(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

(43) 国际公布日
2019 年 6 月 27 日 (27.06.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/119635 A1

(51) 国际专利分类号:
G06Q 30/02 (2012.01)

(21) 国际申请号: PCT/CN2018/076181

(22) 国际申请日: 2018 年 2 月 10 日 (10.02.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201711364792.0 2017年12月18日 (18.12.2017) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO.,LTD.) [CN/CN]; 中国广东省深圳市福田区八卦岭工业区平安大厦六楼, Guangdong 518000 (CN)。

(72) 发明人: 安欣 (AN, Xin); 中国广东省深圳市福田区八卦岭工业区平安大厦六楼, Guangdong 518000 (CN)。 许开河 (XU, KaiHe); 中国广东省深圳市福田区八卦岭工业区平安大厦六楼, Guangdong 518000 (CN)。 王建明 (WANG, JianMing); 中国广东省深圳市福田区八卦岭工业区平安大厦六楼, Guangdong 518000 (CN)。 肖京 (XIAO, Jing); 中国广东省深圳市福田区八卦岭工业区平安大厦六楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市沃德知识产权代理事务所（普通合伙） (SHENZHEN WORLD INTELLECTUAL PROPERTY AGENCY (GENERAL PARTNERSHIP)); 中国广东省深圳市福田区园岭街道八卦四路 10 号中浩大厦 1528-1530 室于志光, Guangdong 518000 (CN)。

(54) **Title:** SEED USER DEVELOPMENT METHOD, ELECTRONIC DEVICE AND COMPUTER-READABLE STORAGE MEDIUM

(54) **发明名称:** 种子用户拓展方法、电子设备及计算机可读存储介质

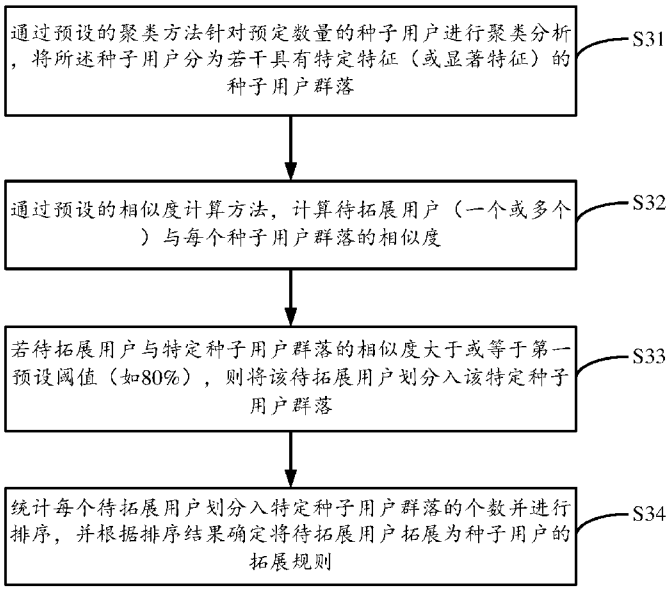


图 3

(57) **Abstract:** Disclosed is a seed user development method, the method comprising the steps of: carrying out clustering analysis on a pre-determined number of seed users by using a pre-set clustering method, and dividing the seed users into several seed user communities having specific features; calculating, by using a pre-set similarity calculation method, the similarity between a user to be developed and each seed user community; if the similarity between the user to be developed and the specific seed user community is greater than or equal to a first pre-set threshold value, classifying the user to be developed into the specific seed user community; and

WO 2019/119635 A1

[见续页]

(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, 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))。

compiling statistics of the number of specific seed user communities, into which each user to be developed is classified, and ranking same, and determining, according to a ranking result, a development rule for developing a user to be developed into a seed user. The present application can reduce the calculation complexity of seed user development and improve the accuracy of development.

(57) 摘要：本申请公开了一种种子用户拓展方法，该方法包括步骤：通过预设的聚类方法针对预定数量的种子用户进行聚类分析，将所述种子用户分为若干具有特定特征的种子用户群落；通过预设的相似度计算方法，计算待拓展用户与每个种子用户群落的相似度；若待拓展用户与特定种子用户群落的相似度大于或等于第一预设阈值，则将该待拓展用户划分入该特定种子用户群落；统计每个待拓展用户划分入特定种子用户群落的个数并进行排序，并根据排序结果确定将待拓展用户拓展为种子用户的拓展规则。本申请可以减少种子用户拓展的计算复杂度、提高拓展的准确度。

种子用户拓展方法、电子设备及计算机可读存储介质

本申请要求于2017年12月18日提交中国专利局、申请号为201711364792.0、发明名称为“种子用户拓展方法、电子设备及计算机可读存储介质”的中国专利申请的优先权，其全部内容通过引用结合在申请中。

技术领域

本申请涉及计算机信息技术领域，尤其涉及一种种子用户拓展方法、电子设备及计算机可读存储介质。

背景技术

目前，对于种子用户拓展通常是通过寻找相似用户的方法，该方法通过计算每个用户间的相似程度进行用户拓展。然而，如果用户数据为大数据量级，则计算复杂度成指数增长，如何减少计算复杂度并高效进行种子用户拓展是目前急需解决的技术问题。故，现有技术中的种子用户拓展方法设计不够合理，亟需改进。

发明内容

有鉴于此，本申请提出一种种子用户拓展方法、电子设备及计算机可读存储介质，通过无监督学习聚类方法与距离相似度算法相结合，减少了种子用户拓展的计算复杂度，并提高了拓展的准确度。

