

(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
2017 年 5 月 18 日 (18.05.2017)



(10) 国际公布号
WO 2017/080170 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2016/083164
- (22) 国际申请日: 2016 年 5 月 24 日 (24.05.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510772095.3 2015 年 11 月 12 日 (12.11.2015) CN
- (71) 申请人: 乐视控股(北京)有限公司 (LE HOLDINGS (BEIJING) CO., LTD.) [CN/CN]; 中国北京市朝阳区姚家园路 105 号 3 号楼 10 层 1102, Beijing 100025 (CN)。 乐视云计算有限公司 (LECLOUD COMPUTING CO., LTD.) [CN/CN]; 中国北京市海淀区中关村南大街 5 号 1 区 689 号楼 1153, Beijing 100081 (CN)。
- (72) 发明人: 张幼明 (ZHANG, Youming); 中国北京市海淀区中关村南大街 5 号 1 区 689 号楼 1153, Beijing 100081 (CN)。
- (74) 代理人: 北京商专永信知识产权代理事务所(普通合伙) (SBZL IP LAW OFFICE); 中国北京市海淀区知春路 1 号 9 层 901 室, Beijing 100083 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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)。

[见续页]

(54) Title: GROUP USER PROFILING METHOD AND SYSTEM

(54) 发明名称: 群体用户画像方法及系统

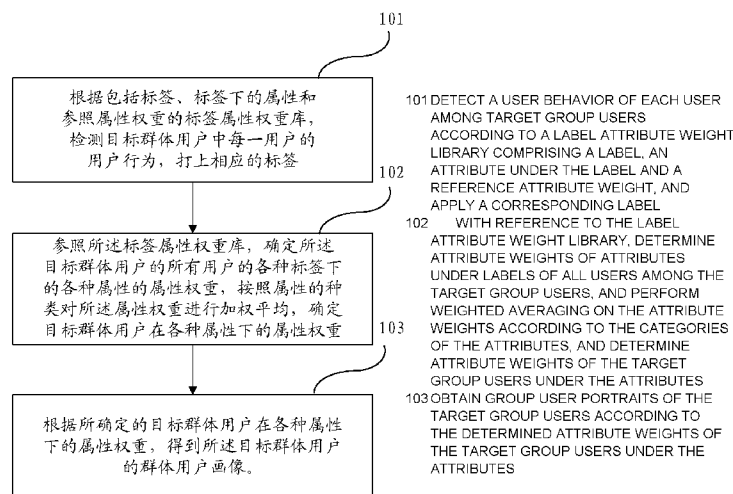


图 1

(57) Abstract: A group user profiling method and system. The method comprises: detecting a user behavior of each user among target group users according to a label attribute weight library comprising a label, an attribute under the label and a reference attribute weight, and applying a corresponding label (101); with reference to the label attribute weight library, determining attribute weights of attributes under labels of all users among the target group users, and performing weighted averaging on the attribute weights according to the categories of the attributes, and determining attribute weights of the target group users under the attributes (102); and obtaining group user portraits of the target group users according to the determined attribute weights of the target group users under the attributes (103). The method and the system can effectively improve the precision of group user portraits.

(57) 摘要:

[见续页]

WO 2017/080170 A1

根据细则 4.17 的声明:

— 关于申请人有权要求在先申请的优先权(细则 4.17(iii))

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种群体用户画像方法和系统，所述方法包括：根据包括标签、标签下的属性和参照属性权重的标签属性权重库，检测目标群体用户中每一用户的用户行为，打上相应的标签（101）；参照所述标签属性权重库，确定所述目标群体用户的所有用户的各种标签下的各种属性的属性权重，按照属性的种类对所述属性权重进行加权平均，确定目标群体用户在各种属性下的属性权重（102）；根据所确定的目标群体用户在各种属性下的属性权重，得到所述目标群体用户的群体用户画像（103）。所述方法和系统能有效提高群体用户画像的精准度。

群体用户画像方法及系统

技术领域

5 本发明实施例涉及用户画像领域，尤其涉及一种群体用户画像方法及系统。

背景技术

在互联网逐渐步入大数据时代后，消费者的一切行为在企业面前似乎都
10 将是“可视化”的。企业的专注点也开始益聚焦于怎样利用大数据来为精准营销服务，进而深入挖掘潜在的商业价值。于是，“用户画像”的概念也就应运而生。

大数据使得企业能够通过互联网便利地获取用户更为广泛的反馈信息，为进一步精准、快速地分析用户行为习惯、消费习惯等重要商业信息，提供
15 了足够的数据基础。伴随着对人的了解逐步深入，一个概念悄然而生：用户画像（UserProfile），完美地抽象出一个用户的信息全貌，可以看作企业应用大数据的根基。

用户画像，即用户信息标签化，就是企业通过收集与分析消费者社会属性、生活习惯、消费行为等主要信息的数据之后，完美地抽象出一个用户的
20 商业全貌，可以看作是企业应用大数据技术的基本方式。

用户画像分为个体用户画像和群体用户画像。前者主要用于个性化定制，而后者用于对群体用户的定位。目前所有的群体用户画像都是基于个人的用

户画像，即先画个人用户的画像（确定每个人的各属性的属性值），而后再画成群体的画像（统计各属性值在各属性中所占的比例）。

用户画像领域中常见的几个专业术语有属性、属性值以及在群体用户画像中常用的属性权重，其中属性是指：进行用户画像需要统计的维度，如性别下的男和女，年龄下的少年、青年、中年、老年，收入等级下的贫困，中低，中等，富裕等；属性权重在个体用户画像中解释为可能性，例如男女权重 0.8:0.2，解释为该用户 80%可能为男性 20%可能为女性，而在群体用户画像中则解释为在所统计的群体中存在 80%的比例的用户为男性，存在 20%比例的用户为女性。

现有技术中个体用户画像主要有两种实现方案：第一种实现方案是：通过用户的注册信息直接画出用户画像的方法和对用户的行为进行监测，而后为用户打上各种标签；第二种实现方案是：后台工作人员利用个人经验对所有标签进行分析推导得出用户画像的方法。

关于上述个体用户画像的第一种实现方案，存在以下缺点：目前很多的网站/媒体的访问并不需要提前注册，故这些网站/媒体也并不清楚自己的用户的属性；另外，有些用户也不愿意注册用户信息，即使用户注册了信息，也很难保证注册信息的准确性（比如，涉及用户的个人隐私、时间因素等），这样就很难得到准确的用户画像。

关于上述个体用户画像的第二种实现方案，存在以下缺点：过于依赖后台工作人员个人因素会导致得到的用户画像结果的差异性很大，而且也没有考虑到标签的时效性，会导致最终得到的用户画像不够精确。

