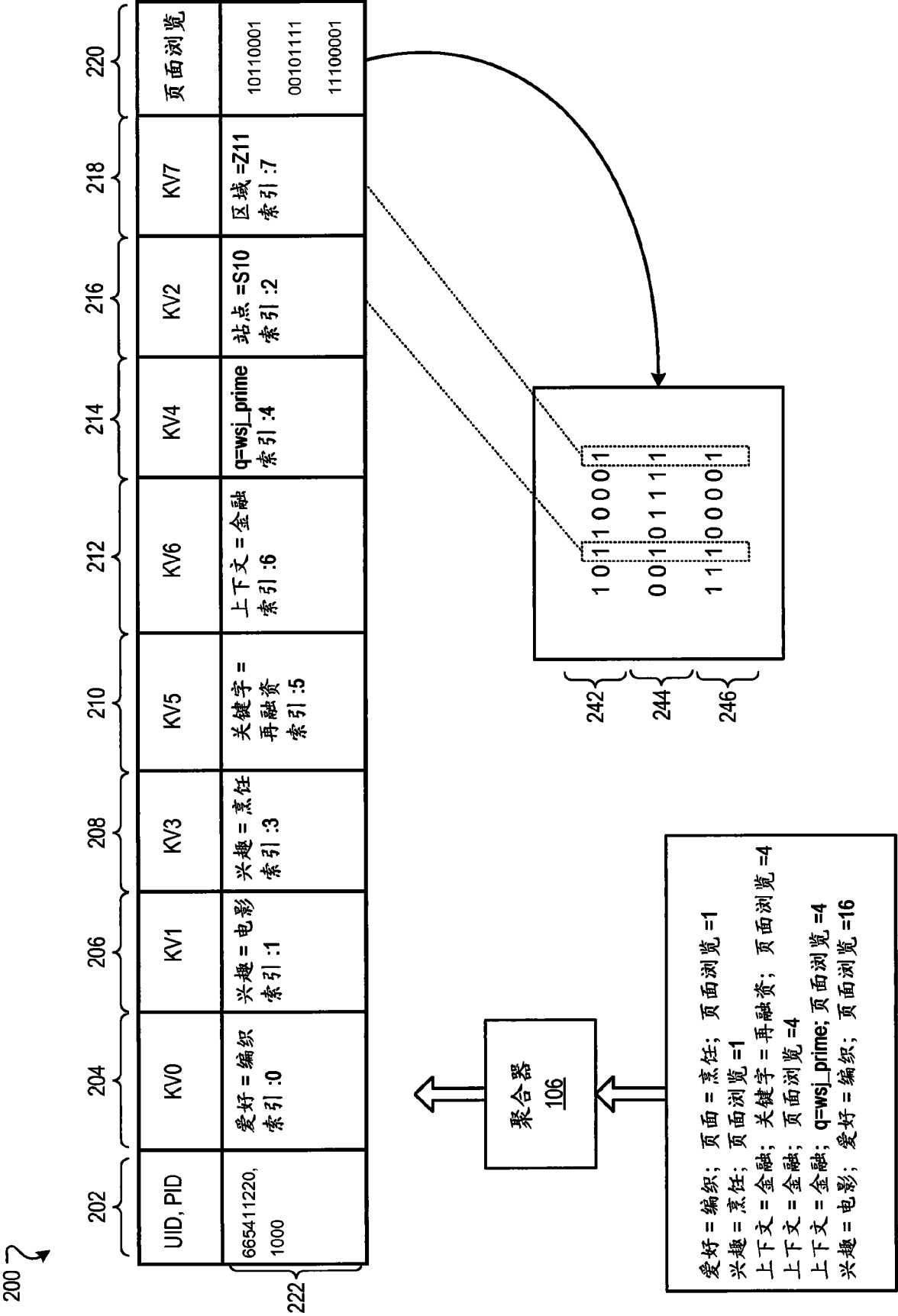


图2