首先，为实现上述目的，本申请提出一种电子设备，所述电子设备包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的种子用户拓展系统，所述种子用户拓展系统被所述处理器执行时实现如下步骤：

通过预设的聚类方法针对预定数量的种子用户进行聚类分析，将所述种子用户分为若干具有特定特征的种子用户群落；

通过预设的相似度计算方法，计算待拓展用户与每个种子用户群落的相似度；

若待拓展用户与特定种子用户群落的相似度大于或等于第一预设阈值，则将该待拓展用户划分入该特定种子用户群落；及

统计每个待拓展用户划分入特定种子用户群落的个数并进行排序，并根据排序结果确定将待拓展用户拓展为种子用户的拓展规则。

优选地，所述特定特征包括用户的地理位置、是否为注册用户、是否购买过特定产品。

优选地，所述计算待拓展用户与每个种子用户群落的相似度包括：计算待拓展用户与每个种子用户群落中心点的相似度，作为待拓展用户与每个种子用户群落的相似度。

优选地，所述拓展规则设置为：根据个数从高到低的顺序，选取指定数量的待拓展用户，将选取的待拓展用户拓展为种子用户。

优选地，所述拓展规则设置为：若待拓展用户划分入特定种子用户群落的个数大于或等于第二预设阈值，则将该待拓展用户拓展为种子用户，其中，所述第二预设阈值设置为所有种子用户群落总个数的预定比例。

此外，为实现上述目的，本申请还提供一种种子用户拓展方法，该方法应用于电子设备，所述方法包括：

通过预设的聚类方法针对预定数量的种子用户进行聚类分析，将所述种子用户分为若干具有特定特征的种子用户群落；

通过预设的相似度计算方法，计算待拓展用户与每个种子用户群落的相似度；

若待拓展用户与特定种子用户群落的相似度大于或等于第一预设阈值，则将该待拓展用户划分入该特定种子用户群落；及

统计每个待拓展用户划分入特定种子用户群落的个数并进行排序，并根

据排序结果确定将待拓展用户拓展为种子用户的拓展规则。

优选地，所述特定特征包括用户的地理位置、是否为注册用户、是否购买过特定产品；及

所述计算待拓展用户与每个种子用户群落的相似度包括：

计算待拓展用户与每个种子用户群落中心点的相似度，作为待拓展用户与每个种子用户群落的相似度。

优选地，所述拓展规则设置为：根据个数从高到低的顺序，选取指定数量的待拓展用户，将选取的待拓展用户拓展为种子用户。

优选地，所述拓展规则设置为：若待拓展用户划分入特定种子用户群落的个数大于或等于第二预设阈值，则将该待拓展用户拓展为种子用户，其中，所述第二预设阈值设置为所有种子用户群落总个数的预定比例。

进一步地，为实现上述目的，本申请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种子用户拓展系统，所述种子用户拓展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种子用户拓展方法的步骤。

相较于现有技术，本申请所提出的电子设备、种子用户拓展方法及计算机可读存储介质，通过无监督学习聚类方法与距离相似度算法相结合，减少了种子用户拓展的计算复杂度，并提高了拓展的准确度。

附图说明

图1是本申请电子设备一可选的硬件架构的示意图；

图2是本申请电子设备中种子用户拓展系统一实施例的程序模块示意图；

图3为本申请种子用户拓展方法一实施例的实施流程示意图。

附图标记：

电子设备	2
存储器	21
处理器	22
网络接口	23
种子用户拓展系统	20
分析模块	201
计算模块	202
拓展模块	203
流程步骤	S31-S34

本申请目的的实现、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

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用以解释本申请，并不用于限定本申请。基于本申请中的实施例，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，都属于本申请保护的范围。

需要说明的是，在本申请中涉及“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，也不在本申请要求的保护范围之内。

进一步需要说明的是，在本文中，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

首先，本申请提出一种电子设备 2。

参阅图 1 所示，是本申请电子设备 2 一可选的硬件架构的示意图。本实施例中，所述电子设备 2 可包括，但不限于，可通过系统总线相互通信连接存储器 21、处理器 22、网络接口 23。需要指出的是，图 1 仅示出了具有组件 21-23 的电子设备 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。

其中，所述电子设备 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，该电子设备 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。

所述存储器 21 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 21 可以是所述电子设备 2 的内部存储单元，例如该电子设备 2 的硬盘或内存。在另一些实施例中，所述存储器 21 也可以是所述电子设备 2 的外部存储设备，例如该电子设备 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 21 还可以既包括所述电子设备 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 21 通常用于存储安装于所述电子设备 2 的操作系统和各类应用软件，例如所述种子用户拓展系统 20 的程序代码等。此外，所述存储器 21 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。

所述处理器 22 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 22 通常用于控制所述电子设备 2 的总体操作，例如执行与所述电子设备 2 进行数据交互或者通信相关的控制和处理等。本实施例中，所述处理器 22 用于运行所述存储器 21 中存储的程序代码或者处理数据，例如运行所述的种子用户拓展系统 20 等。

所述网络接口 23 可包括无线网络接口或有线网络接口，该网络接口 23 通常用于在所述电子设备 2 与其他电子设备之间建立通信连接。例如，所述网络接口 23 用于通过网络将所述电子设备 2 与外部数据平台相连，在所述电子设备 2 与外部数据平台之间建立数据传输通道和通信连接。所述网络可以是企业内部网（Intranet）、互联网（Internet）、全球移动通讯系统（Global System of Mobile communication，GSM）、宽带码分多址（Wideband Code Division Multiple Access，WCDMA）、4G 网络、5G 网络、蓝牙（Bluetooth）、Wi-Fi 等无线或有线网络。