现有技术中的群体用户画像方法是基于现有技术中所得的个体用户画像构建而成的群体用户画像，由此使得个体用户画像的误差在被统计成群体用

户画像时被叠加放大，导致群体用户画像结果的精确度不高。

发明内容

本发明实施例提供一种群体用户画像方法，解决现有技术中因个人差异
5 为用户画像而造成的差异而使得群体用户画像不够精确的技术问题。

本发明一实施例提供一种群体用户画像方法，包括：

根据包括标签、标签下的属性和参照属性权重的标签属性权重库，检测
目标群体用户中每一用户的用户行为，打上相应的标签；

参照所述标签属性权重库，确定所述目标群体用户的所有用户的各种标
10 签下的各种属性的属性权重，按照属性的种类对所述属性权重进行加权平均，
确定目标群体用户在各种属性下的属性权重；

根据所确定的目标群体用户在各种属性下的属性权重，得到所述目标群
体用户的群体用户画像。

本发明一实施例提供一种群体用户画像系统，包括：

15 行为检测单元，配置以：根据包括标签、标签下的属性和参照属性权重
的标签属性权重库，检测目标群体用户中每一用户的用户行为，打上相应的
标签；

属性权重确定单元，配置以：参照所述标签属性权重库，确定所述目标
群体用户的所有用户的各种标签下的各种属性的属性权重，按照属性的种类
20 对所述属性权重进行加权平均，确定目标群体用户在各种属性下的属性权
重；

用户画像生成单元，配置以：根据所确定的目标群体用户在各种属性下
的属性权重，得到所述目标群体用户的群体用户画像。

在本发明实施例所提供的群体用户画像方法及系统中，提出了一套通过建立标签属性权重库，查询匹配标签属性权重库，确定目标群体用户的群体属性权重，最终得到群体用户画像；其中标签属性权重库中的参照属性权重指代的是大众平均化的标签属性权重，将此参照属性权重应用于群体用户画像中，降低了个体用户画像的误差对群体用户画像的干扰，提高了群体用户画像的精确性。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图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的前提下，还可以根据这些附图获得其他的附图。

图 1 示出了本发明一实施例的群体用户画像方法的流程图；

图 2 示出了图 1 中方法中的步骤 S102 的一种具体实施方式的详细执行图；

图 3 示出了本发明一实施例的群体用户画像系统的示意图；

图 4 为实施本发明实施例的群体用户画像方法及系统的架构图；

图 5 为可以应用于实现本发明实施例的终端设备或服务器的结构示意图。

具体实施方式

为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作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。

参见图 1 示出的是本发明一具体实施例的群体用户画像方法的流程图，所述方法包括：

S101：根据包括标签、标签下的属性和参照属性权重的标签属性权重库，用户画像服务器检测目标群体用户中每一用户的用户行为，打上相应的标签；

S102：参照所述标签属性权重库，用户画像服务器确定所述目标群体用户的所有用户的各种标签下的各种属性的属性权重，按照属性的种类对所述属性权重进行加权平均，确定目标群体用户在各种属性下的属性权重；

S103：根据所确定的目标群体用户在各种属性下的属性权重，用户画像服务器得到所述目标群体用户的群体用户画像。

在本发明实施例所提供的群体用户画像方法中，提出了一套通过建立标签属性权重库，查询匹配标签属性权重库，确定目标群体用户的群体属性权重，最终得到群体用户画像；其中标签属性权重库中的参照属性权重指代的是大众平均化的标签属性权重，将此参照属性权重应用于群体用户画像中，降低了个体用户画像的误差对群体用户画像的干扰，提高了群体用户画像的精确性。

在本实施例方法的一种优选实施例中，在所述根据包括标签、标签下的属性和参照属性权重的标签属性权重库，用户画像服务器检测目标群体用户

中每一用户的用户行为，打上相应的标签之前，还包括在用户画像服务器中建立标签属性权重库；关于建立标签属性权重库的执行，可以包括如下子步骤：

选择与所述目标群体用户相关的参照群体用户；

- 5 用户画像服务器根据参照群体用户中每一用户的用户行为，打上相应的标签；

用户画像服务器根据所述参照群体用户中每一用户的标签，确定参照群体用户中每一用户在各种属性下的属性权重；

- 10 用户画像服务器将每一用户在各种属性下的属性权重赋予到与属性相对应的标签中；

用户画像服务器按照标签的种类对参照群体用户中所有用户的属性权重加权平均，确定参照群体用户在每种标签下各种属性的参照属性权重；

用户画像服务器根据标签、标签下的属性和参照属性权重建立标签属性权重库。

- 15 由此所得的标签属性权重库中的参照属性权重指代的是大众平均化的标签属性权重，将此参照属性权重应用于目标群体用户画像中，降低了个体用户画像的误差对群体用户画像的干扰，提高了群体用户画像的精确性。

在本发明方法的一种实施方式中，关于建立标签属性权重库的具体执行过程还可以包括：

- 20 用户画像服务器根据参照群体用户中每一用户的历史表现，基于用户属性挖掘模型，确定每一用户在各种属性下的属性权重；常见的用户属性挖掘模型包括 svm、贝叶斯、聚类、加权平均等各种算法模型。

在现有技术中，对群体用户的各个个体用户的属性进行推导，再将个体用户已确定的属性应用在群体用户中，得到群体用户画像；在本发明实施例方法中，用户画像服务器基于参照群体用户中每一用户的历史表现对每一用户的属性权重进行推导，是对用户属性的一种模糊化处理，并将所得的属性权重赋予到相应的标签中，使得标签相对于属性具有更高的参考价值，并能保障基于标签所获得的群体用户画像更加精确。

作为示例，用户画像服务器将标签释义分类，获取对应于各种属性的各个匹配标签，可以通过关键词分类工具来实现，所述匹配标签与属性之间具有一定的逻辑关系即可，不一定要要求所述匹配标签一定能够推导得出属性。

例如，当匹配标签为“化妆品”时，即可推导得出相对应的属性为用户性别属性，可以理解的是相对应于属性的属性权重即为男女比例。

在本发明方法的一些可选的实施方式中，标签属性权重库是由多个子标签属性权重库构成，不同的子标签属性权重库与不同维度的属性相对应，例如：年龄标签属性权重库与用户年龄维度属性相对应、收入等级标签属性权重库与用户收入等级维度属性相对应、消费等级标签属性权重库与用户消费等级维度属性相对应、消费喜好标签属性权重库与用户消费喜好维度属性相对应等，由用户各个不同维度的属性共同构成了用户画像。

