

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号

WO 2019/019348 A1

(43) 国际公布日

2019 年 1 月 31 日 (31.01.2019)

WIPO | PCT

- (51) 国际专利分类号:
G06Q 30/06 (2012.01)
- (21) 国际申请号: PCT/CN2017/103989
- (22) 国际申请日: 2017 年 9 月 28 日 (28.09.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201710626280.0 2017 年 7 月 27 日 (27.07.2017) CN
- (71) 申请人: 上海壹账通金融科技有限公司 (ONECONNECT FINANCIAL TECHNOLOGY CO., LTD. (SHANGHAI)) [CN/CN]; 中国上海市徐汇区凯滨路 166 号 9 层, 10 层, Shanghai 200000 (CN)。
- (72) 发明人: 柴春燕 (CHAI, Chunyan); 中国上海市徐汇区凯滨路 166 号 9 层, 10 层, Shanghai 200000 (CN)。
- (74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区花城大道 85 号 3901 房, Guangdong 510623 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: PRODUCT INFORMATION PUSHING METHOD AND APPARATUS, STORAGE MEDIUM, AND COMPUTER DEVICE

(54) 发明名称: 产品信息推送方法、装置、存储介质和计算机设备

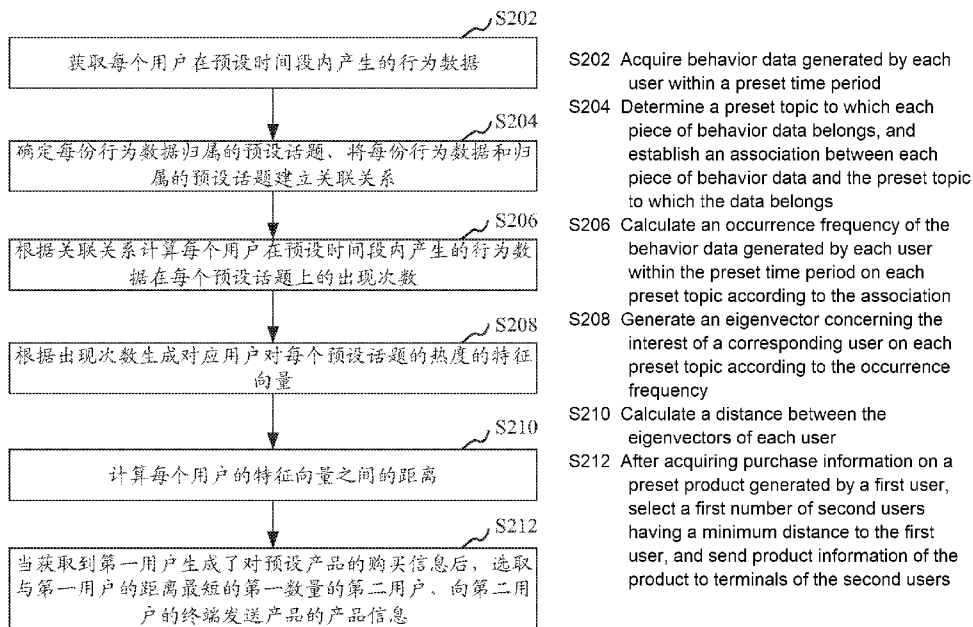


图 2

(57) Abstract: A product information pushing method, comprising: acquiring behavior data generated by each user within a preset time period (S202); determining a preset topic to which each piece of behavior data belongs, and establishing an association between each piece of behavior data and the preset topic to which the data belongs (S204); calculating an occurrence frequency of the behavior data generated by each user within the preset time period on each preset topic according to the association (S206); generating an eigenvector concerning the interest of a corresponding user on each preset topic according to the occurrence frequency (S208); calculating a distance between the eigenvectors of each user (S210); after acquiring purchase information on a preset product generated by a first user, selecting a first number of second users having a minimum distance to the first user, and sending product information of the product to terminals

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

of the second users (S212).

(57) 摘要: 一种产品信息推送方法, 包括: 获取每个用户在预设时间段内产生的行为数据 (S202); 确定每份行为数据归属的预设话题, 将每份行为数据和归属的预设话题建立关联关系 (S204); 根据所述关联关系计算每个用户在所述预设时间段内产生的行为数据在每个预设话题上的出现次数 (S206); 根据所述出现次数生成对应用户对每个预设话题的热度的特征向量 (S208); 计算每个用户的特征向量之间的距离 (S210); 当获取到第一用户生成了对预设产品的购买信息后, 选取与所述第一用户的距离最短的第一数量的第二用户, 向所述第二用户的终端发送所述产品的产品信息 (S212)。

产品信息推送方法、装置、存储介质和计算机设备

本申请要求于 2017 年 07 月 27 日提交中国专利局，申请号为 2017106262800，发明名称为“产品信息推送方法、装置、存储介质和计算机设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本申请涉及信息处理技术领域，特别是涉及一种产品信息推送方法、装置、存储介质和计算机设备。

背景技术

10 随着大数据应用的发展，即时通信和社交应用平台中存储了大量的用户信息，比如用户的年龄、爱好、职业等个人信息，以及用户发送的即时通信消息、文章等行为信息，而这些用户信息对商品广告等信息的精准推送，有着重要的参考价值。

传统的技术方案中，通常都是对所获取的海量用户进行简单的筛选，比如筛选出浏览过与服务商准备推送的产品信息相同或相似的用户，作为目标用户，并向其投放相关服务商的产品的推送信息。而现实中，存在众多的产品，比如保险或基金等金融产品，在用户不具有强烈的购买欲望时，是很少去浏览这些产品或类似产品的信息，或者即时浏览了这些产品信息，也并不一定具有购买意愿。因此，仅通过这种简单的筛选所确定的用于推送信息的目标用户不够精准。

15

20

发明内容

根据本申请公开的各种实施例，提供一种产品信息推送方法、装置、存储介质和计算机设备。

一种产品信息推送方法，包括：获取每个用户在预设时间段内产生的行为数据；确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系；根据所述关联关系计算每个用户在所述预设时间段内产生的行为数据在每个预设话题上的出现次数；根据所述出现次数生成对应用户对每个预设话题的热度的特征向量；计算每个用户的特征向量之间的距离；及当获取到第一用户生成了对预设产品的购买信息后，选取与所述第一用户的距离最短的第一数量的第二用户，向所述第二用户的终端发送所述产品的产品信息。

一种产品信息推送装置，所述装置包括：行为数据获取模块，用于获取每个用户在预设时间段内产生的行为数据；话题确定模块，用于确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系；次数计算模块，用于根据所述关联关系计算每个用户在所述预设时间段内产生的行为数据在每个预设话题上的出现次数；特征向量生成模块，用户根据所述出现次数生成对应用户对每个预设话题的热度的特征向量；距离计算模块，用于计算每个用户的特征向量之间的距离；及信息推送模块，用于当获取到第一用户生成了对预设产品的购买信息后，选取与所述第一用户的距离最短的第一数量的第二用户，向所述第二用户的终端发送所述产品的产品信息。

一个或多个存储有计算机可读指令的非易失性存储介质，所述计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行如下步骤：获取每个用户在预设时间段内产生的行为数据；确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系；根据所述关联关系计算每个用户在所述预设时间段内产生的行为数据在每个预设话题上的出现次数；根据所述出现次数生成对应用户对每个预设话题的热度的特征向量；计算每个用户的特征向量之间的距离；及当获取到第一用户生成了对预设产品的购买信息后，选取与所述第一用户的距离最短的第一数量的第二用户，向所述第二用户的终端发送所述产品的产品信息。

一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行如下步骤：获取每个用户在预设时间段内产生的行为数据；确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系；根据所述关联关系计算每个用户在所述预设时间段内产生的行为数据在每个预设话题上的出现次数；根据所述出现次数生成对应用户对每个预设话题的热度的特征向量；计算每个用户的特征向量之间的距离；及当获取到第一用户生成了对预设产品的购买信息后，选取与所述第一用户的距离最短的第一数量的第二用户，向所述第二用户的终端发送所述产品的产品信息。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征、目的和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为一个实施例中产品信息推送方法的应用环境图；

图 2 为一个实施例中产品信息推送方法的流程图；

图 3 为一个实施例中确定每份行为数据归属的预设话题的流程图；

图 4 为一个实施例中中心向量的计算的流程图；

图 5 为一个实施例中产品信息推送装置的结构框图；

图 6 为另一个实施例中产品信息推送装置的结构框图；

图 7 为一个实施例中话题确定模块的结构框图；

图 8 为又一个实施例中产品信息推送装置的结构框图；及

图 9 为一个实施例中计算机设备的内部结构图。

具体实施方式

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。本申请所使用的术语“第一”、

5 “第二”等可在本文中用于描述各种元件，但这些元件不受这些术语限制。这些术语仅用于将第一个元件与另一个元件区分。举例来说，在不脱离本申请的范围的情况下，可以将第一用户称为第二用户，且类似地，可将第二用户称为第一用户。第一用户和第二用户两者都是用户，但其不是同一用户。