至此，已经详细介绍了本申请各个实施例的应用环境和相关设备的硬件结构和功能。下面，将基于上述应用环境和相关设备，提出本申请的各个实施例。

参阅图 2 所示，是本申请电子设备 2 中种子用户拓展系统 20 一实施例的程序模块图。本实施例中，所述的种子用户拓展系统 20 可以被分割成一个或多个程序模块，所述一个或者多个程序模块被存储于所述存储器 21 中，并由一个或多个处理器（本实施例中为所述处理器 22）所执行，以完成本申请。例如，在图 2 中，所述的种子用户拓展系统 20 可以被分割成分析模块 201、

计算模块 202、以及拓展模块 203。本申请所称的程序模块是指能够完成特定功能的一系列计算机程序指令段，比程序更适合于描述所述种子用户拓展系统 20 在所述电子设备 2 中的执行过程。以下将就各程序模块 201-203 的功能进行详细描述。

所述分析模块201，用于通过预设的聚类方法（如无监督学习K-means聚类方法）针对预定数量（可以是大数据量级）的种子用户进行聚类分析，将所述种子用户分为若干具有特定特征（或显著特征）的种子用户群落。其中，所述特定特征包括，但不限于，用户的地理位置、是否为注册用户、是否购买过特定产品（如产险）等。

举例而言，假设聚类个数为k（即种子用户群落的个数），所有种子用户的预定数量为N，则K-means聚类方法包括如下步骤：

（A1）首先从N个数据对象（即N个种子用户）中任意选择 k 个数据对象作为聚类中心（即初始聚类中心）；

（A2）针对剩余的数据对象，计算每个剩余的数据对象到每个聚类中心的相似度（如欧式距离，距离越小代表相似度越高），并根据相似度的大小分别将每个剩余的数据对象分配给与其最相似的聚类中心所代表的聚类；

（A3）重新计算已经得到的每个聚类的聚类中心（即每个聚类中所有数据对象的均值）；

（A4）迭代A2至A3步骤，直至预设的标准测度函数开始收敛为止。在本实施例中，可以采用均方差作为预设的标准测度函数。

所述计算模块202，用于通过预设的相似度计算方法，计算待拓展用户（一个或多个）与每个种子用户群落（即具有特定特征的种子用户群落）的相似度。其中，所述预设的相似度计算方法可以是：欧氏距离、夹角余弦、汉明距离等相似度计算方法。

优选地，在本实施例中，所述计算待拓展用户与每个种子用户群落的相

似度包括：计算待拓展用户与每个种子用户群落中心点的相似度（计算方法与用户之间相似度计算方法一致），作为待拓展用户与每个种子用户群落的相似度。其中，种子用户群落是多个具有相似特征的用户集合，每个集合向一个中心点聚合，该中心点即种子用户群落的中心点。由于无需与种子用户群落中的每个用户计算相似度，从而可以大幅降低计算复杂度。

举例而言，采用夹角余弦方法计算待拓展用户与某个种子用户群落中心点的相似度可以采用如下公式1所示。

$$\cos(\vec{a},\vec{b}) = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}|\cdot|\vec{b}|} \qquad \text{公式1}$$

其中，$\cos(\vec{a},\vec{b})$代表待拓展用户a和某个种子用户群落中心点b的余弦相似度，$\vec{a}$代表待拓展用户a的评分向量，$\vec{b}$代表某个种子用户群落中心点b的评分向量。

所述拓展模块203，用于若待拓展用户与特定种子用户群落的相似度大于或等于第一预设阈值（如80%），则将该待拓展用户划分入该特定种子用户群落。

举例而言，假设聚类分析后得到的种子用户群落包括三个：B1、B2、B3，待拓展用户A与种子用户群落B1的相似度S1为60%、与种子用户群落B2的相似度S2为85%（大于第一预设阈值）、与种子用户群落B3的相似度S3为90%（大于第一预设阈值），则将待拓展用户A划分入特定种子用户群落B2和B3。

所述拓展模块203，还用于统计每个待拓展用户划分入特定种子用户群落的个数并进行排序，并根据排序结果确定将待拓展用户拓展为种子用户的拓展规则。其中，排序越高代表相似度越高。

优选地，在本实施例中，所述拓展规则设置为：根据个数从高到低的顺序，选取指定数量（如前2位）的待拓展用户，将选取的待拓展用户拓展为种

子用户。例如，假设待拓展用户A同时划分入5个种子用户群落，待拓展用户B同时划分入3个种子用户群落，待拓展用户C同时划分入2个种子用户群落，则将待拓展用户A和B拓展为种子用户。

优选地，在其它实施例中，所述拓展规则还可以设置为：若待拓展用户划分入特定种子用户群落的个数大于或等于第二预设阈值，则将该待拓展用户拓展为种子用户，其中，所述第二预设阈值可以设置为所有种子用户群落（即具有特定特征的种子用户群落）总个数的预定比例（如50%）。例如，假设所有种子用户群落总个数为4，预定比例为50%，则所述第二预设阈值为2。

通过上述程序模块201-203，本申请所提出的种子用户拓展系统20，通过无监督学习聚类方法与距离相似度算法相结合，减少了种子用户拓展的计算复杂度，并提高了拓展的准确度。

此外，本申请还提出一种种子用户拓展方法。

参阅图3所示，是本申请种子用户拓展方法一实施例的实施流程示意图。在本实施例中，根据不同的需求，图3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。