可以理解的是，标签的一种表现形式为匹配关键词，所述匹配关键词对应着用户行为；由于用户在各个数据源对应的网页上执行诸如浏览产品操作、购买产品操作、关注产品操作或收藏产品操作时，均可触发日志信息的生成，相应地日志信息的生成时间用于说明用户执行上述诸如浏览产品操作、购买产品操作、关注产品操作或收藏产品操作所对应的时间；针对上述

用户行为的情形，可以选择将产品信息或产品的分类信息作为用户行为与用户标签相匹配的匹配关键词。

本发明主要提供一种群体用户画像的技术方案，主要包括：根据用户行为，对所有用户（包括参照群体用户和目标群体用户）打标签。根据经验分析，

5 对参照群体用户中的一个参照用户进行模糊化推理，推出该参照用户每个属性的属性权重；将每个属性的属性权重赋予到该用户的每个标签中；依次类推，直至参照群体用户中所有的参照用户均完成相同的步骤；对每个标签的所有属性权重加权平均，得到最终的属性权重，由此建立标签的属性权重库。将一个测试用户的所有标签放入库中，得到它们的各属性的权重值，然后后

10 加权平均，得到该测试用户最终的各属性的权重值，依次类推，直到目标群体用户中所有的测试用户均完成相同的步骤；对所有测试用户的各属性权重值加权平均，得到群用户画像。

在本发明实施例方法的一种优选实施例中，在根据标签、标签下的属性和参照属性权重建立标签属性权重库后，在用户画像服务器中定期补充参照

15 群体用户中的用户数量，对参照属性权重进行修正更新。

当参照群体用户中参照用户的数量越多，参照属性权重就越精确，通过定期向用户画像服务器中补充参照群体用户的用户数量，并对参照属性权重进行修正更新，由此实现了参照属性权重的定期修正，更保障了群体用户画像的实时精准性。

20 作为图 1 所示实施例方法的进一步优化，查询标签属性权重库，若在标签属性权重库中没有找到与测试群体用户所生成的标签相匹配的标签时，则将此标签及此标签所对应的属性和属性权重添加至标签属性权重库；

由此实现了标签属性权重库的学习与扩充；在一种情况下，目标群体用户的所有属性均可参照参照群体用户的参照属性权重，更加提高了群体用户画像的精准度。

在本发明实施例方法的一种优选实施例中，在图 1 所示的实施例方法的

5 S103 之后还包括：

在完成群体用户画像后，用户画像服务器根据群体用户画像为群体用户进行个性化信息推送，继续检测群体用户在收到所述个性化信息推送后的行为以重新确定用户属性权重。

在本实施例中，基于群体用户画像为群体用户进行个性化信息推送，根据群体用户对个性化信息推送反馈的用户行为，重新确定群体用户的属性权重，实现了对群体用户属性权重和群体用户画像的校准，同时也避免了对群体用户信息推送的一成不变。

继续参照图 2，作为图 1 所示实施例方法的进一步优化，关于图 1 中的 S102 的执行，可以包括如下子步骤：

15 S1021：用户画像服务器将所述目标群体用户中的一个用户的所有标签放入所述标签属性权重库中，确定所述一个用户的所有标签的各种属性的属性权重；

S1022：用户画像服务器按照属性的种类对所述属性权重进行加权平均，确定所述一个用户在各种属性下的属性权重；

20 S1023：以此类推，用户画像服务器确定所述目标群体用户中的所有用户在各种属性下的属性权重；

S1024：用户画像服务器对所述目标群体用户中的所有用户在各种属性下的属性权重加权平均，确定目标群体用户在各种属性下的属性权重。

在本实施例中，一方面，通过目标群体用户的多种标签共同确定目标群体用户属性，由此提高了所获得的群体用户属性的精确度；另一方面，通过对目标群体用户中的所有用户在各种属性下的属性权重进行加权平均，由此避免了因个体用户画像差异对群体用户画像的干扰，提高了群体用户画像的精确度。

更具体地，关于 S102 步骤的更具体的执行过程可以是：

将所述目标群体用户中的一个用户的所有标签放入所述参照所述标签属性权重库中，用户画像服务器确定所述一个用户的所有标签的各种属性的属性权重；例如：将所述目标群体用户中的一个用户的所有标签放入所述参照所述标签属性权重库中，用户画像服务器将所述一个用户的所有标签以标签为 key 遍历所述标签属性权重库，确定所述一个用户的所有标签所对应的各种属性及各种属性的属性权重；

用户画像服务器按照属性的种类对所述属性权重进行加权平均，确定所述一个用户在各种属性下的属性权重；作为示例，用户画像服务器将所述一个用户对应于用户年龄属性的所有标签的属性权重进行加权平均，将得到的平均值作为所述一个用户在用户年龄属性下的属性权重，以此类推，可以得到所述一个用户在各种属性（例如，用户消费等级属性和用户性别属性）下的属性权重；

以此类推，用户画像服务器确定所述目标群体用户中的所有用户在各种属性下的属性权重；作为示例，用户画像服务器对目标群体用户中的其他用户重复上述步骤，由此得到目标群体用户相在各种属性下的属性权重，可以得到所述目标群体的所有用户在用户年龄属性、用户消费等级属性和用户性别属性等各种属性下的属性权重；

用户画像服务器对所述目标群体用户中的所有用户在各种属性下的属性权重加权平均，确定目标群体用户在各种属性下的属性权重；作为示例，用户画像服务器对所述目标群体用户中所有用户在用户年龄属性下的属性权重进行加权平均，将所得到的平均值作为所述群体用户在用户年龄属性下的属性权重；以此类推，用户画像服务器得到目标群体用户在用户消费等级属性和用户性别属性各种属性下的属性权重。

关于图 1 中的 S103 步骤的执行，用户画像服务器根据所确定的目标群体用户在各种属性下的属性权重，得到所述目标群体用户的群体用户画像。作为示例，用户画像服务器将所确定的目标群体用户在各种属性下的属性权重应用于目标群体用户的属性中，例如：当确定目标群体用户在性别维度下的属性权重为 0.7:0.3（男：女），相应地可确定目标群体用户中 70% 的用户为男性和 30% 的用户为女性。

可以理解的是，属性权重是对用户属性的一种模糊化的处理的体现形式，在用户画像服务器确定群体用户的属性权重后，将模糊的属性权重转化为清晰的关于群体用户属性的人群比例，由此确定群体用户画像；本发明实施例所述的群体用户画像方法操作方便，并且通过本实施例方法所得的群体用户画像具有较高的精确度。