本申请实施例所提供的产品信息推送方法可应用于如图 1 所示的应用环
10 境中。参考图 1，服务器 120 与多个用户终端 110-1~120-x 通过网络相连接。每个用户终端 110 可通过服务器 120 来生成行为数据，服务器 120 可接收并存储每个用户终端 110 所产生的行为数据，并确定每份行为数据归属的预设话题，将每份行为数据和归属的话题建立关联关系；根据关联关系计算每个用户在预设时间段内产生的行为数据在每个话题上的出现次数；根据出现次
15 数生成对应用户对每个话题的热度的特征向量；计算每个用户的特征向量之间的距离；当获取到第一用户生成了对预设产品的购买信息后，选取与第一用户的距离最短的第一数量的第二用户，向第二用户的终端发送产品的产品信息。

在一个实施例中，如图 2 所示，提供了产品信息推送方法，该方法可应
20 用于如图 1 所示的服务器中，或者还可应用于终端等计算机设备上，该方法包括：

步骤 S202，获取每个用户在预设时间段内产生的行为数据。

本实施例中，行为数据为用户通过即时通信应用或社交应用等发表的信息，该行为数据包括用户终端通过即时通信应用或社交应用发送的聊天消息、
25 文章或日志等形式的数据。当该行为数据的形式为文章或日志等相应形式的文本数据时，可将每一份文章或日志作为对应的一份行为数据。当行为数据的形式为用户发送的文字聊天消息或语音聊天消息等独立的消息时，可每小

时或每天或每周等合适时长内，所产生的所有的消息，汇总为对应的一份行为数据。每份行为数据中还可携带生成该信息的用户的用户标识，或者建立该行为数据与该用户标识之间的关联关系，使得根据其中携带的用户标识或关联关系，确定每条行为数据的所属用户。并根据所确定的用户，区分并汇总每个用户在预设时间段内产生的行为数据。

预设时间段可为所设置的任意时间段，比如可设置为最近一年或两年等。其中，该行为数据可预先存储在计算机设备的本地存储器中，还可为存储在其它设备中。当存储在本地时，计算机设备可从本地存储器中读取每个用户的行为数据；当存储在其它设备上时，可向对应设备发送行为数据获取请求，并接收对应设备根据该请求所返回的行为数据。

步骤 S204，确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系。

本实施例中，计算机设备还预先设置了多种话题，该话题可包括购物、销售、房价、保险、金融、恋爱、游戏、健康等其中的任意几种。服务器可对每份行为数据进行语义解析，根据解析出的语义，确定每个行为数据归属的话题。具体地，可针对每份行为数据进行词句的划分，针对划分的每个词句进行语音解析，根据解析出的每个词句的语义，分析出由该行为数据归属的话题。计算机设备中还预设了话题计算模型，比如可根据预设的 LDA(Latent Dirichlet Allocation, 潜在狄立克雷分配)话题模型计算出每份行为数据所归属的话题，并建立该行为数据和所归属的话题之间的关联关系。其中，一份行为数据可归属一种或几种话题，进一步地，还可不归属于任意一种预设的话题。

步骤 S206，根据关联关系计算每个用户在预设时间段内产生的行为数据在每个预设话题上的出现次数。

本实施例中，计算机设备可根据所建立的每份行为数据和话题之间的关联关系，统计每个话题与每个用户的所有行为数据中，具有关联关系的行为数据的数量，该数量即为在该用户预设时间段内产生的行为数据在每个话题

上的出现次数。该出现次数可反映出用户对每个话题的感兴趣程度，进而可用于反映出每个用户的用户特征。可以理解地，当行为数据越多，所计算出的出现次数所反映的用户特征的准确性越大。

5 在一个实施例中，还可对每份行为数据进行筛选，确定筛选后的确定每份行为数据归属的预设话题。

具体地，可根据每份行为数据所包含的文字内容以及数量等几种因素，来确定是选取该行为数据还是剔除该行为数据。比如当一份行为数据的文字数量小于一定的数量，和/或解析出的该行为数据所表达的信息相对于每种预设话题，均没有任何意义时，则可剔除该行为数据。并针对所筛选保留的行为数据，确定其归属的话题。同时被舍弃的行为数据可不与任何话题相关联，从而无需对被舍弃的话题进行解析，降低计算机设备的资源消耗。

步骤 S208，根据出现次数生成对应用户对每个预设话题的热度的特征向量。

本实施例中，计算机设备可根据对每个话题的出现次数，生成对应的特征向量，并建立该特征向量与对应用户的用户标识之间的关联关系。该特征向量的维度即为话题的种数，特征向量的每个元素，用于反映相应用户对对应一种话题的热度。当预设的话题数量为话题 1~话题 n 这 n 种话题时，对应的特征向量 S 可计为 $[s_1, s_2, s_3 \dots s_n]_i$ 。其中，特征向量的角标 i 用于反映该特征向量为用户 i 的特征向量， s_1 至 s_n 分别用于反映对应用户 i 产生的行为数据中，话题 1~话题 n 的讨论热度。

在一个实施例中，特征向量中的元素的数值可为对该话题的讨论次数或频率等。如可将每个话题的出现次数直接设置为特征向量中对应元素的数值，或者可将该次数进行相应的运算处理，将处理后的值作为对应元素的数值。比如还可对每个出现次数进行归一化处理，将归一化处理后的值作为对应元素的数值。通过进行归一化处理，可降低因不同的用户的行为数据的份数的差异过大，而导致后续进行距离计算时，使得每个距离之间的差异性过大。

步骤 S210，计算每个用户的特征向量之间的距离。

本实施例中，计算机设备可按照向量之间的距离的计算方式来计算每个用户的特征向量和其他用户的特征向量之间的距离，该距离反映了对应两个用户之间的相似度。距离数值越小，则表示两者之间的相似度越大。计算机设备可获取两个未计算距离的用户的特征向量，两所获取的其中一个特征向量每个参数，与对应另一个获取的特征向量中的对应参数相减，得到差值，然后将每个差值进行平方求和，所得到的平方和数值即为相应两个用户的特征向量之间的距离。

具体地，用户 k 的特征向量和用户 j 的特征向量之间的距离 d_{kj} ，可通过公式 $d_{kj} = \sum_{i=1}^n (s_{ik} - s_{ij})^2$ 来计算第一距离。其中， n 表示特征标签向量维度，该维度即为预设的提数量， s_{ik} 和 s_{ij} 分别表示用户 k 的特征向量和用户 j 的特征向量中的第 i 个参数。通过上述的计算公式，分别可计算出每两个用户的特征向量之间的距离。

步骤 S212，当获取到第一用户生成了对预设产品的购买信息后，选取与第一用户的距离最短的第一数量的第二用户，向第二用户的终端发送产品的产品信息。

本身实施例中，第一用户为购买了预设产品的用户。第二用户为被确定推荐该预设产品的产品信息的用户。该预设产品可为任意的产品，包括实体产品和虚拟产品。实体产品可包括零食、数码以及服饰等任意类型的产品；虚拟产品可包括虚拟游戏类的产品和金融类的产品等，金融类的产品可为保险、基金、股票等任意一种产品。产品信息包括相应产品的产品名称以及购买方式和购买价格等相应的产品推荐信息。

计算机设备可获取第一用户的购买信息，所获取的购买信息可为第一用户向该计算机设备发送的购买信息，还可为从存储了第一用户的购买信息的其他设备上所获取的购买信息。购买信息中包括相应购买产品的产品标识以及用户标识，使得可根据该用户标识读取对应用户的特征向量，根据该产品标

志可获取对应产品的产品信息。

在读取相应的第一用户的用户标识后，可根据该用户标识，读取与对应第一用户的距离最短的第一数量的用户的用户标识，所读取的用户标识即为对应第二用户的用户标识，并向该第二用户的终端发送相应的产品信息。其中，第一数量可为预设的固定数值，还可为根据相应的距离大小所确定的数值。比如，可计算与第一用户的特征向量的距离小于预设距离的其他特征向量的数量，将所计算出的数量设置为第一数量。

在一个实施例中，在步骤 S210 之前，还包括：根据所计算出的每个用户的特征向量与其他用户的特征向量之间的距离，对其他用户进行排序。比如可按照距离的大小进行顺序排序。举例来说，设存在 M 个用户，针对用户 i ，其他用户相对于该用户 i 的距离，按照从小到大的排序，分别为：用户 3、用户 9、用户 6、用户 80 等等。

计算机设备可从相对于第一用户的该排名中，读取相对于该第一用户的距离最小的前第一数量的用户的用户标识，所读取的用户标识即为第二用户的用户标识。比如，当该第一用户为上述的用户 i ，第一数量为 3 时，则可读取上述的用户 3、用户 9、用户 6 的用户标识，并向该用户 3、用户 9、用户 6 的终端发送对应产品的产品信息。

