

(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
2020 年 4 月 30 日 (30.04.2020)



(10) 国际公布号
WO 2020/082596 A1

- (51) 国际专利分类号:
G06F 16/9535 (2019.01)
- (21) 国际申请号: PCT/CN2018/124669
- (22) 国际申请日: 2018 年 12 月 28 日 (28.12.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201811237829.8 2018年10月23日 (23.10.2018) CN
- (71) 申请人: 深圳壹账通智能科技有限公司(ONE CONNECT SMART TECHNOLOGY CO., LTD. (SHENZHEN)) [CN/CN]; 中国广东省深圳市前海深港合作区前湾一路1号A栋201室, Guangdong 518052 (CN)。
- (72) 发明人: 张仁娟(ZHANG, Renjuan); 中国广东省深圳市前海深港合作区前湾一路1号A栋201室, Guangdong 518052 (CN)。
- (74) 代理人: 北京英特普罗知识产权代理有限公司(INTELLECPRO CHINA LIMITED); 中国北京市西城区车公庄大街9号五栋大楼C座11层, Beijing 100044 (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, 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))。

(54) **Title:** DATA PROCESSING-BASED AUTOMATIC USER PROFILE GENERATING METHOD AND SYSTEM

(54) 发明名称: 基于数据处理的用户画像自动生成方法和系统

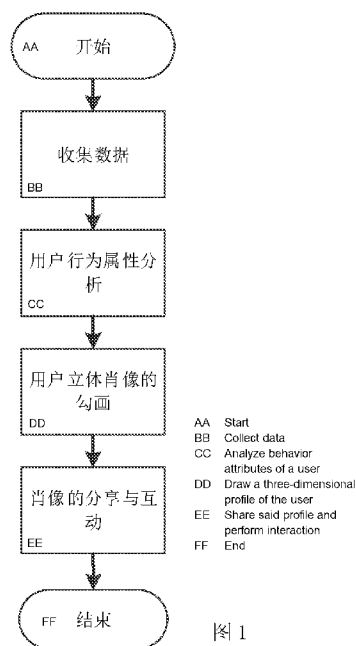


图 1

(57) **Abstract:** The present disclosure relates to a data processing-based automatic user profile generating method and system. Said method comprises the following steps: step 1, collecting user information, the user information comprising static information and behavior information about a user; step 2, generating a specific label of the user according to the collected user information; step 3, quantizing the generated label according to a predetermined rule, and calculating attribute values of the user; and step 4, generating a user profile according to the calculated attribute values of the user. The beneficial effects of the present application mainly lie in: 1, reducing user operations and improving user experiences; and 2, providing enjoyment, and facilitating, by means of interaction between users, promotion of a product and increasing the daily active users of the product.

(57) **摘要:** 本公开涉及一种基于数据处理的用户画像自动生成方法和系统, 所述方法包括如下步骤: 步骤1、收集用户信息, 所述用户信息包括用户的静态信息和行为信息; 步骤2、根据所收集的用户信息, 生成所述用户的特定标签; 步骤3、根据预定规则, 对所生成的标签进行量化, 计算得到所述用户的属性值; 步骤4、根据计算出的所述用户的属性值, 生成用户画像。本申请的有益效果主要在于: 1、减少用户操作及提高用户体验; 2、增加趣味性, 用户之间的互动可有助于进行产品的推广和增加产品的日活。

WO 2020/082596 A1

基于数据处理的用户画像自动生成方法和系统

本申请申明享有 2018 年 10 月 23 日递交的申请号为 201811237829.8、名称为“基于数
据处理的用户画像自动生成方法和系统”的中国专利申请的优先权，该中国专利申请的整
体内容以参考的方式结合在本申请中。

技术领域

本申请涉及基于互联网应用服务技术领域，尤其涉及一种基于数据处理的用户画像自
动生成方法和系统。

背景技术

随着互联网爆炸式的发展，每天都在产生着大量数据，如何快速抓取信息并生成用户
画像，也成了重要的课题。

在传统的方法中，需要对用户行为进行人工分类、判定，并在后期提供有针对性的服
务或其它处理应对。在海量数据的场景下，数据往往维度多、数据量大，人工很难将用户
行为相关的指标统计全面。此外，由于人会出现疲劳等情况，这种传统的人工识别的方法
准确率并不高。