本发明提供的一种群体用户画像的方案，主要在于：用户画像服务器不需要获取准确的用户注册信息，对群体用户中的个体用户属性的确定不需要十分的准确与具体，只要能推导得到群体用户中所有个体用户的属性权重即可，例如：在用户画像服务器对群体用户的性别属性进行推导的时候，只需要对群体用户中每个个体用户的性别进行模糊推理，不需要完全确定个体用户的性别是男或女，只需要得到个体用户在性别属性下的属性权重，所述属

性权重的大小与此个体用户性别为男或女可能性的强弱相对应；同时在用户画像服务器中的标签属性权重数据库建立之时，选定一个与目标群体用户相关的参照群体用户，在一种情况下，标签的参照属性权重会随着参照群体用户中参照用户的数量的增多而修正并使得属性权重趋于大众平均化水平，故

5 参照属性权重代表的是群体用户属性在大众平均化下的属性权重，将此参照属性权重应用于群体用户画像中，修正了因个体误差对群体用户画像造成的干扰。

其中上述图 1 的方法，可以根据本发明的下述系统（参照图 3）予以操作来实现。

10 参见图 3 示出的是本发明一实施例的群体用户画像系统的结构示意图，具体包括：

行为检测单元，配置以：根据包括标签、标签下的属性和参照属性权重的标签属性权重库，检测目标群体用户中每一用户的用户行为，打上相应的标签；

15 属性权重确定单元，配置以：参照所述标签属性权重库，确定所述目标群体用户的所有用户在被行为检测单元所打上的各种标签下的各种属性的属性权重，按照属性的种类对所述属性权重进行加权平均，确定目标群体用户在各种属性下的属性权重；

用户画像生成单元，配置以：根据属性权重确定单元所确定的目标群体

20 用户在各种属性下的属性权重，得到所述目标群体用户的群体用户画像。

在本发明实施例所提供的群体用户画像系统中，提出了一套通过建立标签属性权重库，查询匹配标签属性权重库，确定目标群体用户的群体属性权重，最终得到群体用户画像；其中标签属性权重库中的参照属性权重指代的

是大众平均化的标签属性权重，将此参照属性权重应用于群体用户画像中，降低了个体用户画像的误差对群体用户画像的干扰，提高了群体用户画像的精确性。

本实施例中的群体用户画像系统为一个服务器或者服务器集群，其中每个单元可以是单独的服务器或者服务器集群，此时，上述单元之间的交互表现为各单元所对应的服务器或者服务器集群之间的交互，所述多个服务器或服务器集群共同构成本发明的群体用户画像系统。

具体地，上述多个服务器或服务器集群共同构成本发明的群体用户画像系统包括：

行为检测服务器或服务器集群，配置以：根据包括标签、标签下的属性和参照属性权重的标签属性权重库，检测目标群体用户中每一用户的用户行为，打上相应的标签；

属性权重确定服务器或服务器集群，配置以：参照所述标签属性权重库，确定所述目标群体用户的所有用户在被行为检测单元所打上的各种标签下的各种属性的属性权重，按照属性的种类对所述属性权重进行加权平均，确定目标群体用户在各种属性下的属性权重；

用户画像生成服务器或服务器集群，配置以：根据属性权重确定单元所确定的目标群体用户在各种属性下的属性权重，得到所述目标群体用户的群体用户画像。

在一种替代实施例中，可以是上述多个单元中的几个单元共同组成一个服务器或者服务器集群。例如：行为检测单元和属性权重确定单元共同组成

第一服务器或者第一服务器集群，用户画像生成单元构成第二服务器或者第二服务器集群。

此时，上述单元之间的交互表现为第一服务器和第二服务器之间的交互或者第一服务器集群和第二服务器集群之间的交互，所述第一服务器和第二服务器或第一服务器集群至第二服务器集群共同构成本发明的群体用户画像系统。

作为图 3 所述实施例系统的进一步优化，所述系统还包括与行为检测单元相连接的标签属性权重库建立单元，所述标签属性权重库建立单元包括：

10 参照群体用户确定模块，配置以选择与所述目标群体用户相关的参照群体用户；

测试标签生成模块，配置以根据参照群体用户中每一用户的用户行为，打上相应的标签；

属性权重确定模块，配置以根据所述参照群体用户中每一用户的标签，
15 确定参照群体用户中每一用户在各种属性下的属性权重；

权重标签赋予模块，配置以将每一用户在各种属性下的属性权重赋予到与属性相对应的标签中；

标签属性确定模块，配置以按照标签的种类对参照群体用户中所有用户的属性权重加权平均，确定参照群体用户在每种标签下各种属性的参照属性
20 权重；

属性权重库建立模块，配置以根据标签、标签下的属性和参照属性权重建立标签属性权重库。

本实施例中的标签规则库建立单元可以为一个服务器或者服务器集群，其中每个模块可以是单独的服务器或者服务器集群，此时，上述各模块之间的交互表现为各模块所对应的服务器或者服务器集群之间的交互，上述多个服务器或者服务器集群共同构成上述标签规则库建立单元以用于构成本发明的群体用户画像系统。

在一种替代实施例中，可以是上述多个模块中的几个模块共同组成一个服务器或者服务器集群。

通过本实施例，由此建立的标签属性权重库中的参照属性权重指代的是大众平均化的标签属性权重，将此参照属性权重应用于目标群体用户画像中，降低了个体用户画像的误差对群体用户画像的干扰，提高了群体用户画像的精确性。

作为本发明实施例系统的一种优选实施例，所述标签属性权重库建立单元还包括权重模型确定模块，所述权重模型确定模块配置以：根据参照群体用户中每一用户的历史表现，基于用户属性挖掘模型，确定每一用户在各种属性下的属性权重。常见的用户挖掘模型有：常见的用户属性挖掘模型包括 svm、贝叶斯、聚类、加权平均等各种算法模型。

本实施例中的标签属性权重库建立单元可以为一个服务器或者服务器集群，其中的权重模型确定模块可以是单独的服务器或者服务器集群，此时，上述单独的服务器或者服务器集群所构成的权重模型确定模块用于构成上述标签属性权重库建立单元以用于构成本发明的群体用户画像系统。

在本实施例中，基于参照群体用户中每一用户的历史表现对每一用户的属性权重进行推导，是对用户属性的一种模糊化处理，并将所得的属性权重赋予到相应的标签中，使得标签相对于属性具有更高的参考价值，进一步地使得基于标签所获得的群体用户画像更加精确。