上述的产品信息推送方法，通过获取每个用户在预设时间段内产生的行为数据；确定每份行为数据归属的预设话题，从而计算出每个用户的所有行为数据在每个预设话题上的出现次数，根据该次数来生成对应用户对每个预设话题的热度的特征向量，再根据所生成的特征向量来计算每个用户的距离，使得根据该距离大小来体现出每个用户之间的相似度的大小，并当获取到第一用户生成了对预设产品的购买信息后，选取与第一用户的距离最短的第一数量的第二用户，向第二用户的终端发送产品的产品信息，从而提高了向用户进行产品信息推送的精准性。

在一个实施例中，在步骤 S212 之前，上述方法还包括：获取每个用户的

个人信息；根据个人信息计算每个用户之间的相似度。选取与第一用户的距离最短的第一数量的第二用户，包括：从与第一用户的相似度大于相似度阈值的用户中选取与第一用户的距离最短的第一数量的第二用户。

本实施例中，该个人信息包括用户的性别、年龄、职业、爱好等字段，
5 计算机设备可预先在本地的存储器中存储有每个用户的个人信息，还可向其它设备发送用户的个人信息获取请求，并从其它设备中获取个人信息。计算机设备可根据个人信息中的每个字段，来计算每个用户之间的相似度。该相似度用于反映用户之间的相似性，可以理解地，具有相同或相似的字段的用户之间的相似度较大。

10 具体地，可根据上述个人信息中的性别、年龄、职业、爱好等其中的任意一种或几种字段，计算每个用户之间的相似度，并根据该相似度来划分用户组，将相似度超过预设的相似度阈值的用户划分为同一用户组之中，使得在与第一用户处于相同用户组之中进行特征向量之间的距离的计算和/或第二用户的选取。计算机设备可从相似度大于相似度阈值的用户中，选取与第一用户的距离最短的第一数量的第二用户。即可在处于相同用户组中来选取
15 与第一用户的距离最短的第一数量的第二用户。

在一个实施例中，可预先划分出预设数量的类别的用户组，并设置每个类别的用户组中所需包含的具体的字段特征，判断每个用户的个人信息中的字段与所划分出的每个类别的用户组中的字段特征的匹配度，并根据该匹配
20 度的大小，将其划分至最高的匹配度对应的用户组中，并判定处于相同用户组中的用户之间的相似度大于该相似度阈值。举例来说，可根据年龄的年龄段、职业所归属的职业类别、爱好的归属类别以及性别等其中的一种或几种的组合，设置出相应数量的用户组。计算每个用户的个人信息中的年龄、职业、爱好以及性别与被设置的用户组的匹配度。

25 本实施例中，通过进一步结合用户的个人信息，根据用户的个人信息来计算每个用户的相似度，并从与第一用户的相似度大于相似度阈值的用户中选取与第一用户的距离最短的第一数量的第二用户，使得所选取出的用于进

行产品信息推荐的第二用户的精准度更高。

在一个实施例中，如图 3 所示，确定每份行为数据归属的预设话题，包括：

5 步骤 S302，提取每份行为数据中的关键词句。

本实施例中，计算机设备可对每份行为数据中的内容进行词句切分，对切分后的词句进行筛选，将筛选出的词句设置为行为数据的关键词句。

计算机设备可预先建立词句的语义数据库，该语义数据库中包含了大量的词句（即词语和句子）。根据行为数据的所属语言，按照相应的语法树的规则，并结合语义数据库中所记录的句子，将行为数据中的每个句子切分成
10 相应数量的词句。并确定每个词语在该句子中的词性以及在该行为数据中所处的位置，比如确定某个词语为名词，并为该句子中的主语。词句在行为数据中所处的位置包括处于标题、处于行为数据正文中、处于章节的题目中、行为数据作者、发表时间等。

15 在进行词句切分的过程中，若一个句子中的连续排列在一块的多个词语，在数据库中对对应存在一个完整的词语，则将该多个词语组成一个词语，使切割后的词句保持一个整体。

在完成词句切分后，可根据所切分的词句的词性，进行筛选，以删除对于行为数据的话题判定具有干扰或意义不大的词句。具体地，可将词性判断
20 为停用词或助词等词性的词句进行删除。比如，停用词为“the”、“is”、“at”、“that”、“是”、“的”等，助词比如为“也”、“者”、“乎”等。通过对切分后的词句的筛选，既可减少计算机设备对数据处理的计算量，又可排除被删除的词句的干扰，提高了话题判定的准确性。

步骤 S304，计算每个关键词句和每个预设话题的第一关联度。

25 本实施例中，计算机设备可预先设置了多个话题，在获取某一关键词句后，可首先检测是否预先计算出该关键词句或者与该关键词句相近的词句，和每个话题的关联度，若是，则可不必再次计算，直接将该关键词句或相近

的词句与每个话题的关联度，设置为该关键词句和每个预设话题的第一关联度。若不存在，则可对每个关键词句进行语义解析，获取每个关键词语与每个预设话题的第一关联度。第一关联度可用百分比来表示，其关联度的大小处于 0~1。举例来说，比如预设话题包括购物、销售、房价、保险、金融、恋爱、游戏、健康等。则可针对获取的每个关键词句，分别计算其与上述的购物、销售、房价、保险、金融、恋爱、游戏、健康等话题的第一关联度。进一步地，根据每个关键词句和所有预设话题的第一关联度可构成相应的第一关联度向量。可以理解地，该第一关联度向量是与上述的特征向量的维度相同的向量，其中的每一个元素分别表示对应关键词句和其中一种预设话题的关联度。

具体地，计算机设备可预设有语义分析模型，并可根据该语义分析模型来计算出所获取的每个关键词句和每个话题的关联度。其中，该语义分析模型可包括根据潜在语义分析（Latent Semantic Index, LSI）、潜在狄立克雷分配（Latent Dirichlet Allocation, LDA）或者 Word2vec 等所设置出的语义分析模型。

步骤 S306，根据第一关联度计算出对应行为数据和每个预设话题的第二关联度。

计算机设备可对每个话题，将其与每个关键词句的第一关联度进行加权求和，从而可计算出该行为数据对每个话题的第二关联度。具体地，可将每个构成的第一关联度向量中，处于相同维度上的元素进行加权求和，根据所计算出的每个元素的加权和，生成该第二关联度向量，该第二关联度向量中的每个元素即为对应行为数据和相应话题的第二关联度。其中，每个关键词句的权值可根据检测出的该词句在行为数据中的位置、词性，以及行为数据中的关键词句的数量等一种或多种因素来确定。比如当该关键词句处于行为数据中的标题中时，可设置相对较大的权值；当该行为数据中的每个关键词句的数量较多时，为每个关键词句设置的权值则相对较小。

步骤 S308，将超过关联度阈值的第二关联度对应的预设话题确定为行为

数据归属的预设话题。

本实施例中，计算机设备中还设置了相应的关联度阈值，关联度阈值可为自定义设置的数值。计算机设备可将每个第二关联度与该关联度阈值进行比较，提取超过该关联度阈值的第二关联度的话题，将其与相应行为数据建立对应关系，使得将其设置为该行为数据所归属的话题。举例来说，当关联度阈值为 0.5 时，则可将所有大于 0.5 的第二关联度对应的话题，与该行为数据之间建立关联关系，将其设置为该行为数据所归属的话题。

本实施例中，通过对行为数据划分关键词句，然后计算每个关键词句和每个预设话题的第一关联度，根据所有的第一关联度来确定行为数据所归属的话题，可进一步提高对行为数据进行话题归属划分的精准性。

在一个实施例中，根据出现次数生成对应用户对每个预设话题的热度的特征向量，包括：根据每个用户的每份行为数据与归属的预设话题的出现次数，以及与归属的话题的第二关联度，计算出每个用户的特征向量。

本实施例中，计算机设备可针对每个话题，计算出每个用户的行为数据中，对归属于该话题的数量，该数量即为上述的出现次数。并结合归属该话题的行为数据第二关联度，设置相应的权值，根据该出现次数和权值，计算出该特征向量中对应于该话题的参数数值。该参数数值可为该权值和次数的乘积。其中，该权值可为根据每个第二关联度所计算出的平均值。

举例来说，假设存在话题 i ，针对用户 i 的行为数据中，被确定为归属至该话题 i 的行为数据的数量为 10 个，即该出现次数为 10。计算机设备可分别提取所计算出的该 10 个行为数据相对于该话题 i 的第二关联度，根据所提取出的 10 个第二关联度计算出相应的权值，将该权值和次数的乘积设置为用户 i 的特征向量 S_i 中用于反映对该话题 i 的热度的参数 s_i 的数值。针对每个用户以及每个话题，可均采用上述的方法，从而可计算出每个用户的特征向量。