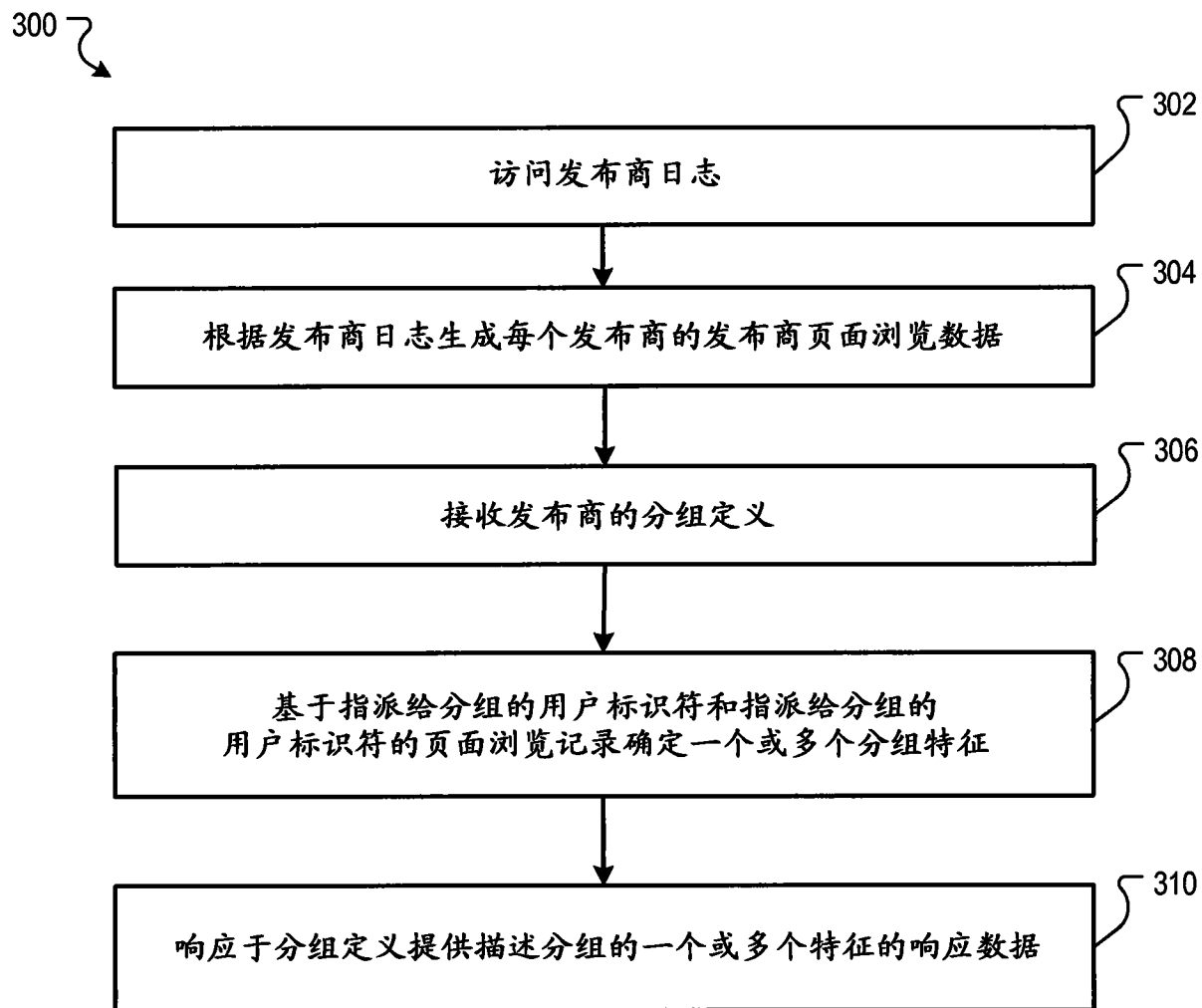


图3

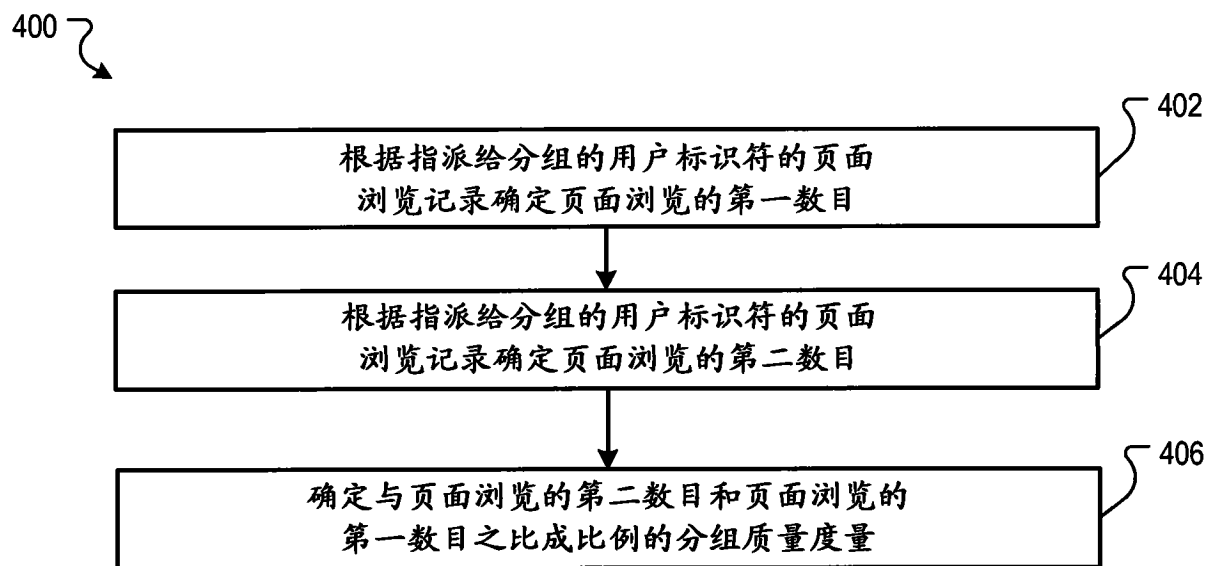