5

在本发明实施例系统的一种优选实施例中，所述标签属性权重库建立单元还包括扩充修正模块，所述扩充修正模块配置以：在根据标签、标签下的属性和参照属性权重建立标签属性权重库后，定期补充参照群体用户中的用户数量，对参照属性权重进行修正更新。

10 本实施例中的标签属性权重库建立单元可以为一个服务器或者服务器集群，其中的扩充修正模块可以是单独的服务器或者服务器集群，此时，上述单独的服务器或者服务器集群所构成的扩充修正模块用于构成上述标签属性权重库建立单元以用于构成本发明的群体用户画像系统。

15 在本实施例中，当参照群体用户中参照用户的数量越多，参照属性权重就越精确，通过定期补充参照群体用户中的用户数量，对参照属性权重进行修正更新，由此实现了参照属性权重的定期修正，更保障了群体用户画像的精准性。

20 作为本发明实施例系统的一种优选实施例，所述系统还包括与用户画像单元相连接的信息推送单元，所述信息推送单元配置以：

在完成群体用户画像后，根据群体用户画像为群体用户进行个性化信息推送，继续检测群体用户在收到所述个性化信息推送后的行为以重新确定用户属性权重。

本实施例中的信息推送单元和用户画像生成单元可以分别为一个服务器或者服务器集群，此时，用户画像生成单元和信息推送单元之间的交互表现为其分别所对应的服务器或者服务器集群之间的交互，上述服务器或者服务器集群组建成信息推送单元以用于构成本发明的群体用户画像系统。

5 在本实施例中，基于群体用户画像为群体用户进行个性化信息推送，根据群体用户对个性化信息推送反馈的用户行为，重新确定群体用户的属性权重，实现了对群体用户属性权重和群体用户画像的校准，同时也避免了对群体用户信息推送的一成不变。

10 作为本发明实施例系统的一种优选实施例，图 3 中的属性权重确定单元具体包括：

个体权重确定模块，配置以将所述目标群体用户中的一个用户的所有标签放入所述参照所述标签属性权重库中，确定所述一个用户的所有标签的各种属性的属性权重；

15 个体权重均衡模块，配置以按照属性的种类对所述属性权重进行加权平均，确定所述一个用户在各种属性下的属性权重；

群体权重确定模块，配置以重复调用个体权重确定模块及个体权重平均模块，并以此类推，确定所述目标群体用户中的所有用户在各种属性下的属性权重；

20 群体权重均衡模块，配置以对所述目标群体用户中的所有用户在各种属性下的属性权重加权平均，确定目标群体用户在各种属性下的属性权重。

本实施例中的属性权重确定单元可以为一个服务器或者服务器集群，其中的每个模块可以是单独的服务器或者服务器集群，此时，上述各模块之间

的交互表现为各模块所对应的服务器或者服务器集群之间的交互，上述多个服务器或者服务器集群共同构成上述属性权重确定单元以用于构成本发明的群体用户画像系统。

在一种替代实施例中，可以是上述多个模块中的几个模块共同组成一个

5 服务器或者服务器集群。

在本实施例中，一方面，通过目标群体用户的多种标签共同确定目标群体用户属性，由此提高了所获得的群体用户属性的精确度；另一方面，通过对目标群体用户中的所有用户在各种属性下的属性权重进行加权平均，由此避免了因个体用户画像差异对群体用户画像的干扰，提高了群体用户画像的

10 精确度。

如图 4 所示，为本实施本发明的实施例的群体用户画像的方法和系统的架构图，包括用户画像服务器 40、群体用户 $A_1 \sim A_n$ ，多个访问服务器 $C_1 \sim C_i$ ，本架构图中访问服务器 $C_1 \sim C_i$ 在服务完成群体用户 $A_1 \sim A_n$ 中的多个待画像的

15 个体用户 M 通过客户端（客户端至少为智能终端）所发送的访问请求之后，用户画像服务器 40 根据各访问服务器 $C_1 \sim C_i$ 中各群体用户 $A_1 \sim A_n$ 的服务访问请求的缓存信息执行本发明如图 1 所示的群体用户画像方法，以获取群体用户 $A_1 \sim A_n$ 较精准的群体用户画像结果。

20 参见图 5 示出了适于用来实现本申请实施例的终端设备或服务器的计算机系统的结构示意图，其中计算机系统包括中央处理单元（CPU）501，其可以根据存储在只读存储器（ROM）502 中的程序或者从存储部分 508 加载到

随机访问存储器（RAM）503 中的程序而执行各种适当的动作和处理。在 RAM503 中，还存储有系统操作所需的各种程序和数据。CPU 501、ROM 502 以及 RAM503 通过总线 504 彼此相连。输入/输出（I/O）接口 505 也连接至总线 504。

5 以下部件连接至 I/O 接口 505：包括键盘、鼠标等的输入部分 506；包括诸如阴极射线管（CRT）、液晶显示器（LCD）等以及扬声器等的输出部分 507；包括硬盘等的存储部分 508；以及包括诸如 LAN 卡、调制解调器等的网络接口卡的通信部分 509。通信部分 509 经由诸如因特网的网络执行通信处理。驱动器 510 也根据需要连接至 I/O 接口 505。可拆卸介质 511，诸如磁盘、光
10 盘、磁光盘、半导体存储器等等，根据需要安装在驱动器 510 上，以便于从其上读出的计算机程序根据需要被安装入存储部分 508。

特别地，根据本发明的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本发明的实施例包括一种计算机程序产品，其包括有形地包含在机器可读介质上的计算机程序，上述计算机程序包含用于执
15 行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信部分 509 从网络上被下载和安装，和/或从可拆卸介质 511 被安装。

在本发明一方面的应用上，本发明实施例中的群体用户画像系统可以是作为功能元件的形式内嵌于网站服务器中；作为本发明的另一方面的应用，
20 本发明实施例中的群体用户画像系统还可以内嵌于云计算服务器中，此云计算服务器连接于网站服务器和用户终端之间。

需要说明的是，在不冲突的情况下，本发明中的实施例及优选实施例中所涉及到的技术特征彼此之间可以相互组合；术语“包括”、“包含”，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、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包
5 括……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

本发明实施例中可以通过硬件处理器来实现相关功能模块和单元。

以上所描述的装置实施例仅仅是示意性的，其中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
10 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模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下，即可以理解并实施。

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各实
15 施方式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。基于这样的理解，上述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，该计算机软件产品可以存储在计算机可读存储介质中，如 ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备）执行各个实施
20 例或者实施例的某些部分所述的方法。