本实施例中，通过进一步结合所计算出的第二关联度来计算出用户的特征性向量，使得所计算出的用户的特征向量更能量化表达出用户对每个话题

的热度，从而根据该特征向量所确定的第二用户更精准。

在一个实施例中，在步骤 S208 之后，还包括根据距离生成第二数量的聚类，将每个用户划分至对应的一个聚类之中。

5 本实施例中，计算机设备还预先设置了第二数量的类别，用于将每个用户划分至对应的一个类别之中，被划分至同一类别中的用户构成对应一个聚类。其中，该第二数量可为任意设置的合适的小于用户总数的数值，第二数量越大，则对用户划分的类别越细致。计算机设备可预先针对每个聚类设置相应的中心向量，中心向量是用于表征某一类别的信息对应的种子用户所共
10 有的特征信息，中心向量的形式与特征向量相同，具有与特征向量相同长度的维度，其每个维度上的参数表征的含义与特征向量对应的参数表征的含义相同。

 计算机设备可计算每个特征向量与每个中心向量之间的距离，将特征向量划分至对应距离最小的一个聚类之中，从而完成了对每个用户的类别的划
15 分，被划分至同一聚类中的特征向量所对应的用户即处于同一聚类。

 选取与第一用户的距离最短的第一数量的第二用户的步骤，包括：从与第一用户处于相同聚类的用户中选取与第一用户的距离最短的第一数量的第二用户。

 当获取到第一用户生成了对预设产品的购买信息后，确定该第一用户所
20 被划分的聚类，并从该聚类中读取特征向量与该第一用户的特征向量的距离最短的，第一数量的用户的用户标识。通过从处于相同聚类的用户中进行用户的选取，可进一步提高了对用户选取的精准度。

 在一个实施例中，可根据每个用户的特征向量，按照预设的聚类算法计算出每个聚类的中心距离。该聚类算法可为 K-means、K-medoids 或 Clara 等
25 任意一种聚类算法。如图 4 所示，中心向量的计算过程包括：

 步骤 S402，确定待形成的每个聚类的初始中心向量，将初始中心向量设置为对应待形成的聚类的当前中心向量。

本实施例中，该初始中心向量可为预先设置的特征向量。该初始中心向量可根据每个聚类的所具有的特征而形成的历史经验值而设置的特征向量，以减少计算机设备的计算量。进一步地还可从每个用户的特征向量中来选取第二数量的特征向量，将选取的每个特征向量设置为对应一个聚类的初始中心向量，使每个聚类均具有一个中心向量，将初始中心向量设置为当前中心向量。同样的，也可以按照该历史经验来选取。将确定初始中心向量即为当前中心向量，用于进行下述步骤的迭代计算，以确定最终聚类 and 每个所确定的最终聚类的中心向量。

步骤 S404，计算每个特征向量与每个当前中心向量的距离，将每个特征向量的用户划分至最小距离对应的聚类中，生成当前聚类。

在生成了当前中心向量之后，可分别计算每个特征向量与每个当前中心向量的距离，根据所计算出的距离将用户进行划分，从而形成了对应第二数量的聚类，将该聚类设置为当前聚类，用于进行下述的迭代。具体地，可针对每个用户的特征向量，计算其与每个当前中心向量距离，确定该第二数量的距离中的最小距离，将该特征向量的用户划分至最小距离所对应的聚类中。针对每个用户的特征向量，重复该过程，以完成对所有用户的划分，从而实现一次聚类的形成。

步骤 S406，计算当前聚类中，每个聚类新的中心向量。

本实施例中，可根据最新形成的聚类，针对每个聚类中的用户的特征向量，计算出该聚类的中心向量，并将新计算出的中心向量替换上一次计算出的中心向量，以进行迭代计算。

其中，可按照预设的中心向量计算模型来计算出每个聚类的中心向量的计算方式。比如可对每个聚类的特征向量，将特征向量中相同位置的参数进行求平均值等运算，以形成对应聚类的新的中心向量中，该参数的数值。

步骤 S408，检测每个新的中心向量是否收敛，若否，则将每个新的中心向量设置为当前中心向量，并继续执行计算每个特征向量与每个当前中心向量的距离，直至每个新的中心向量均收敛。当每个新的中心向量均收敛时，

将该当前聚类设置为确定的最终聚类，将新的中心向量设置为对应确定的最终聚类的中心向量。

本实施例中，可判断每个新的中心向量与前一次的中心向量之间的差值是否小于预设数值。若是，则判定所计算出的中心点收敛，否则，判断不收敛。当判断不收敛时，可将当前计算出的每个聚类的中心点作为对应待形成聚类的当前中心向量，返回上述步骤 S404，继续执行步骤 S404~S408，以再次形成新的聚类和新的聚类的中心向量，直至判定每个最新形成的聚类的中心向量收敛。

具体地，该差值可为距离值。计算机设备中对应预设有一个距离阈值，该预设距离阈值用于作为判断中心向量是否收敛的标准。通过将所计算出的差值与该距离阈值进行比较，当小于该距离阈值时，则判定对应中心向量收敛。当其中的一个或多个中心向量不收敛时，可继续执行步骤 S404，将最新的中心向量设置为对应待形成聚类的初始中心向量，以进行重新聚类。若每个差值均小于距离阈值，则判定所计算出的新的中心点收敛。

当在判断出所有聚类的中心点已均经收敛时，则可终止迭代，将该当前聚类设置为确定的最终聚类，将新的中心向量设置为对应确定的最终聚类的中心向量，完成对所有用户的聚类以及每个聚类的中心向量的计算。

本实施例中，通过判断所形成的中心向量是否收敛，若否，则重新聚类，计算重新形成的聚类的中心向量，并进行中心向量的收敛判断，直至每个中心向量均收敛，将该当前聚类设置为确定的最终聚类，将新的中心向量设置为对应确定的最终聚类的中心向量，完成对所有用户的聚类以及每个聚类的中心向量的计算。可提高对聚类形成的准确度，提高所计算出的聚类中心的准确度，进而提高了对第二用户选取的精准度。

在一个实施例中，针对根据用户的个人信息所形成的用户组，对每个用户组中按照上述的聚类过程形成相应的聚类，以进一步提高对用户划分的准确度，提高对第二用户选取的精准度。

在一个实施例中，如图 5 所示，提供了一种产品信息推送装置，该装置包括：

行为数据获取模块 502，用于获取每个用户在预设时间段内产生的行为数据。

5 话题确定模块 504，用于确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系。

次数计算模块 506，用于根据关联关系计算每个用户在预设时间段内产生的行为数据在每个预设话题上的出现次数。

10 特征向量生成模块 508，用户根据出现次数生成对应用户对每个预设话题的热度的特征向量。

距离计算模块 510，用于计算每个用户的特征向量之间的距离。

信息推送模块 512，用于当获取到第一用户生成了对预设产品的购买信息后，选取与第一用户的距离最短的第一数量的第二用户，向第二用户的终端发送产品的产品信息。

15 在一个实施例中，如图 6 所示，上述装置还包括：

用户划分模块 514，用于获取每个用户的个人信息；根据个人信息计算每个用户之间的相似度。

信息推送模块 512 还用于从与第一用户的相似度大于相似度阈值的用户中选取与第一用户的距离最短的第一数量的第二用户。

20 在一个实施例中，如图 7 所示，预设话题确定模块 504 包括：

关键词句提取单元 702，用于提取每份行为数据中的关键词句；

关联度计算单元 704，用于计算每个关键词句和每个预设话题的第一关联度；根据第一关联度计算出对应行为数据和每个预设话题的第二关联度；

25 话题确定单元 706，用于将超过关联度阈值的第二关联度的预设话题确定为行为数据归属的预设话题。

在一个实施例中，特征向量生成模块 508 还用于根据每个用户的每份行为数据与归属的预设话题的出现次数，以及与归属的预设话题的第二关联度，

计算出每个用户的特征向量。

在一个实施例中，如图 8 所示，上述装置还包括：

聚类模块 516，用于根据距离生成第二数量的聚类，将每个用户划分至对应的一个聚类之中。

5 信息推送模块 512 还用于从与第一用户处于相同聚类的用户中选取与第一用户的距离最短的第一数量的第二用户。

10 在一个实施例中，该聚类模块 516 还用于确定待形成的每个聚类的初始中心向量，将初始中心向量设置为对应待形成的聚类的当前中心向量；计算每个特征向量与每个当前中心向量的距离，将每个特征向量的用户划分至最小距离对应的聚类中，生成当前聚类；计算当前聚类中，每个聚类新的中心向量；检测每个新的中心向量是否收敛，若否，则将每个新的中心向量设置为当前中心向量，并继续执行上述的计算每个特征向量与每个当前中心向量的距离，直至每个新的中心向量均收敛；当每个新的中心向量都收敛时将该当前聚类设置为确定的最终聚类，将新的中心向量设置为对应确定的最终聚类的中心向量。

上述装置中的各个模块和单元可全部或部分通过软件、硬件及其组合来实现。其中，网络接口可以是以太网卡或无线网卡等。上述各模块和单元可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。该处理器可以为中央处理单元（CPU）、微处理器、单片机等。