在互联网逐渐步入大数据时代后，用户的行为在服务商面前都将是可视化的。服务商
的关注点日也开始益聚焦于怎样利用大数据来精准营销，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。大数据使得服务商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，提供了足够的数据基础。伴随着对人的了解逐步深入，用户画像（UserProfile）的概
念应运而生，其用于通过用户标签抽象出用户的信息全貌，可以看作服务商应用大数据的
根基。典型的用户画像是将用户信息标签化，就是服务商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商业全貌，可以看
作是服务商应用大数据技术的基本方式。

然而，目前，对用户标签的识别、用户画像的建立，主要还是通过人工干预和计算机
简单处理转换的方式完成，存在有以下不足：1、耗时很长；2、人工成本高；3、生成的结
果不够直观；4、信息录入有因人工导致的错误风险。

另外，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。

在现有技术中，对用户行为进行分类和预测的方法较为单一，参考效果不理想。由于用户行为包括线上和线下行为，数据来源复杂，存在这样的需求：开发能够针对不同的数据来源、结合多种分类预测技术而综合判定和预测用户属性、并生成用户画像的方案。

5 发明内容

考虑到现有技术的上述问题，发明人做出了本申请，本申请不需要用户上传图片，系统可以根据获取的大数据分析人物，大致勾画出人物画像并给出人物的幸运分，在人物画像的周围显示人物的智慧值，财力值，能力值，体力值等，用户也可以将人物画像分析分享给微信好友，好友可以打开并查看对方分享的人物剖析结果，经过授权后，同时可以查询自己的人物画像分析结果，从而有利于 APP 推广和增加 APP 的日活。

根据本申请的实施例，提供了一种基于数据处理的用户画像自动生成方法，其特征在于包括如下步骤：

步骤 1、收集用户信息，所述用户信息包括用户的静态信息和行为信息；

步骤 2、根据所收集的用户信息，生成所述用户的特定标签；

15 步骤 3、根据预定规则，对所生成的标签进行量化，计算得到所述用户的属性值；

步骤 4、根据计算出的所述用户的属性值，生成用户画像。

根据本申请的实施例，还提供了一种用于执行所述方法的用户画像自动生成系统，其包括：

20 用户信息收集模块，用于收集用户信息，所述用户信息包括用户的静态信息和行为信息；

用户信息分析模块，用于对用户信息进行分析，生成所述用户的特定标签；

用户属性计算模块，用于按照特定规则，对每种标签进行量化，再进行加权相加，获得用户的属性值；

用户画像生成模块，用于根据用户的标签和属性值，生成用户画像。

25 根据本申请的实施例，还提供了一种计算机可读存储介质，所述计算机可读存储介质上存储有用于上述方法的计算机可读指令，所述计算机可读指令被处理器执行时，执行如下步骤：

步骤 1、收集用户信息，所述用户信息包括用户的静态信息和行为信息；

步骤 2、根据所收集的用户信息，生成所述用户的特定标签；

30 步骤 3、根据预定规则，对所生成的标签进行量化，计算得到所述用户的属性值；

步骤 4、根据计算出的所述用户的属性值，生成用户画像。

本申请的有益效果主要在于：

- 1、减少用户操作及提高用户体验；
- 2、增加趣味性，用户之间的互动可有助于进行产品的推广和增加产品的日活；
- 3、考虑到数据的不同来源，进行差异化的处理，提高了处理的精细度和准确度；
- 4、以用户画像作为参照信息，大大提高相关业务的办理速度；
- 5、能够根据样本数据来源的不同，选择不同的分类模型级联和/或并联，使得客户关系管理更加精准。

附图说明

- 图 1 为示出根据本申请的实施例的基于数据处理的用户画像自动生成方法的概念示意图；
- 图 2 为示出根据本申请的实施例的基于数据处理的用户画像自动生成方法的流程示意图；
- 图 3 为根据本申请的实施例的基于数据处理的用户画像自动生成系统的功能模块组成示意图；
- 图 4 为根据本申请实施例的安装了应用程序的系统的运行环境的示意图。

具体实施方式

下面，结合附图对技术方案的实施作进一步的详细描述。

- 本领域的技术人员能够理解，尽管以下的说明涉及到有关本申请的实施例的很多技术细节，但这仅为用来说明本申请的原理的示例、而不意味着任何限制。本申请能够适用于不同于以下例举的技术细节之外的场合，只要它们不背离本申请的原理和精神即可。