最后应说明的是：以上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。

权 利 要 求 书

1、一种群体用户画像方法，包括：

根据包括标签、标签下的属性和参照属性权重的标签属性权重库，检测目标群体用户中每一用户的用户行为，打上相应的标签；

5 参照所述标签属性权重库，确定所述目标群体用户的所有用户的各种标签下的各种属性的属性权重，按照属性的种类对所述属性权重进行加权平均，确定目标群体用户在各种属性下的属性权重；

根据所确定的目标群体用户在各种属性下的属性权重，得到所述目标群体用户的群体用户画像。

10

2、根据权利要求 1 所述的方法，其特征在于，在所述根据包括标签、标签下的属性和参照属性权重的标签属性权重库，检测目标群体用户中每一用户的用户行为，打上相应的标签之前，还包括建立标签属性权重库：

选择与所述目标群体用户相关的参照群体用户；

15 根据参照群体用户中每一用户的用户行为，打上相应的标签；

根据所述参照群体用户中每一用户的标签，确定参照群体用户中每一用户在各种属性下的属性权重；

将每一用户在各种属性下的属性权重赋予到与属性相对应的标签中；

按照标签的种类对参照群体用户中所有用户的属性权重加权平均，确定

20 参照群体用户在每种标签下各种属性的参照属性权重；

根据标签、标签下的属性和参照属性权重建立标签属性权重库。

3、根据权利要求 1 所述的方法，其特征在于，所述参照所述标签属性权重库，确定所述目标群体用户的所有用户的各种标签下的各种属性的属性权重，按照属性的种类对所述属性权重进行加权平均，确定目标群体用户在各种属性下的属性权重包括：

5 将所述目标群体用户中的一个用户的所有标签放入所述标签属性权重库中，确定所述一个用户的所有标签的各种属性的属性权重；

按照属性的种类对所述属性权重进行加权平均，确定所述一个用户在各种属性下的属性权重；

10 以此类推，确定所述目标群体用户中的所有用户在各种属性下的属性权重；

对所述目标群体用户中的所有用户在各种属性下的属性权重加权平均，确定目标群体用户在各种属性下的属性权重。

15 4、根据权利要求 2 所述的方法，其特征在于，在根据标签、标签下的属性和参照属性权重建立标签属性权重库后，定期补充参照群体用户中的用户数量，对参照属性权重进行修正更新。

5、根据权利要求 2 所述的方法，其特征在于，所述确定参照群体用户中每一用户在各种属性下的属性权重包括：

20 根据参照群体用户中每一用户的历史表现，基于用户属性挖掘模型，确定每一用户在各种属性下的属性权重。

6、一种群体用户画像系统，包括：

行为检测单元，配置以：根据包括标签、标签下的属性和参照属性权重的标签属性权重库，检测目标群体用户中每一用户的用户行为，打上相应的标签；

属性权重确定单元，配置以：参照所述标签属性权重库，确定所述目标
5 群体用户的所有用户的各种标签下的各种属性的属性权重，按照属性的种类对所述属性权重进行加权平均，确定目标群体用户在各种属性下的属性权重；

用户画像生成单元，配置以：根据所确定的目标群体用户在各种属性下的属性权重，得到所述目标群体用户的群体用户画像。

10

7、根据权利要求 6 所述的系统，其特征在于，所述系统还包括标签属性权重库建立单元，所述标签属性权重库建立单元配置以：

选择与所述目标群体用户相关的参照群体用户；

根据参照群体用户中每一用户的用户行为，打上相应的标签；

15 根据所述参照群体用户中每一用户的标签，确定参照群体用户中每一用户在各种属性下的属性权重；

将每一用户在各种属性下的属性权重赋予到与属性相对应的标签中；

按照标签的种类对参照群体用户中所有用户的属性权重加权平均，确定参照群体用户在每种标签下各种属性的参照属性权重；

20 根据标签、标签下的属性和参照属性权重建立标签属性权重库。

8、根据权利要求 6 所述的系统，其特征在于，所述属性权重确定单元配置以：

将所述目标群体用户中的一个用户的所有标签放入所述标签属性权重库中，确定所述一个用户的所有标签的各种属性的属性权重；

按照属性的种类对所述属性权重进行加权平均，确定所述一个用户在各种属性下的属性权重；

5 以此类推，确定所述目标群体用户中的所有用户在各种属性下的属性权重；

对所述目标群体用户中的所有用户在各种属性下的属性权重加权平均，确定目标群体用户在各种属性下的属性权重。

10 9、根据权利要求 7 所述的系统，其特征在于，所述标签属性权重库还包括扩充修正模块，所述扩充修正模块配置以：在根据标签、标签下的属性和参照属性权重建立标签属性权重库后，定期补充参照群体用户中的用户数量，对参照属性权重进行修正更新。

15 10、根据权利要求 7 所述的系统，其特征在于，所述标签属性权重库还包括属性权重确定模块，所述属性权重确定模块配置以：根据参照群体用户中每一用户的历史表现，基于用户属性挖掘模型，确定每一用户在各种属性下的属性权重。