20 在一个实施例中，本申请提供的产品信息推送装置可以实现为一种计算机程序的形式，该计算机程序可在如图 9 所示的计算机设备上运行，所述计算机设备的非易失性存储介质可存储组成该产品信息推送装置的各个程序模块。比如可包括如图 5 所示的行为数据获取模块 502、话题确定模块 504、次数计算模块 506、特征向量生成模块 508、距离计算模块 510 及信息推送模块 512。各个程序模块中包括计算机可读指令，该计算机可读指令用于使所述计算机设备执行本说明书中描述的本申请各个实施例的数据处理方法中的步

骤，例如，计算机设备可以通过如图 5 所示的产品信息推送装置中的行为数据获取模块 502 接收获取每个用户在预设时间段内产生的行为数据；通过话题确定模块 504 确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系；通过次数计算模块 506 根据关联关系计算每个用户在预设时间段内产生的行为数据在每个预设话题上的出现次数；通过特征向量生成模块 508 根据出现次数生成对应用户对每个预设话题的热度的特征向量；通过距离计算模块 510 计算每个用户的特征向量之间的距离；及通过信息推送模块 512 当获取到第一用户生成了对预设产品的购买信息后，选取与第一用户的距离最短的第一数量的第二用户，向第二用户的终端发送产品的产品信息。

在一个实施例中，提供了一个或多个包含计算机可读指令的非易失性计算机可读存储介质，当该计算机可读指令被一个或多个处理器执行时，使得该处理器执行上述各实施例所提供的产品信息推送方法的步骤。

在一个实施例中，当该计算机可读指令被一个或多个处理器执行时，使得该处理器执行以下步骤：获取每个用户在预设时间段内产生的行为数据；确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系；根据关联关系计算每个用户在预设时间段内产生的行为数据在每个预设话题上的出现次数；根据出现次数生成对应用户对每个预设话题的热度的特征向量；计算每个用户的特征向量之间的距离；当获取到第一用户生成了对预设产品的购买信息后，选取与第一用户的距离最短的第一数量的第二用户，向第二用户的终端发送产品的产品信息。

在一个实施例中，当该计算机可读指令被一个或多个处理器执行时，使得该处理器在执行选取与第一用户的距离最短的第一数量的第二用户之前，还包括执行以下步骤：获取每个用户的个人信息；根据个人信息计算每个用户之间的相似度；所实现的选取与第一用户的距离最短的第一数量的第二用户，包括：从与第一用户的相似度大于相似度阈值的用户中选取与第一用户的距离最短的第一数量的第二用户。

在一个实施例中，当该计算机可读指令被一个或多个处理器执行时，使得该处理器执行的确定每份行为数据归属的预设话题，包括：提取每份行为数据中的关键词句；计算每个关键词句和每个预设话题的第一关联度；根据第一关联度计算出对应行为数据和每个预设话题的第二关联度；将超过关联度阈值的第二关联度的预设话题确定为行为数据归属的预设话题。

在一个实施例中，当该计算机可读指令被一个或多个处理器执行时，使得该处理器执行的根据出现次数生成对应用户对每个预设话题的热度的特征向量，包括：根据每个用户的每份行为数据与归属的预设话题的出现次数，以及与归属的预设话题的第二关联度，计算出每个用户的特征向量。

在一个实施例中，当该计算机可读指令被一个或多个处理器执行时，使得该处理器在执行计算每个用户的特征向量之间的距离之后，还包括执行以下步骤：根据距离生成第二数量的聚类，将每个用户划分至对应的一个聚类之中；所实现的选取与第一用户的距离最短的第一数量的第二用户，包括：从与第一用户处于相同聚类的用户中选取与第一用户的距离最短的第一数量的第二用户。

在一个实施例中，当该计算机可读指令被一个或多个处理器执行时，还使得该处理器执行以下步骤：确定待形成的每个聚类的初始中心向量，将初始中心向量设置为对应待形成的聚类的当前中心向量；计算每个特征向量与每个当前中心向量的距离，将每个特征向量的用户划分至最小距离对应的聚类中，生成当前聚类；计算当前聚类中，每个聚类新的中心向量；检测每个新的中心向量是否收敛，若否，则将每个新的中心向量设置为当前中心向量，并继续执行上述的计算每个特征向量与每个当前中心向量的距离，直至每个新的中心向量均收敛；当每个新的中心向量都收敛时将该当前聚类设置为确定的最终聚类，将新的中心向量设置为对应确定的最终聚类的中心向量。

在一个实施例中，提供了一种计算机设备，包括存储器及处理器，存储器中储存有计算机可读指令，该指令被处理器执行时，使得处理器执行上述

各实施例所提供的产品信息推送方法的步骤。

该计算机设备可为终端或服务器，如图 9 所示，提供了该计算设备的一种内部结构示意图。该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中，该处理器用于提供计算和控制能力，支撑整个计算机设备的运行。存储器用于存储数据、指令代码等。存储器上存储至少一个计算机可读指令，该计算机可读指令可被处理器执行，以实现本申请实施例中提供的适用于该计算机设备的产品信息推送方法。存储器可包括磁碟、光盘、只读存储记忆体（Read-Only Memory, ROM）等非易失性存储介质。例如，在一个实施例中，存储器包括非易失性存储介质及内存储器。该非易失性存储介质存储有操作系统、数据库和计算机可读指令。该数据库中存储有用于实现以上各个实施例所提供的一种产品信息推送方法相关的数据，比如可存储有用户的行为数据。该计算机可读指令可被处理器所执行，以用于实现以上各个实施例所提供的一种产品信息推送方法。该内存储器为非易失性存储介质中的操作系统、数据库和计算机可读指令提供高速缓存的运行环境。网络接口可以是以太网卡或无线网卡等，用于与外部的终端或计算机设备进行通信，如将所生成的对比结果发送至预设的测试终端。当该计算机设备为服务器时，还可由独立的服务器或者是多个服务器组成的服务器集群来实现。本领域技术人员可以理解，图 8 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

在一个实施例中，该指令被处理器执行时，使得处理器执行以下步骤：获取每个用户在预设时间段内产生的行为数据；确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系；根据关联关系计算每个用户在预设时间段内产生的行为数据在每个预设话题上的出现次数；根据出现次数生成对应用户对每个预设话题的热度的特征向量；计算每个用户的特征向量之间的距离；当获取到第一用户生成了对预设产品的购买信息

后，选取与第一用户的距离最短的第一数量的第二用户，向第二用户的终端发送产品的产品信息。

在一个实施例中，该指令被处理器执行时，使得处理器执行选取与第一用户的距离最短的第一数量的第二用户之前，还包括执行以下步骤：获取每个用户的个人信息；根据个人信息计算每个用户之间的相似度；所实现的选取与第一用户的距离最短的第一数量的第二用户，包括：从与第一用户的相似度大于相似度阈值的用户中选取与第一用户的距离最短的第一数量的第二用户。

在一个实施例中，该指令被处理器执行时，使得处理器执行的确定每份行为数据归属的预设话题，包括：提取每份行为数据中的关键词句；计算每个关键词句和每个预设话题的第一关联度；根据第一关联度计算出对应行为数据和每个预设话题的第二关联度；将超过关联度阈值的第二关联度的预设话题确定为行为数据归属的预设话题。

在一个实施例中，该指令被处理器执行时，使得处理器执行的根据出现次数生成对应用户对每个预设话题的热度的特征向量，包括：根据每个用户的每份行为数据与归属的预设话题的出现次数，以及与归属的预设话题的第二关联度，计算出每个用户的特征向量。

在一个实施例中，该指令被处理器执行时，使得处理器在执行计算每个用户的特征向量之间的距离之后，还包括执行以下步骤：根据距离生成第二数量的聚类，将每个用户划分至对应的一个聚类之中；所实现的选取与第一用户的距离最短的第一数量的第二用户，包括：从与第一用户处于相同聚类的用户中选取与第一用户的距离最短的第一数量的第二用户。

在一个实施例中，该指令被处理器执行时，还使得处理器执行以下步骤：确定待形成的每个聚类的初始中心向量，将初始中心向量设置为对应待形成的聚类的当前中心向量；计算每个特征向量与每个当前中心向量的距离，将每个特征向量的用户划分至最小距离对应的聚类中，生成当前聚类；计算当前聚类中，每个聚类新的中心向量；检测每个新的中心向量是否收敛，若否，

则将每个新的中心向量设置为当前中心向量，并继续执行上述的计算每个特征向量与每个当前中心向量的距离，直至每个新的中心向量均收敛；当每个新的中心向量都收敛时将该当前聚类设置为确定的最终聚类，将新的中心向量设置为对应确定的最终聚类的中心向量。

5

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读取存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，所述的存储介质可为
10 磁碟、光盘、只读存储记忆体（Read-Only Memory, ROM）等。

以上所述实施例的各技术特征可以进行任意的组合，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。

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发明专利范围的限制。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
15 变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种产品信息推送方法，包括：