图4

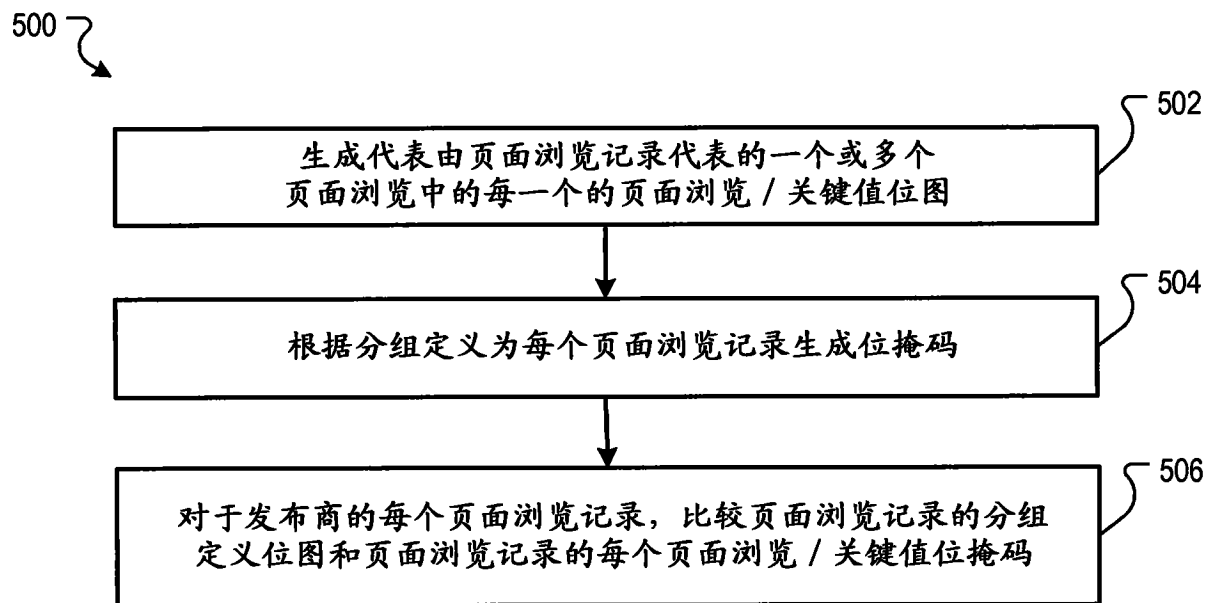