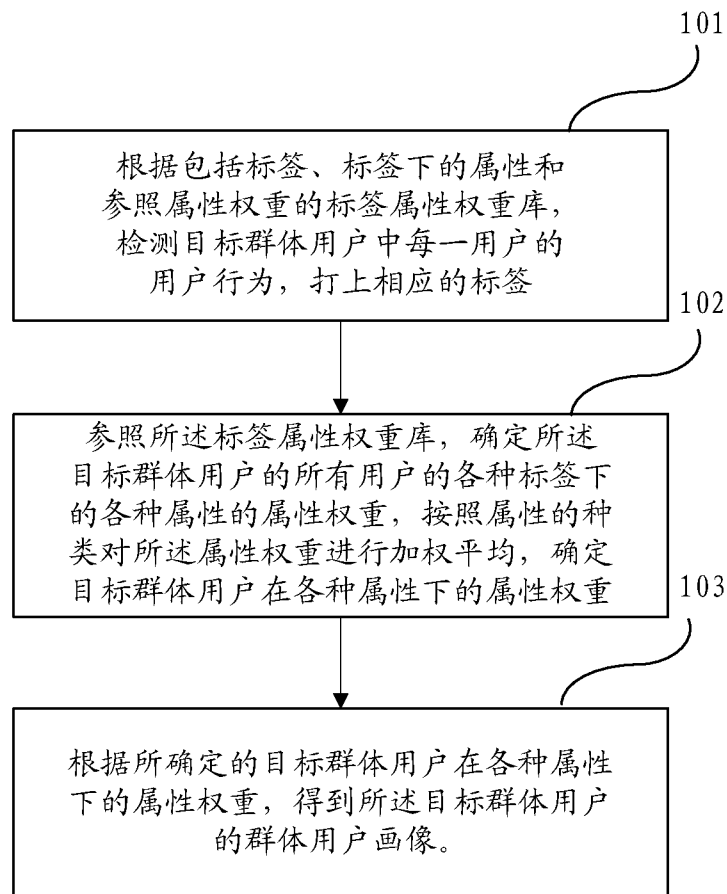


图 1

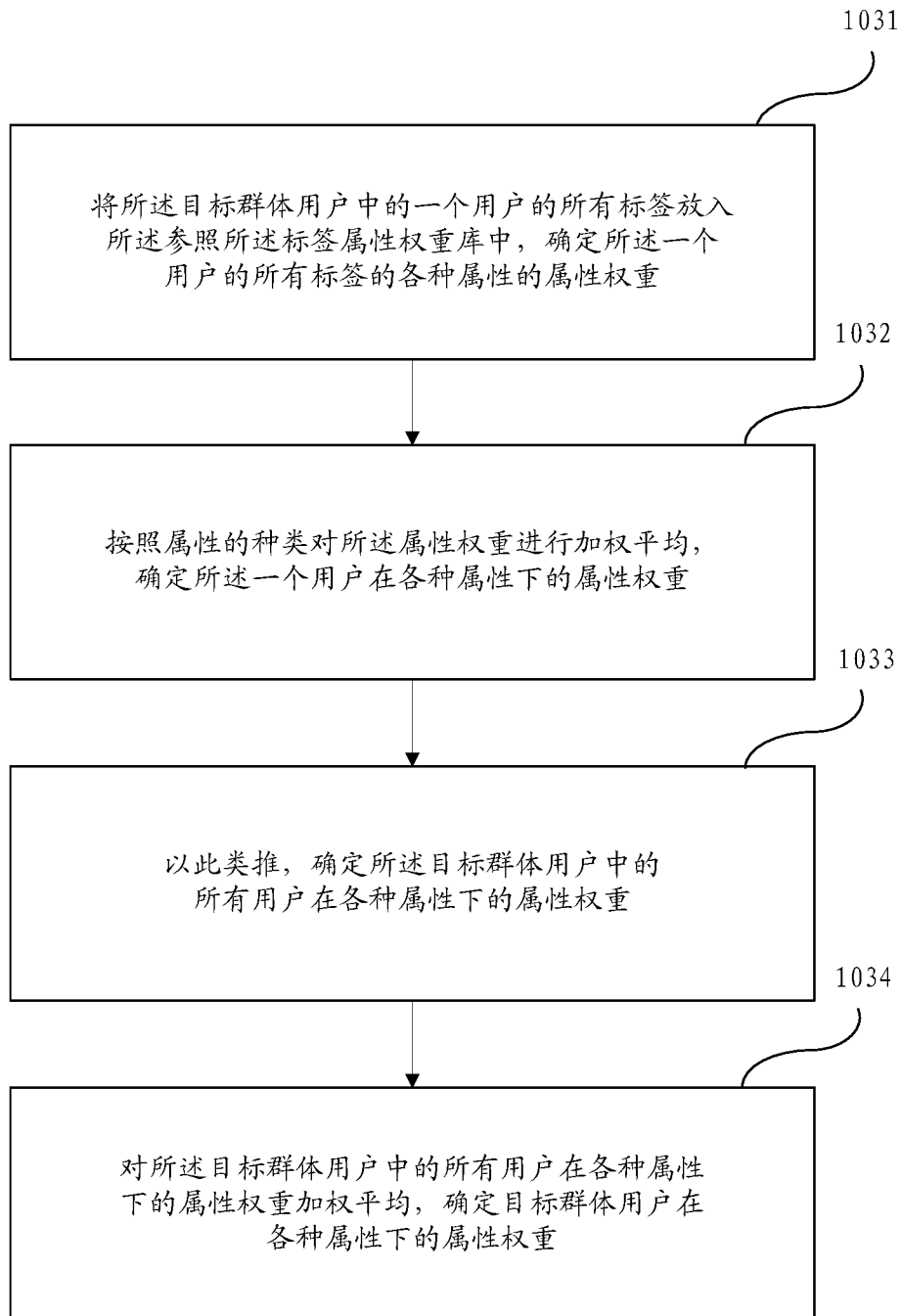


图 2



图 3

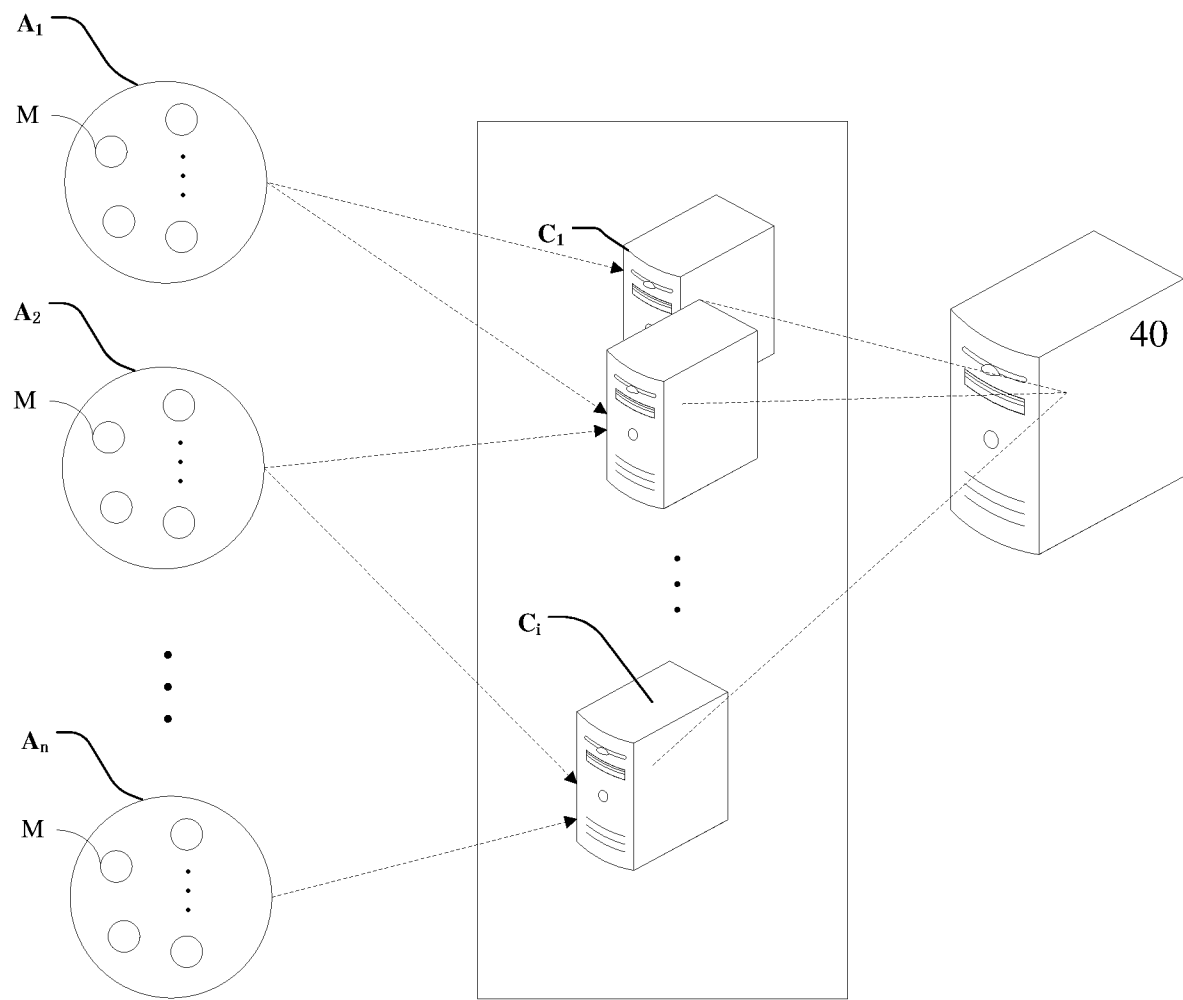


图 4

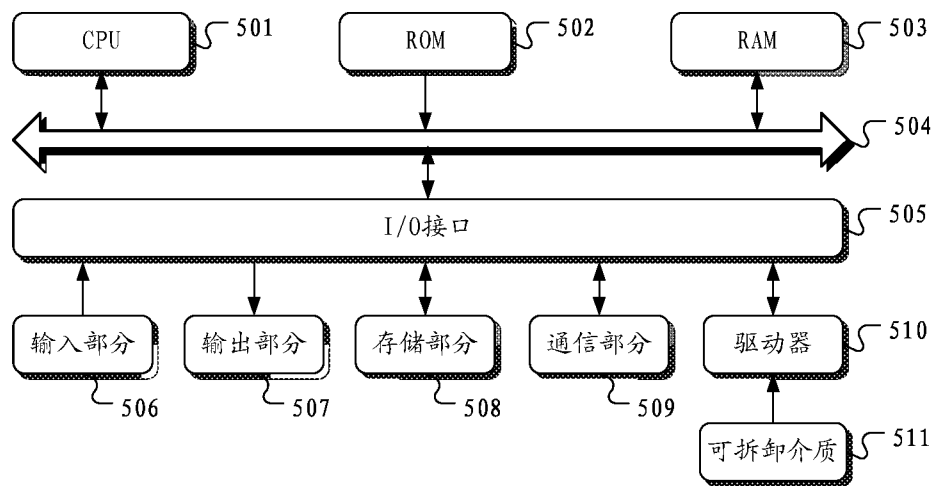


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/083164

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) 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)

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)

CNKI, CNPAT, WPI, EPODOC, IEEE: probability; profile, portrayal, user, crowd, group, character, characteristics, tag, label, attribute?, weight, membership

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103778555 A (BEIJING GEO POLYMERIZATION TECHNOLOGY CO., LTD.), 07 May 2014 (07.05.2014), description, paragraphs [0041]-[0054]	1-10
A	CN 103577549 A (FUDAN UNIVERSITY), 12 February 2014 (12.02.2014), the whole document	1-10
A	CN 104867037 A (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD. et al.), 26 August 2015 (26.08.2015), the whole document	1-10
A	US 2008275861 A1 (GOOGLE INC.), 06 November 2008 (06.11.2008), the whole document	1-10
A	EP 2456238 A1 (ALCATEL LUCENT), 23 May 2012 (23.05.2012), the whole document	1-10

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

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 17 August 2016 (17.08.2016)	Date of mailing of the international search report 26 August 2016 (26.08.2016)
Name and mailing address of the ISA/CN: 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 CAO, Shumei Telephone No.: (86-10) 010-62413688

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2016/083164

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103778555 A	07 May 2014	None	
CN 103577549 A	12 February 2014	None	