另外，为了避免使本说明书的描述限于冗繁，在本说明书中的描述中，可能对可在现有技术资料中获得的部分技术细节进行了省略、简化、变通等处理，这对于本领域的技术人员来说是可以理解的，并且这不会影响本说明书的公开充分性。

- 下文中，将描述用于进行本申请的实施例。注意，将以下面的次序给出描述：1、发明构思的概要（图 1）；2、基于数据处理的用户画像自动生成方法（图 2）；3、基于数据处理的用户画像自动生成系统（图 3）；4、根据本申请的实施例的安装了应用程序的系统（图 4）。

1、发明构思的概要

如图 1 所示，本申请的构思主要包括以下方面。

一、大量数据收集

通过用户在网络或 APP 上的行为记录，收集用户在网上行为数据、用户信息、用户偏好记录、用户的交易数据等信息。

二、用户行为属性的分析

通过收集到的大量用户信息进行处理，通过机器学习对用户的行为、偏好进行猜测，同时也通过大概率事件将用户行为属性的信息划分为标签，再将标签通过算法得出用户的智慧值、财力值、能力值、体力值及魅力值等信息

三、用户立体画像的勾画

根据通过对用户行为属性分析划分的标签，按照标签提取，聚合分析后我们可以勾画出用户的立体画像，画像生成后可在周围展示其计算出的智慧值、财力值、能力值、体力值及魅力值。

四、画像的分享

生成画像后用户也可通过微信分享给好友，好友通过授权后可查看用户的画像值，并可给画像值打分。也可自己进行生成画像。增加用户与好友之间的互动。

下面，结合实施例来说明上述发明构思的实现。

2、基于数据处理的用户画像自动生成方法

图 2 为根据本申请的实施例的基于数据处理的用户画像自动生成方法的流程示意图。

如图 2 所示，本申请的实施例提供了一种基于数据处理的用户画像自动生成方法，主要包括以下步骤：

步骤 S100、收集用户信息，所述用户信息包括用户的静态信息和行为信息；

步骤 S200、对用户信息进行分析，生成所述用户的特定标签；

其中，所述标签可包括性别标签、学历标签、消费标签、运动标签、职位标签、购物标签、上网查阅标签、兴趣爱好标签以及活动地方标签，等等；

步骤 S300、按照特定规则，对每种标签进行量化，再进行加权相加，获得用户的属性值；

例如，学历标签被可量化为六个档次，最高档次为 100 分，最少档次为 16.67，其他标签类似；

再将标签通过算法得出用户的智慧值、财力值、能力值、体力值等信息，

每个属性值的算法如下：

智慧值=学历标签占 80%+购物标签 15%+其他类的标签占 5%；

财力值=消费标签占 60%+购物标签 30%+其他类标签占 10%；

能力值=职位标签占 40%+学历标签占 30%+消费标签占 20%+其他类标签占 10%；

体力值=运动标签占 50%+购物标签占 20%+其他标签占 30%；

步骤 S400、根据用户的标签和属性值，生成用户画像；

其中，用户画像包含量化分值较高的标签、以及属性值。

作为示例，所述静态信息是指用户相对稳定（不易随时间变化）的信息，其可包括公共数据（大数据）、或者机构内部的客户信息（用户注册的数据），例如，包括人口属性、商业属性等方面数据；

5 其中，人口属性包括：地域、年龄、性别、文化、职业、收入、生活习惯、消费习惯等；

其中，所述静态信息可直接对应于特定标签，例如，用户的性别可直接对应于标签“男性”或“女性”，用户的生育记录可对应于标签“妈妈”，等等。

10 其中，所述行为信息是指在网络或 APP 上的行为信息，包括用户的网站浏览行为信息、用户交易行为信息，等等，

其中，用户浏览行为包括用户浏览的活跃频率、产品喜好、使用习惯等；

可选地，步骤 S100 包括：

15 S101、对静态和行为信息数据进行预处理（数据清洗、筛选），根据预处理后的网络访问信息中获取用户在每个预设行为类别中的行为数据，使获取的同一类别的行为数据具有相同的格式。

S102、根据行为信息确定用户的行为类别，如下：

计算用户行为数据在预设的各个维度上的属性；

根据用户行为数据的来源、以及与所述来源相对应的所述属性，选择相应的分类模型；

根据所选的分类模型，对用户行为数据进行分类。

20 在步骤 S101 中，为提取出每个类别的行为数据，可对该网络访问信息进行预处理。对网络访问信息的预处理包括对网络访问信息进行变量采集、变量区间处理、极大极小规则处理、缺失值处理和格式处理等；