图5

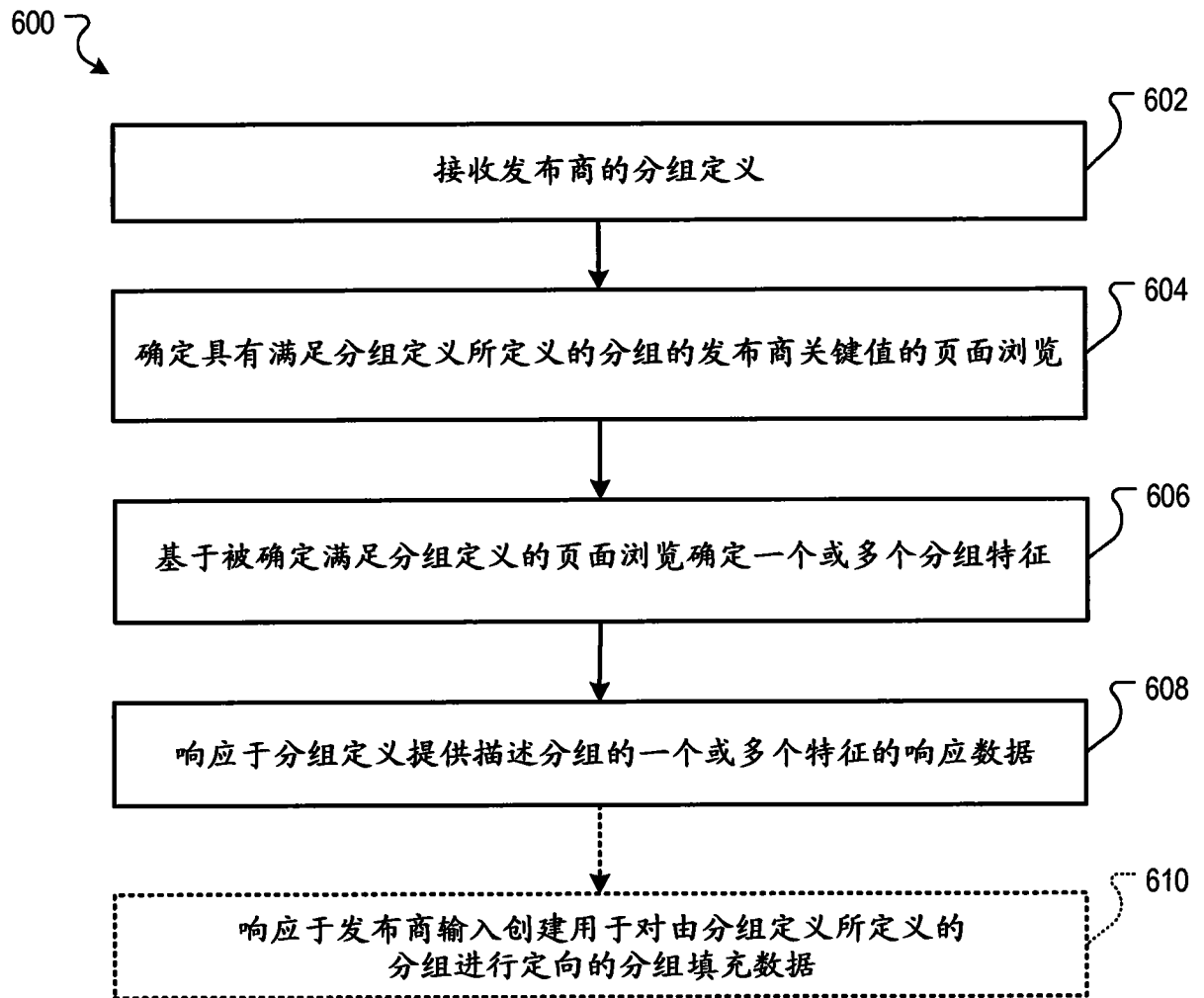


图6