获取每个用户在预设时间段内产生的行为数据；

确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系；

5 根据所述关联关系计算每个用户在所述预设时间段内产生的行为数据在每个预设话题上的出现次数；

根据所述出现次数生成对应用户对每个预设话题的热度的特征向量；

计算每个用户的特征向量之间的距离；及

10 当获取到第一用户生成了对预设产品的购买信息后，选取与所述第一用户的距离最短的第一数量的第二用户，向所述第二用户的终端发送所述产品的产品信息。

2、根据权利要求1所述的方法，其特征在于，在所述选取与所述第一用户的距离最短的第一数量的第二用户之前，还包括：

获取每个用户的个人信息；及

15 根据所述个人信息计算每个用户之间的相似度；

所述选取与所述第一用户的距离最短的第一数量的第二用户，包括：

从与所述第一用户的相似度大于相似度阈值的用户中选取与所述第一用户的距离最短的第一数量的第二用户。

20 3、根据权利要求1所述的方法，其特征在于，所述确定每份行为数据归属的预设话题，包括：

提取每份行为数据中的关键词句；

计算每个关键词句和每个预设话题的第一关联度；

根据所述第一关联度计算出对应行为数据和每个预设话题的第二关联度；及

25 将超过关联度阈值的第二关联度的预设话题确定为行为数据归属的预设话题。

4、根据权利要求3所述的方法，其特征在于，所述根据所述出现次数生成对应用户对每个预设话题的热度的特征向量，包括：

根据每个用户的每份行为数据与归属的预设话题的出现次数，以及与归属的预设话题的第二关联度，计算出每个用户的特征向量。

5 5、根据权利要求1所述的方法，其特征在于，在所述计算每个用户的特征向量之间的距离之后，还包括：

根据所述距离生成第二数量的聚类，将每个用户划分至对应的一个聚类之中；

所述选取与所述第一用户的距离最短的第一数量的第二用户，包括：

10 从与所述第一用户处于相同聚类的用户中选取与所述第一用户的距离最短的第一数量的第二用户。

6、一种产品信息推送装置，包括：

行为数据获取模块，用于获取每个用户在预设时间段内产生的行为数据；

15 话题确定模块，用于确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系；

次数计算模块，用于根据所述关联关系计算每个用户在所述预设时间段内产生的行为数据在每个预设话题上的出现次数；

特征向量生成模块，用户根据所述出现次数生成对应用户对每个预设话题的热度的特征向量；

20 距离计算模块，用于计算每个用户的特征向量之间的距离；及

信息推送模块，用于当获取到第一用户生成了对预设产品的购买信息后，选取与所述第一用户的距离最短的第一数量的第二用户，向所述第二用户的终端发送所述产品的产品信息。

7、根据权利要求6所述的装置，其特征在于，所述装置还包括：

25 用户划分模块，用于获取每个用户的个人信息；根据所述个人信息计算每个用户之间的相似度；

所述信息推送模块还用于从与所述第一用户的相似度大于相似度阈值的

用户中选取与所述第一用户的距离最短的第一数量的第二用户。

8、根据权利要求6所述的装置，其特征在于，所述预设话题确定模块包括：

关键词句提取单元，用于提取每份行为数据中的关键词句；

5 关联度计算单元，用于计算每个关键词句和每个预设话题的第一关联度；
根据所述第一关联度计算出对应行为数据和每个预设话题的第二关联度；及
话题确定单元，用于将超过关联度阈值的第二关联度的预设话题确定为
行为数据归属的预设话题。

9、根据权利要求8所述的装置，其特征在于，所述特征向量生成模块还
10 用于根据每个用户的每份行为数据与归属的预设话题的出现次数，以及与归
属的预设话题的第二关联度，计算出每个用户的特征向量。

10、根据权利要求6所述的装置，其特征在于，所述装置还包括：

聚类模块，用于根据距离生成第二数量的聚类，将每个用户划分至对应
的一个聚类之中；

15 所述信息推送模块还用于从与第一用户处于相同聚类的用户中选取与第
一用户的距离最短的第一数量的第二用户。

11、一个或多个存储有计算机可读指令的计算机可读非易失性存储介质，
所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理
20 器执行以下步骤：