CN 104867037 A	26 August 2015	HK 1212073 A1	03 June 2016
US 2008275861 A1	06 November 2008	US 2012054205 A1	01 March 2012
		WO 2008134772 A1	06 November 2008
		KR 20100051593 A	17 May 2010
		BR PI0811488 A2	04 November 2014
EP 2456238 A1	23 May 2012	None	

国际检索报告

国际申请号

PCT/CN2016/083164

A. 主题的分类

G06F 17/30(2006.01)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)

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))

CNKI, CNPAT, WPI, EPODOC, IEEE:用户, 画像, 群体, 群组, 人群, 标签, 属性, 特征, 权重, 概率, 隶属度; profile, portrayal, user, crowd, group, character, characteristics, tag, label, attribute?, weight, membership

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 103778555 A (北京集奥聚合科技有限公司) 2014年 5月 7日 (2014 - 05 - 07) 说明书第[0041]-[0054]段	1-10
A	CN 103577549 A (复旦大学) 2014年 2月 12日 (2014 - 02 - 12) 全文	1-10
A	CN 104867037 A (北京京东尚科信息技术有限公司 等) 2015年 8月 26日 (2015 - 08 - 26) 全文	1-10
A	US 2008275861 A1 (GOOGLE INC.) 2008年 11月 6日 (2008 - 11 - 06) 全文	1-10
A	EP 2456238 A1 (ALCATEL LUCENT) 2012年 5月 23日 (2012 - 05 - 23) 全文	1-10

☐ 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

2016年 8月 17日

国际检索报告邮寄日期

2016年 8月 26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

授权官员

曹妹妹

传真号 (86-10)62019451

电话号码 (86-10)010-62413688

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/083164

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	103778555	A	2014年 5月 7日	无			
CN	103577549	A	2014年 2月 12日	无			
CN	104867037	A	2015年 8月 26日	HK	1212073	A1	2016年 6月 3日
US	2008275861	A1	2008年 11月 6日	US	2012054205	A1	2012年 3月 1日
				WO	2008134772	A1	2008年 11月 6日
				KR	20100051593	A	2010年 5月 17日
				BR	PI0811488	A2	2014年 11月 4日
EP	2456238	A1	2012年 5月 23日	无			

表 PCT/ISA/210 (同族专利附件) (2009年7月)