步骤S31，通过预设的聚类方法（如无监督学习K-means聚类方法）针对预定数量（可以是大数据量级）的种子用户进行聚类分析，将所述种子用户分为若干具有特定特征（或显著特征）的种子用户群落。其中，所述特定特征包括，但不限于，用户的地理位置、是否为注册用户、是否购买过特定产品（如产险）等。

举例而言，假设聚类个数为k（即种子用户群落的个数），所有种子用户的预定数量为N，则K-means聚类方法包括如下步骤：

（A1）首先从N个数据对象（即N个种子用户）中任意选择 k 个数据对象作为聚类中心（即初始聚类中心）；

（A2）针对剩余的数据对象，计算每个剩余的数据对象到每个聚类中心的相似度（如欧式距离，距离越小代表相似度越高），并根据相似度的大小分别将每个剩余的数据对象分配给与其最相似的聚类中心所代表的聚类；

（A3）重新计算已经得到的每个聚类的聚类中心（即每个聚类中所有数据对象的均值）；

（A4）迭代A2至A3步骤，直至预设的标准测度函数开始收敛为止。在本实施例中，可以采用均方差作为预设的标准测度函数。

步骤S32，通过预设的相似度计算方法，计算待拓展用户（一个或多个）与每个种子用户群落（即具有特定特征的种子用户群落）的相似度。其中，所述预设的相似度计算方法可以是：欧氏距离、夹角余弦、汉明距离等相似度计算方法。

优选地，在本实施例中，所述计算待拓展用户与每个种子用户群落的相似度包括：计算待拓展用户与每个种子用户群落中心点的相似度（计算方法与用户之间相似度计算方法一致），作为待拓展用户与每个种子用户群落的相似度。其中，种子用户群落是多个具有相似特征的用户集合，每个集合向一个中心点聚合，该中心点即种子用户群落的中心点。由于无需与种子用户群落中的每个用户计算相似度，从而可以大幅降低计算复杂度。

举例而言，采用夹角余弦方法计算待拓展用户与某个种子用户群落中心点的相似度可以采用如下公式1所示。

$$\cos(\vec{a},\vec{b})=\frac{\vec{a}\cdot\vec{b}}{|\vec{a}|\cdot|\vec{b}|} \qquad \text{公式1}$$

其中，$\cos(\vec{a},\vec{b})$代表待拓展用户a和某个种子用户群落中心点b的余弦相似度，$\vec{a}$代表待拓展用户a的评分向量，$\vec{b}$代表某个种子用户群落中心点b的评分向量。

步骤S33，若待拓展用户与特定种子用户群落的相似度大于或等于第一预设阈值（如80%），则将该待拓展用户划分入该特定种子用户群落。

举例而言，假设聚类分析后得到的种子用户群落包括三个：B1、B2、B3，待拓展用户A与种子用户群落B1的相似度S1为60%、与种子用户群落B2的相似度S2为85%（大于第一预设阈值）、与种子用户群落B3的相似度S3为90%（大于第一预设阈值），则将待拓展用户A划分入特定种子用户群落B2和B3。

步骤S34，统计每个待拓展用户划分入特定种子用户群落的个数并进行排序，并根据排序结果确定将待拓展用户拓展为种子用户的拓展规则。其中，排序越高代表相似度越高。

优选地，在本实施例中，所述拓展规则设置为：根据个数从高到低的顺序，选取指定数量（如前2位）的待拓展用户，将选取的待拓展用户拓展为种子用户。例如，假设待拓展用户A同时划分入5个种子用户群落，待拓展用户B同时划分入3个种子用户群落，待拓展用户C同时划分入2个种子用户群落，则将待拓展用户A和B拓展为种子用户。

优选地，在其它实施例中，所述拓展规则还可以设置为：若待拓展用户划分入特定种子用户群落的个数大于或等于第二预设阈值，则将该待拓展用户拓展为种子用户，其中，所述第二预设阈值可以设置为所有种子用户群落（即具有特定特征的种子用户群落）总个数的预定比例（如50%）。例如，假设所有种子用户群落总个数为4，预定比例为50%，则所述第二预设阈值为2。

通过上述步骤S31-S34，本申请所提出的种子用户拓展方法，通过无监督学习聚类方法与距离相似度算法相结合，减少了种子用户拓展的计算复杂度，并提高了拓展的准确度。

进一步地，为实现上述目的，本申请还提供一种计算机可读存储介质（如

ROM/RAM、磁碟、光盘），所述计算机可读存储介质存储有种子用户拓展系统20，所述种子用户拓展系统20可被至少一个处理器22执行，以使所述至少一个处理器22执行如下所述的种子用户拓展方法的步骤。

（1）通过预设的聚类方法（如无监督学习K-means聚类方法）针对预定数量（可以是大数据量级）的种子用户进行聚类分析，将所述种子用户分为若干具有特定特征（或显著特征）的种子用户群落。其中，所述特定特征包括，但不限于，用户的地理位置、是否为注册用户、是否购买过特定产品（如产险）等。

举例而言，假设聚类个数为k（即种子用户群落的个数），所有种子用户的预定数量为N，则K-means聚类方法包括如下步骤：

（A1）首先从N个数据对象（即N个种子用户）中任意选择 k 个数据对象作为聚类中心（即初始聚类中心）；

（A2）针对剩余的数据对象，计算每个剩余的数据对象到每个聚类中心的相似度（如欧式距离，距离越小代表相似度越高），并根据相似度的大小分别将每个剩余的数据对象分配给与其最相似的聚类中心所代表的聚类；