获取每个用户在预设时间段内产生的行为数据；

确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题
建立关联关系；

根据所述关联关系计算每个用户在所述预设时间段内产生的行为数据在
25 每个预设话题上的出现次数；

根据所述出现次数生成对应用户对每个预设话题的热度的特征向量；

计算每个用户的特征向量之间的距离；及

当获取到第一用户生成了对预设产品的购买信息后，选取与所述第一用户的距离最短的第一数量的第二用户，向所述第二用户的终端发送所述产品的产品信息。

12、根据权利要求 11 所述的存储介质，其特征在于，所述计算机可读指令被一个或多个处理器执行时，还使得所述一个或多个处理器执行以下步骤：

获取每个用户的个人信息；

根据所述个人信息计算每个用户之间的相似度；及

从与所述第一用户的相似度大于相似度阈值的用户中选取与所述第一用户的距离最短的第一数量的第二用户。

13、根据权利要求 11 所述的存储介质，其特征在于，所述计算机可读指令被一个或多个处理器执行时，还使得所述一个或多个处理器执行以下步骤：

提取每份行为数据中的关键词句；

计算每个关键词句和每个预设话题的第一关联度；

根据所述第一关联度计算出对应行为数据和每个预设话题的第二关联度；及

将超过关联度阈值的第二关联度的预设话题确定为行为数据归属的预设话题。

14、根据权利要求 13 所述的存储介质，其特征在于，所述计算机可读指令被一个或多个处理器执行时，还使得所述一个或多个处理器执行以下步骤：

根据每个用户的每份行为数据与归属的预设话题的出现次数，以及与归属的预设话题的第二关联度，计算出每个用户的特征向量。

15、根据权利要求 11 所述的存储介质，其特征在于，所述计算机可读指令被一个或多个处理器执行时，还使得所述一个或多个处理器执行以下步骤：

根据所述距离生成第二数量的聚类，将每个用户划分至对应的一个聚类之中；及

从与所述第一用户处于相同聚类的用户中选取与所述第一用户的距离最短的第一数量的第二用户。

16、一种计算机设备，包括存储器及处理器，所述存储器中储存有计算机可读指令，所述指令被所述处理器执行时，使得所述处理器执行以下步骤：

获取每个用户在预设时间段内产生的行为数据；

5 确定每份行为数据归属的预设话题，将每份行为数据和归属的预设话题建立关联关系；

根据所述关联关系计算每个用户在所述预设时间段内产生的行为数据在每个预设话题上的出现次数；

根据所述出现次数生成对应用户对每个预设话题的热度的特征向量；

10 计算每个用户的特征向量之间的距离；及

当获取到第一用户生成了对预设产品的购买信息后，选取与所述第一用户的距离最短的第一数量的第二用户，向所述第二用户的终端发送所述产品的产品信息。

17、根据权利要求 16 所述的计算机设备，其特征在于，所述指令被所述
15 处理器执行时，还使得所述处理器执行以下步骤：

获取每个用户的个人信息；

根据所述个人信息计算每个用户之间的相似度；及

从与所述第一用户的相似度大于相似度阈值的用户中选取与所述第一用户的距离最短的第一数量的第二用户。

20 18、根据权利要求 16 所述的计算机设备，其特征在于，所述指令被所述处理器执行时，还使得所述处理器执行以下步骤：

提取每份行为数据中的关键词句；

计算每个关键词句和每个预设话题的第一关联度；

25 根据所述第一关联度计算出对应行为数据和每个预设话题的第二关联度；及

将超过关联度阈值的第二关联度的预设话题确定为行为数据归属的预设话题。

19、根据权利要求 18 所述的计算机设备，其特征在于，所述指令被所述处理器执行时，还使得所述处理器执行以下步骤：

根据每个用户的每份行为数据与归属的预设话题的出现次数，以及与归属的预设话题的第二关联度，计算出每个用户的特征向量。

5 20、根据权利要求 16 所述的计算机设备，其特征在于，所述指令被所述处理器执行时，还使得所述处理器执行以下步骤：

根据所述距离生成第二数量的聚类，将每个用户划分至对应的一个聚类之中；及

10 从与所述第一用户处于相同聚类的用户中选取与所述第一用户的距离最短的第一数量的第二用户。

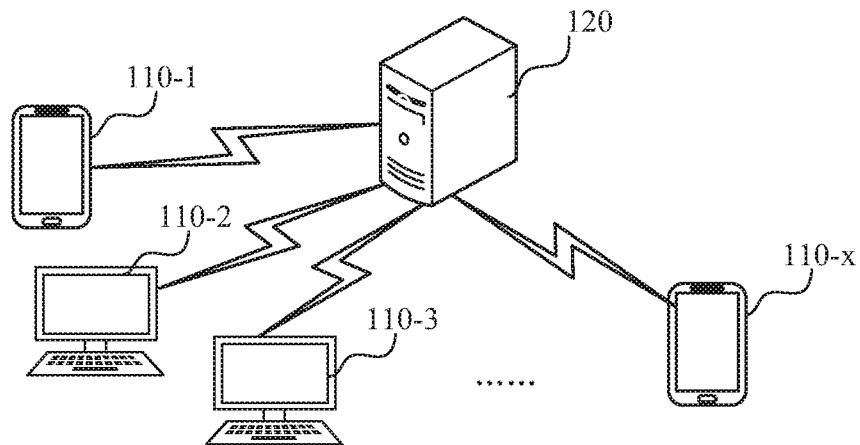


图 1

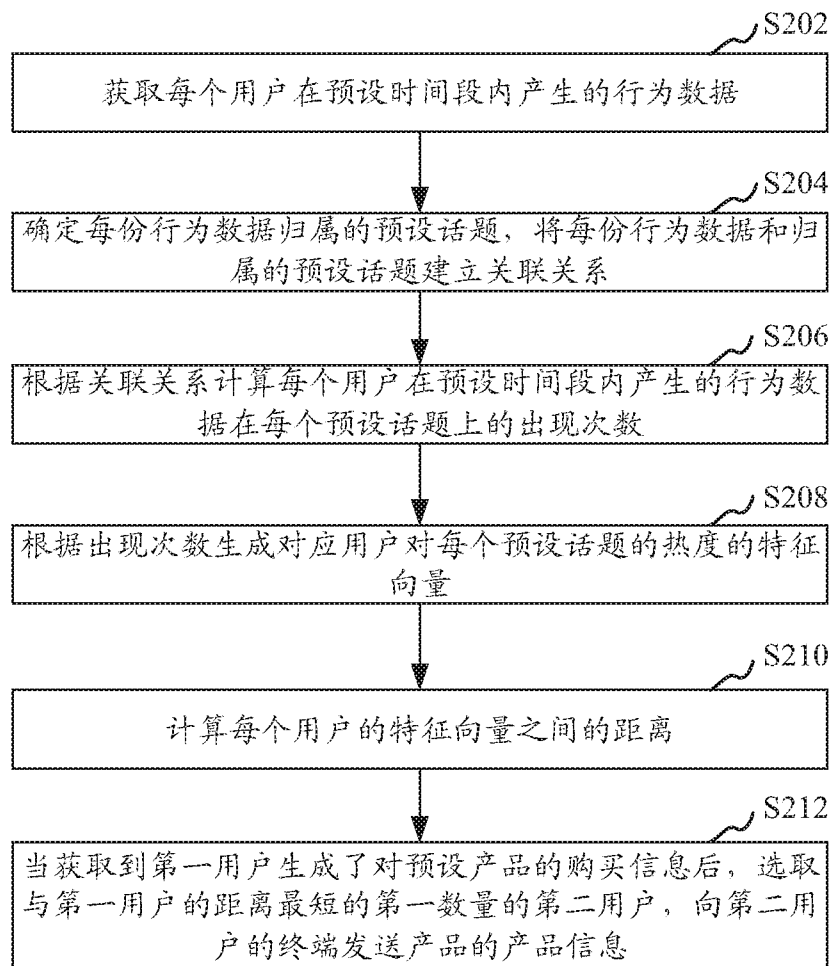


图 2

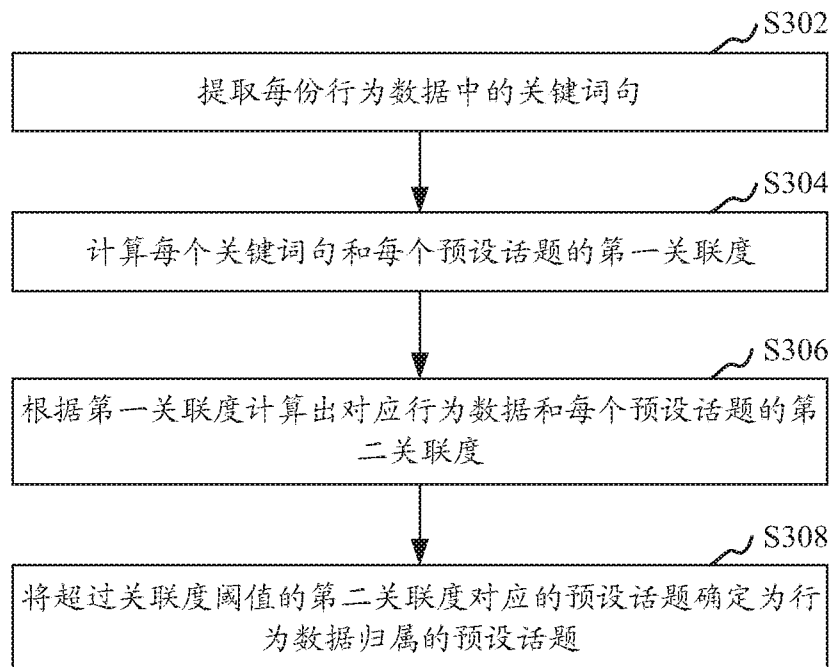


图 3

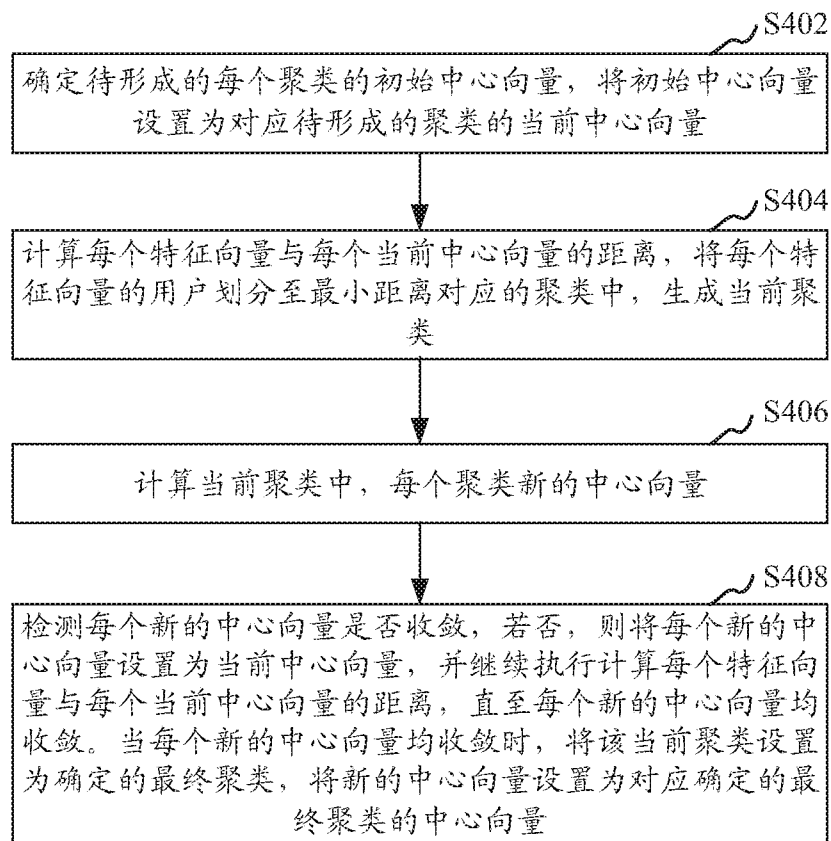


图 4

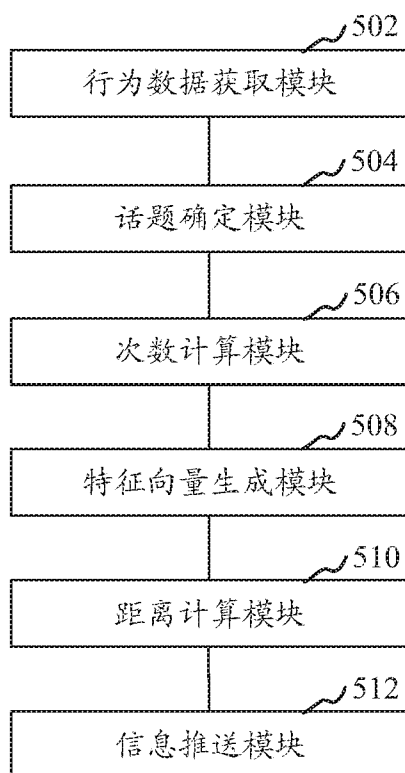


图 5

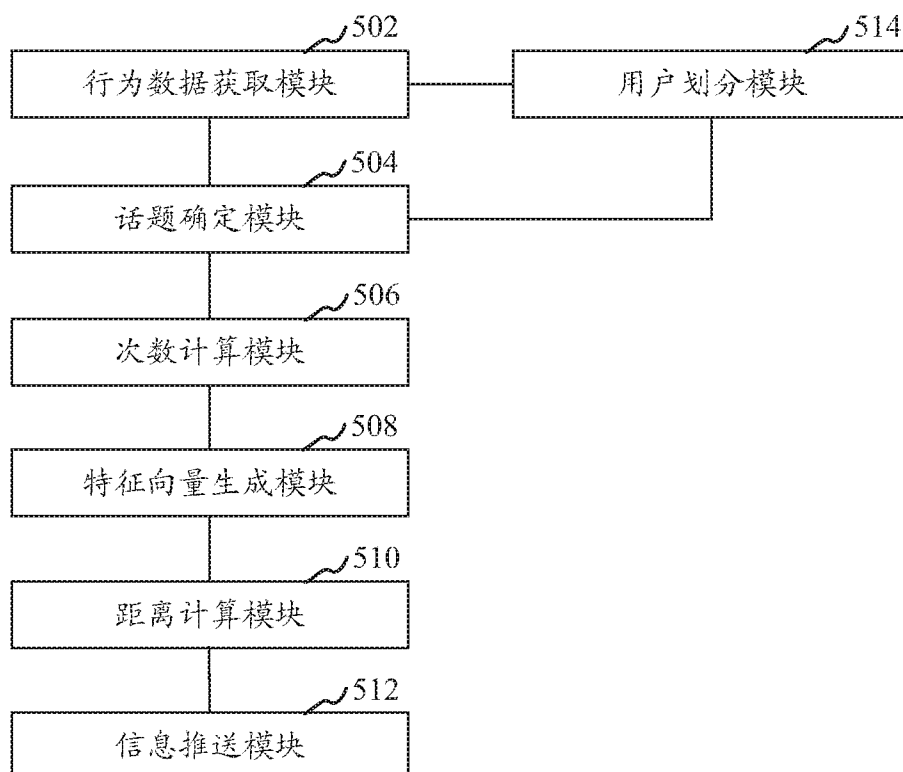


图 6

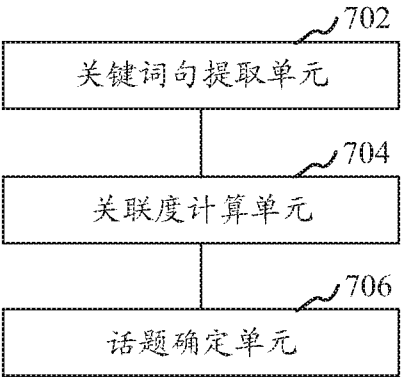


图 7

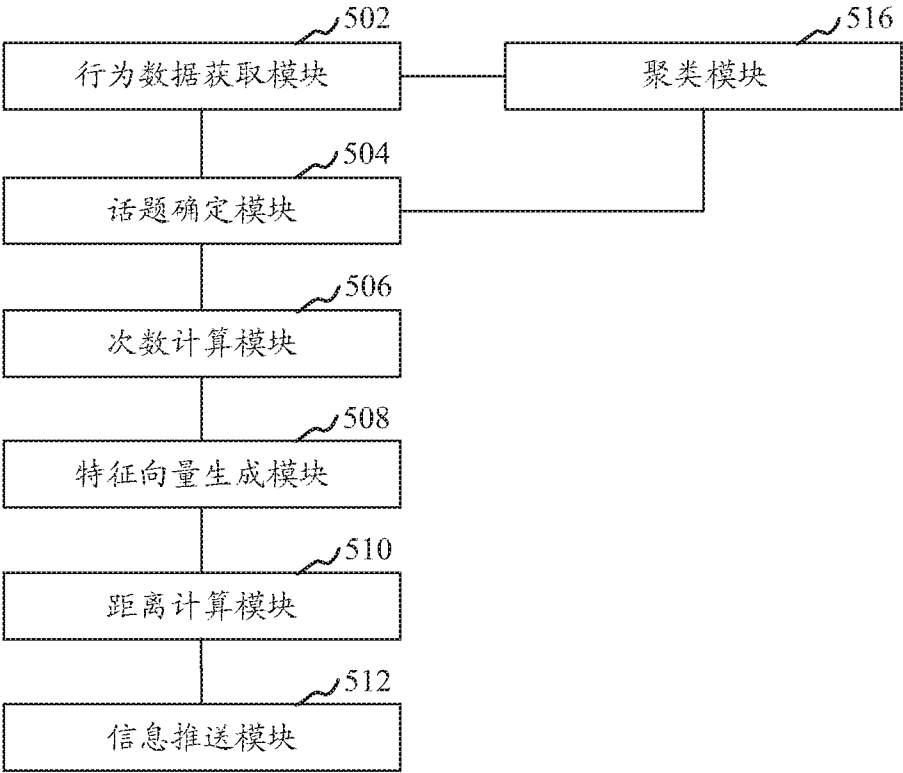


图 8

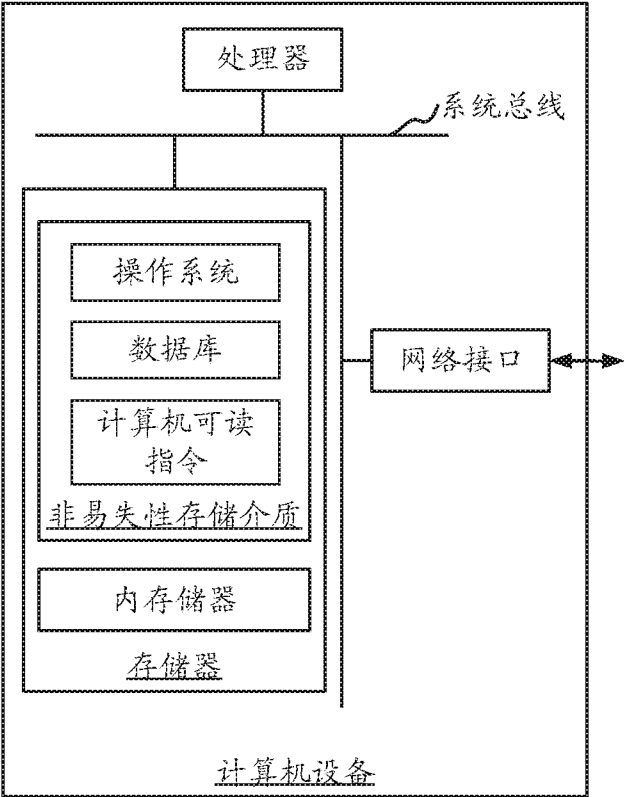


图 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/103989

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 30/06 (2012.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06Q; G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC: 产品, 商品, 推送, 推荐, 行为, 话题, 次数, 热度, 特征向量, 距离, product, commodity, goods, deliver, recommend, behavior, topic, time, count, heat, feature, vector, distance

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106355449 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 25 January 2017 (25.01.2017), the abstract, and description, paragraphs [0036]-[0055] and [0147]	1-20