25 变量采集为从网络访问信息中采集出用户每次网络访问的访问时间、登录时间、浏览信息、搜索信息以及购买信息等等，比如访问一个具体的电商网站时的访问时间、登录时间、浏览信息、搜索信息以及购买信息。服务器在采集出用户每次访问的访问时间、登录时间、浏览信息、搜索信息以及购买信息等信息时，可调用相关的累加器或计算器等对应统计出用户在预设时间段内的登录次数、购买次数、浏览次数和搜索次数、购买金额，等等。

30 变量区间处理为根据业务的规则为每个变量划定区间，并将划定的区间映射为具有业务指标，从而作为后续的数值输入，以计算用户行为熵等特征。例如，用户的上述次数登录次数、购买金额可分别被划分到多个区间中的一个，每个区间对应于具体数值，例如，与次数或金额相关的用户行为可对应于规范化的指标（0 至 100）。

极大极小的规则处理包括对所采集的网络访问信息所包含的数值大小的处理，以降低异常数据对用户的行为分类判断的干扰。具体的，可对所采集的网络访问信息中的用户的年龄进行极大极小的规则处理。比如，对于年龄为-1、0、或 999 岁等等，明显不符合正常用户年龄的数据，对其进行极大极小规则处理。

5 缺失值处理是指所采集网络访问信息中包含的预设行为类别中的行为数据不存在时，可对其进行缺失值处理。如将其标记为“0”，或采用其它信息代替等等。比如，用户采用匿名访问或不登录用户名而直接访问相关的购物网站时，服务器所记录的用户的登录信息则缺失。服务器可对该类信息进行缺失值处理，如可获取用户的访问终端的唯一标识，将该唯一标识作为和用户的登录名进行关联。

10 格式处理包括对网络访问信息中包含的时间信息的格式的处理，使其格式保持相同。比如，对于所记录的用户的登录时间等时间信息，比如所记录到的时间信息包括 20091011 和 2009-10-11 以及 2009 年 10 月 11 日等形式，可将其全部转换成统一格式，如 20091011。

作为示例，根据市场上已有的信息，设定一些用户行为作为基本固定标签。已设定的标签规则举例如下：

15 性别标签：性别标签很简单分为男性标签与女性标签，按照获取用户的行为信息去决定他是男性标签还是女性标签；

学历标签：学历标签分为初中及以下、高中、专科、本科、硕士、博士及以上不同档次；

消费标签：消费分为高档消费、中档消费、低档消费不同档次；

20 运动标签：运动分为散步、跑步、健身等不同档次；

购物标签：购物标签按照购买物进行区分，分为书籍类、衣服包包类、生活用品类、健身及运动器材类等等；

其他标签也类似做此区分。

针对于上述标签，作为示例，步骤 S200 包括：

25 S201、根据用户行为的类别，生成对应场景下的用户标签。

例如，用户对特定网站/网址的访问，表明用户对该网站所对应的特定内容/业务有兴趣、偏好、需求等等。由此可以生成用户在该网站（例如，购物网站）所对应的场景（例如，购物场景）下的用户标签（例如，“运动爱好者”），从而，对于同一用户，可在购物场景下生成并显示一系列对应标签，而在其他场景（例如，社交场景）下生成并显示另外一系列对应标签，从而使后续产生的用户画像具有多维特性，对于同一用户，可根据场景而呈现不同的用户画像。

此外，对用户有些比较复杂的行为，不能按照已设定的区分，可以将对这些行为进行

大概率事件分析的模式进行归类。比如，如果获取的用户行为为用户即在高档购物广场中很贵的餐厅吃饭及购买名牌包包、同一天也在低档的大排档消费，则不好判定此用户属于消费标签中的哪一类，这时，可要按照大概率事件去处理，大概率计算公式如下：

$P(A) = m/n$ 。其中，“(A)”表示事件。“m”表示事件(A)发生的总数。“n”是总事件发生的总数。

若P大于50%以上，即设定为A事件为大概率事件，即将此类事件划分为到该标签的统计事件的范围中，即，所述标签的量化值取决于所关联的特定事件的发生情况。

可选地，在步骤S400之后，在社交平台上，所生成的用户画像可被公开或被发送给其它用户，其中，各个用户可根据相互之间的浏览权限而查看特定的用户画像；

作为示例，在步骤S400之后可包括：

步骤S500、根据各个用户之间的评价权限，用户可给其它用户的属性值进行打分，然后，后台系统可根据打分情况，来对用户的属性值进行修正；例如，通过对其他用户的打分值取一定权重（例如根据打分用户的数量取10~30%的权重），与当前的属性值进行加权求和，生成更新的属性值，从而进一步更新用户画像，其中，后台系统指运行本方法的系统，即，生成用户画像的系统。