（A3）重新计算已经得到的每个聚类的聚类中心（即每个聚类中所有数据对象的均值）；

（A4）迭代A2至A3步骤，直至预设的标准测度函数开始收敛为止。在本实施例中，可以采用均方差作为预设的标准测度函数。

（2）通过预设的相似度计算方法，计算待拓展用户（一个或多个）与每个种子用户群落（即具有特定特征的种子用户群落）的相似度。其中，所述预设的相似度计算方法可以是：欧氏距离、夹角余弦、汉明距离等相似度计算方法。

优选地，在本实施例中，所述计算待拓展用户与每个种子用户群落的相似度包括：计算待拓展用户与每个种子用户群落中心点的相似度（计算方法

与用户之间相似度计算方法一致），作为待拓展用户与每个种子用户群落的相似度。其中，种子用户群落是多个具有相似特征的用户集合，每个集合向一个中心点聚合，该中心点即种子用户群落的中心点。由于无需与种子用户群落中的每个用户计算相似度，从而可以大幅降低计算复杂度。

举例而言，采用夹角余弦方法计算待拓展用户与某个种子用户群落中心点的相似度可以采用如下公式1所示。

$$\cos(\vec{a},\vec{b})=\frac{\vec{a}\cdot\vec{b}}{|\vec{a}|\cdot|\vec{b}|} \qquad \text{公式1}$$

其中，$\cos(\vec{a},\vec{b})$代表待拓展用户a和某个种子用户群落中心点b的余弦相似度，$\vec{a}$代表待拓展用户a的评分向量，$\vec{b}$代表某个种子用户群落中心点b的评分向量。

（3）若待拓展用户与特定种子用户群落的相似度大于或等于第一预设阈值（如80%），则将该待拓展用户划分入该特定种子用户群落。

举例而言，假设聚类分析后得到的种子用户群落包括三个：B1、B2、B3，待拓展用户A与种子用户群落B1的相似度S1为60%、与种子用户群落B2的相似度S2为85%（大于第一预设阈值）、与种子用户群落B3的相似度S3为90%（大于第一预设阈值），则将待拓展用户A划分入特定种子用户群落B2和B3。

（4）统计每个待拓展用户划分入特定种子用户群落的个数并进行排序，并根据排序结果确定将待拓展用户拓展为种子用户的拓展规则。其中，排序越高代表相似度越高。

优选地，在本实施例中，所述拓展规则设置为：根据个数从高到低的顺序，选取指定数量（如前2位）的待拓展用户，将选取的待拓展用户拓展为种子用户。例如，假设待拓展用户A同时划分入5个种子用户群落，待拓展用户B同时划分入3个种子用户群落，待拓展用户C同时划分入2个种子用户群落，则

将待拓展用户A和B拓展为种子用户。

优选地，在其它实施例中，所述拓展规则还可以设置为：若待拓展用户划分入特定种子用户群落的个数大于或等于第二预设阈值，则将该待拓展用户拓展为种子用户，其中，所述第二预设阈值可以设置为所有种子用户群落（即具有特定特征的种子用户群落）总个数的预定比例（如50%）。例如，假设所有种子用户群落总个数为4，预定比例为50%，则所述第二预设阈值为2。

通过上述步骤（1）-（4），本申请所提出的计算机可读存储介质，通过无监督学习聚类方法与距离相似度算法相结合，减少了种子用户拓展的计算复杂度，并提高了拓展的准确度。

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来实现，但很多情况下前者是更佳的实施方式。基于这样的理解，本申请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）执行本申请各个实施例所述的方法。

以上参照附图说明了本申请的优选实施例，并非因此局限本申请的权利范围。上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。另外，虽然在流程图中示出了逻辑顺序，但是在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤。

本领域技术人员不脱离本申请的范围和实质，可以有多种变型方案实现本申请，比如作为一个实施例的特征可用于另一实施例而得到又一实施例。凡是利用本申请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，均同理包括在本申请的专利保护范围内。

权 利 要 求 书

1. 一种电子设备，其特征在于，所述电子设备包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的种子用户拓展系统，所述种子用户拓展系统被所述处理器执行时实现如下步骤：

通过预设的聚类方法针对预定数量的种子用户进行聚类分析，将所述种子用户分为若干具有特定特征的种子用户群落；

通过预设的相似度计算方法，计算待拓展用户与每个种子用户群落的相似度；

若待拓展用户与特定种子用户群落的相似度大于或等于第一预设阈值，则将该待拓展用户划分入该特定种子用户群落；及

统计每个待拓展用户划分入特定种子用户群落的个数并进行排序，并根据排序结果确定将待拓展用户拓展为种子用户的拓展规则。

2. 如权利要求 1 所述的电子设备，其特征在于，所述特定特征包括用户的地理位置、是否为注册用户、是否购买过特定产品。

3. 如权利要求 1 所述的电子设备，其特征在于，所述计算待拓展用户与每个种子用户群落的相似度包括：计算待拓展用户与每个种子用户群落中心点的相似度，作为待拓展用户与每个种子用户群落的相似度。

4. 如权利要求 2 所述的电子设备，其特征在于，所述拓展规则设置为：根据个数从高到低的顺序，选取指定数量的待拓展用户，将选取的待拓展用户拓展为种子用户。

5. 如权利要求 3 所述的电子设备，其特征在于，所述拓展规则设置为：

根据个数从高到低的顺序，选取指定数量的待拓展用户，将选取的待拓展用户拓展为种子用户。

6. 如权利要求 2 所述的电子设备，其特征在于，所述拓展规则设置为：若待拓展用户划分入特定种子用户群落的个数大于或等于第二预设阈值，则将该待拓展用户拓展为种子用户，其中，所述第二预设阈值设置为所有种子用户群落总个数的预定比例。