A	CN 103377250 A (HANGZHOU ZAIYAN NETWORK TECHNOLOGY CO., LTD.) 30 October 2013 (30.10.2013), entire document	1-20
A	CN 104281882 A (INSTITUTE OF INFORMATION ENGINEERING, CHINESE ACADEMY OF SCIENCES) 14 January 2015 (14.01.2015), entire document	1-20
A	CN 103853763 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 11 June 2014 (11.06.2014), entire document	1-20
A	US 2011213786 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 01 September 2011 (01.09.2011), entire document	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	
“A” document defining the general state of the art which is not considered to be of particular relevance	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“E” earlier application or patent but published on or after the international filing date	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	“&” document member of the same patent family

Date of the actual completion of the international search 15 March 2018	Date of mailing of the international search report 28 March 2018
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer JIANG, Lingling Telephone No. (86-10) 53961421

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/103989

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 106355449 A	25 January 2017	None	
CN 103377250 A	30 October 2013	None	
CN 104281882 A	14 January 2015	None	
CN 103853763 A	11 June 2014	None	
US 2011213786 A1	01 September 2011	CN 102169566 A	31 August 2011

国际检索报告

国际申请号

PCT/CN2017/103989

A. 主题的分类

G06Q 30/06 (2012.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06Q; G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNPAT, CNKI, WPI, EPDOC: 产品, 商品, 推送, 推荐, 行为, 话题, 次数, 热度, 特征向量, 距离, product, commodity, goods, deliver, recommend, behavior, topic, time, count, heat, feature, vector, distance

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 106355449 A (腾讯科技深圳有限公司) 2017年 1月 25日 (2017 - 01 - 25) 摘要, 说明书第36-55, 147段	1-20
A	CN 103377250 A (杭州载言网络技术有限公司) 2013年 10月 30日 (2013 - 10 - 30) 全文	1-20
A	CN 104281882 A (中国科学院信息工程研究所) 2015年 1月 14日 (2015 - 01 - 14) 全文	1-20
A	CN 103853763 A (腾讯科技深圳有限公司) 2014年 6月 11日 (2014 - 06 - 11) 全文	1-20
A	US 2011213786 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2011年 9月 1日 (2011 - 09 - 01) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2018年 3月 15日

国际检索报告邮寄日期

2018年 3月 28日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局 (ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

姜玲玲

传真号 (86-10) 62019451

电话号码 (86-10) 010-53961421

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/103989

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	106355449	A	2017年 1月 25日	无			
CN	103377250	A	2013年 10月 30日	无			
CN	104281882	A	2015年 1月 14日	无			
CN	103853763	A	2014年 6月 11日	无			
US	2011213786	A1	2011年 9月 1日	CN	102169566	A	2011年 8月 31日