7. 如权利要求 3 所述的电子设备，其特征在于，所述拓展规则设置为：若待拓展用户划分入特定种子用户群落的个数大于或等于第二预设阈值，则将该待拓展用户拓展为种子用户，其中，所述第二预设阈值设置为所有种子用户群落总个数的预定比例。

8. 一种种子用户拓展方法，应用于电子设备，其特征在于，所述方法包括：

通过预设的聚类方法针对预定数量的种子用户进行聚类分析，将所述种子用户分为若干具有特定特征的种子用户群落；

通过预设的相似度计算方法，计算待拓展用户与每个种子用户群落的相似度；

若待拓展用户与特定种子用户群落的相似度大于或等于第一预设阈值，则将该待拓展用户划分入该特定种子用户群落；及

统计每个待拓展用户划分入特定种子用户群落的个数并进行排序，并根据排序结果确定将待拓展用户拓展为种子用户的拓展规则。

9. 如权利要求 8 所述的种子用户拓展方法，其特征在于，所述特定特征包括用户的地理位置、是否为注册用户、是否购买过特定产品。

10. 如权利要求 8 所述的种子用户拓展方法，其特征在于，所述计算待拓展用户与每个种子用户群落的相似度包括：

计算待拓展用户与每个种子用户群落中心点的相似度，作为待拓展用户与每个种子用户群落的相似度。

11. 如权利要求 9 所述的种子用户拓展方法，其特征在于，所述拓展规则设置为：根据个数从高到低的顺序，选取指定数量的待拓展用户，将选取的待拓展用户拓展为种子用户。

12. 如权利要求 10 所述的种子用户拓展方法，其特征在于，所述拓展规则设置为：根据个数从高到低的顺序，选取指定数量的待拓展用户，将选取的待拓展用户拓展为种子用户。

13. 如权利要求 9 所述的种子用户拓展方法，其特征在于，所述拓展规则设置为：若待拓展用户划分入特定种子用户群落的个数大于或等于第二预设阈值，则将该待拓展用户拓展为种子用户，其中，所述第二预设阈值设置为所有种子用户群落总个数的预定比例。

14. 如权利要求 10 所述的种子用户拓展方法，其特征在于，所述拓展规则设置为：若待拓展用户划分入特定种子用户群落的个数大于或等于第二预设阈值，则将该待拓展用户拓展为种子用户，其中，所述第二预设阈值设置为所有种子用户群落总个数的预定比例。

15. 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种子用户拓展系统，所述种子用户拓展系统可被至少一个处理器执行，所述种子用户拓展系统被所述处理器执行时实现如下步骤：

通过预设的聚类方法针对预定数量的种子用户进行聚类分析，将所述种子用户分为若干具有特定特征的种子用户群落；

通过预设的相似度计算方法，计算待拓展用户与每个种子用户群落的相似度；

若待拓展用户与特定种子用户群落的相似度大于或等于第一预设阈值，则将该待拓展用户划分入该特定种子用户群落；及

统计每个待拓展用户划分入特定种子用户群落的个数并进行排序，并根据排序结果确定将待拓展用户拓展为种子用户的拓展规则。

16. 如权利要求 15 所述的计算机可读存储介质，其特征在于，所述特定特征包括用户的地理位置、是否为注册用户、是否购买过特定产品。

17. 如权利要求 15 所述的计算机可读存储介质，其特征在于，所述计算待拓展用户与每个种子用户群落的相似度包括：计算待拓展用户与每个种子用户群落中心点的相似度，作为待拓展用户与每个种子用户群落的相似度。

18. 如权利要求 16 所述的计算机可读存储介质，其特征在于，所述拓展规则设置为：根据个数从高到低的顺序，选取指定数量的待拓展用户，将选取的待拓展用户拓展为种子用户。

19. 如权利要求 17 所述的计算机可读存储介质，其特征在于，所述拓展规则设置为：根据个数从高到低的顺序，选取指定数量的待拓展用户，将选取的待拓展用户拓展为种子用户。

20. 如权利要求 16 或 17 所述的计算机可读存储介质，其特征在于，所述拓展规则设置为：若待拓展用户划分入特定种子用户群落的个数大于或等于

第二预设阈值，则将该待拓展用户拓展为种子用户，其中，所述第二预设阈值设置为所有种子用户群落总个数的预定比例。

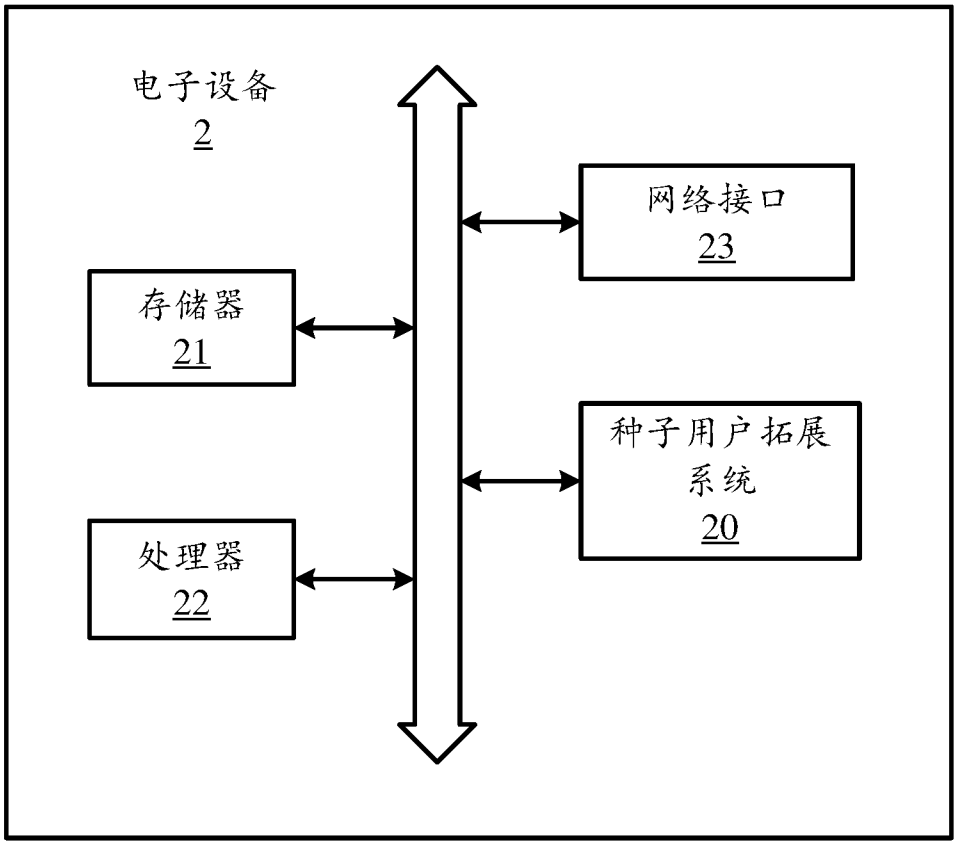


图 1

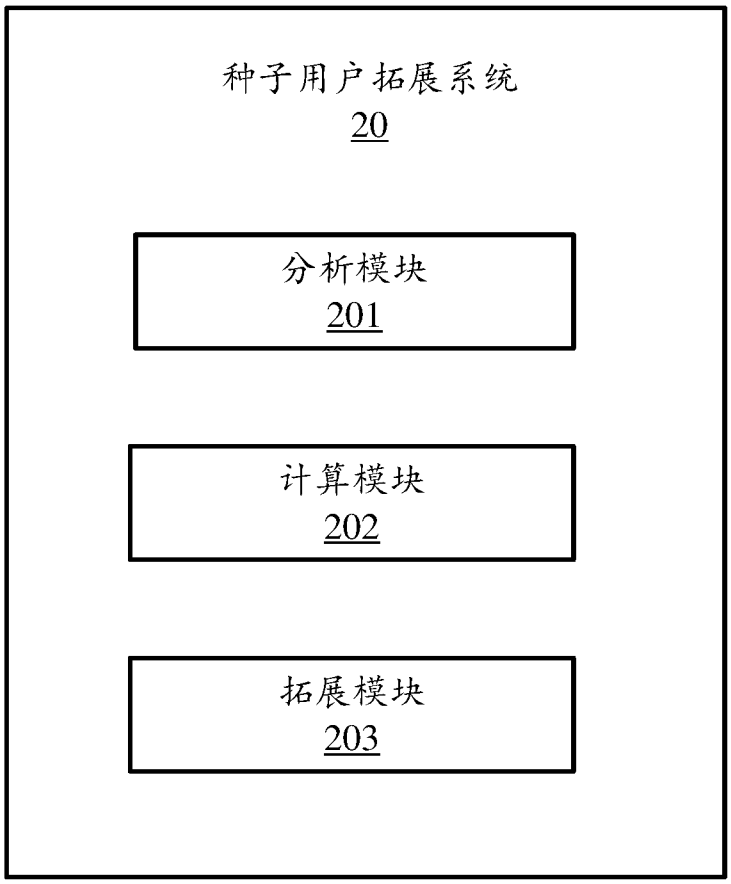


图 2

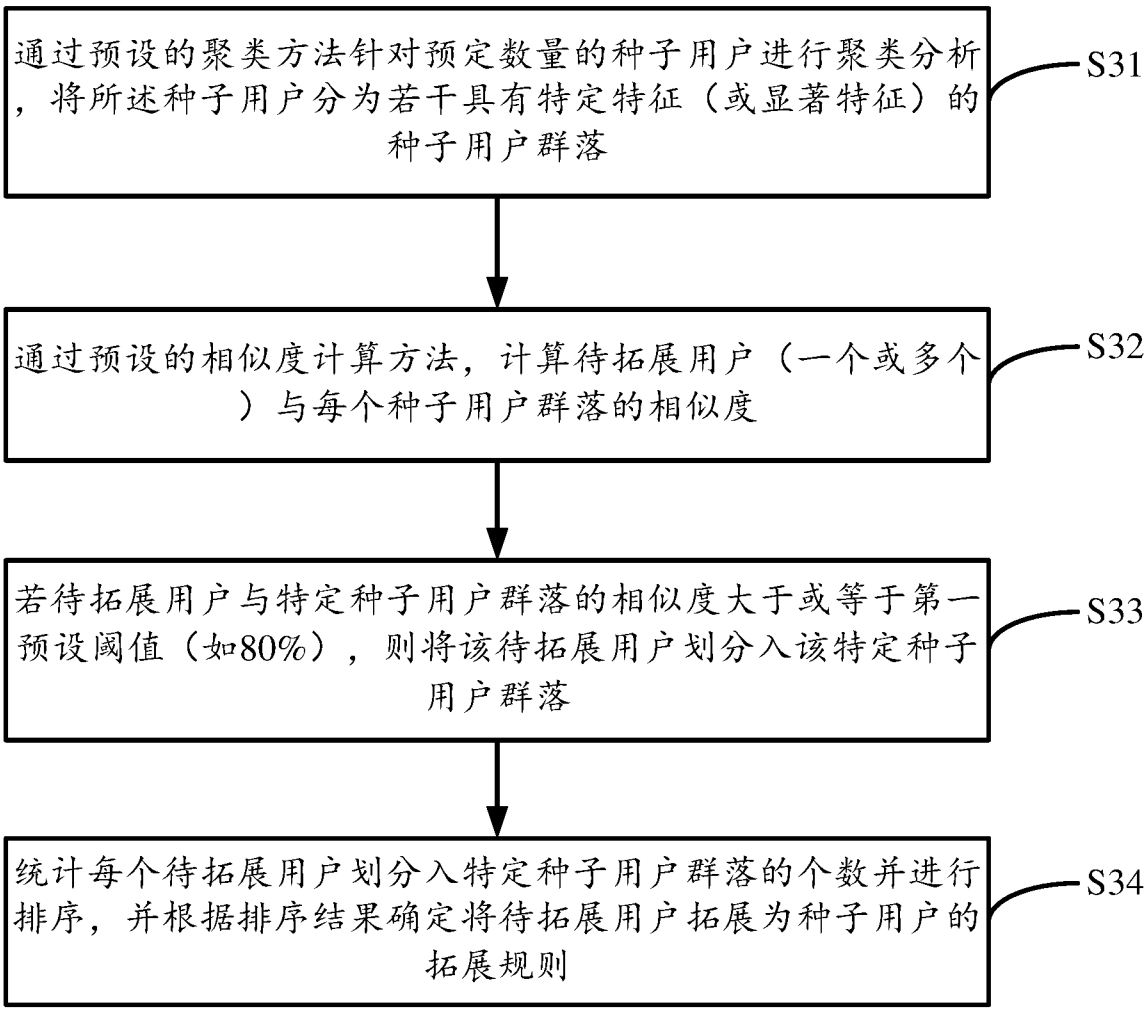

通过预设的聚类方法针对预定数量的种子用户进行聚类分析，将所述种子用户分为若干具有特定特征（或显著特征）的种子用户群落
S31
通过预设的相似度计算方法，计算待拓展用户（一个或多个）与每个种子用户群落的相似度
S32
若待拓展用户与特定种子用户群落的相似度大于或等于第一预设阈值（如80%），则将该待拓展用户划分入该特定种子用户群落
S33
统计每个待拓展用户划分入特定种子用户群落的个数并进行排序，并根据排序结果确定将待拓展用户拓展为种子用户的拓展规则
S34

图 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/076181

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 30/02(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

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)

CNTXT; CNABS; CNKI; TWTXT; TWABS; USTXT; VEN; EPTXT; WOTXT: 种子, 用户, 聚类, 相似度, 排序, 中心点, seeds, user, clustering, similarity, sort, central point

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 106022800 A (BEIJING BAIFENDIAN INFORMATION TECHNOLOGY, INC.) 12 October 2016 (2016-10-12) description, paragraphs 29-32, 156-163, and 222-233, and figure 1D	1-20
A	CN 102737327 A (INTERNATIONAL BUSINESS MACHINES CORPORATION) 17 October 2012 (2012-10-17) entire document	1-20
A	CN 107067045 A (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD. ET AL.) 18 August 2017 (2017-08-18) entire document	1-20
A	US 2015302436 A1 (REYNOLDS, T. J.) 22 October 2015 (2015-10-22) 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

"E" earlier application or patent but published on or after the international filing date

"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)

"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

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 August 2018	**25 September 2018**

Name and mailing address of the ISA/CN	Authorized officer
State Intellectual Property Office of the P. R. China (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/076181

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 106022800 A	12 October 2016	None	
CN 102737327 A	17 October 2012	US 2012254179 A1	04 October 2012
		CN 102737327 B	10 February 2016
		US 8914372 B2	16 December 2014
		US 8918397 B2	23 December 2014
		US 2012290580 A1	15 November 2012
CN 107067045 A	18 August 2017	None	
US 2015302436 A1	22 October 2015	None	

国际检索报告

国际申请号
PCT/CN2018/076181

A. 主题的分类

G06Q 30/02(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

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））

CNTXT;CNABS;CNKI;TWTXT;TWABS;USTXT;VEN;EPTXT;WOTXT:种子，用户，聚类，相似度，排序，中心点，seeds, user, clustering, similarity, sort, central point

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
X	CN 106022800 A (北京百分点信息科技有限公司) 2016年 10月 12日 (2016-10-12) 说明书第29-32、156-163、222-233段，图1D	1-20
A	CN 102737327 A (国际商业机器公司) 2012年 10月 17日 (2012-10-17) 全文	1-20
A	CN 107067045 A (北京京东尚科信息技术有限公司 等) 2017年 8月 18日 (2017-08-18) 全文	1-20
A	US 2015302436 A1 (REYNOLDS, THOMAS J.) 2015年 10月 22日 (2015-10-22) 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	国际检索报告邮寄日期
2018年 8月 30日	2018年 9月 25日

ISA/CN的名称和邮寄地址	受权官员
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	杨牛
传真号 (86-10)62019451	电话号码 86-(20)-28950388

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/076181

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	106022800	A	2016年 10月 12日	无			
CN	102737327	A	2012年 10月 17日	US	2012254179	A1	2012年 10月 4日
				CN	102737327	B	2016年 2月 10日
				US	8914372	B2	2014年 12月 16日
				US	8918397	B2	2014年 12月 23日
				US	2012290580	A1	2012年 11月 15日
CN	107067045	A	2017年 8月 18日	无			
US	2015302436	A1	2015年 10月 22日	无			