3、基于数据处理的用户画像自动生成系统

图3为根据本申请的实施例的基于数据处理的用户画像自动生成系统的功能模块示意图。如该图所示，根据本申请的实施例，提供了一种基于数据处理的用户画像自动生成系统，其主要包括用户信息收集模块、用户信息分析模块、用户属性计算模块、用户画像生成模块。

其中，所述用户信息收集模块用于收集用户信息，所述用户信息包括用户的静态信息和行为信息；

所述用户信息分析模块用于对用户信息进行分析，生成所述用户的特定标签；

其中，所述标签可包括性别标签、学历标签、消费标签、运动标签、职位标签、购物标签、上网查阅标签、兴趣爱好标签以及活动地方标签，等等；

所述用户属性计算模块用于按照特定规则，对每种标签进行量化，再进行加权相加，获得用户的属性值；

例如，学历标签被可量化为六个档次，最高档次为100分，最少档次为16.67，其他标签类似；

再将标签通过算法得出用户的智慧值、财力值、能力值、体力值等信息，

每个属性值的算法如下：

智慧值=学历标签占 80%+购物标签 15%+其他类的标签占 5%;

财力值=消费标签占 60%+购物标签 30%+其他类标签占 5%;

能力值=职位标签占 40%+学历标签占 30%+消费标签占 20%+其他类标签占 10%;

体力值=运动标签占 50%+购物标签占 20%+其他标签占 30%;

5 所述用户画像生成模块用于根据用户的标签和属性值,生成用户画像。

此外,本申请的不同实施例也可以通过软件模块或存储在一个或多个计算机可读介质上的计算机可读指令的方式实现,其中,所述计算机可读指令是当被处理器或设备组件执行时,执行本申请所述的不同的实施例。类似地,软件模块、计算机可读介质和硬件部件的任意组合都是本申请预期的。所述软件模块可以被存储在任意类型的计算机可读存储介
10 质上,例如 RAM、EPROM、EEPROM、闪存、寄存器、硬盘、CD-ROM、DVD 等等。

4、根据本申请的实施例的安装了应用程序的系统

参照图 4,其示出了根据本申请实施例的安装了应用程序的系统的运行环境。

在本实施例中,所述的安装应用程序的系统安装并运行于电子装置中。所述电子装置
15 可以是桌上型计算机、笔记本、掌上电脑及服务器等计算设备。该电子装置可包括但不限于存储器、处理器及显示器。该图仅示出了具有上述组件的电子装置,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。

所述存储器在一些实施例中可以是所述电子装置的内部存储单元,例如该电子装置的硬盘或内存。所述存储器在另一些实施例中也可以是所述电子装置的外部存储设备,例如
20 所述电子装置上配备的插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)等。进一步地,所述存储器还可以既包括所述电子装置的内部存储单元也包括外部存储设备。所述存储器用于存储安装于所述电子装置的应用软件及各类数据,例如所述安装应用程序的系统的计算机可读指令等。所述存储器还可以用于暂时地存储已经输出或者将要输出的数据。

25 所述处理器在一些实施例中可以是中央处理单元(Central Processing Unit, CPU)、微处理器或其他数据处理芯片,用于运行所述存储器中存储的程序代码或处理数据,例如执行所述安装应用程序的系统等。

所述显示器在一些实施例中可以是 LED 显示器、液晶显示器、触控式液晶显示器以及
30 OLED(Organic Light-Emitting Diode, 有机发光二极管)触摸器等。所述显示器用于显示在所述电子装置中处理的信息以及用于显示可视化的客户界面,例如应用菜单界面、应用图标界面等。所述电子装置的部件通过系统总线相互通信。

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,上述实施方式中的

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商品的形式体现出来，该计算机软件商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
5 以是手机，计算机，服务器，空调器，或者网络设备等）执行本申请本申请各个实施例所述的方法。

也就是说，根据本申请的实施例，还提供了一种计算机可读存储介质，所述计算机可读存储介质上存储有用于执行根据本申请的实施例的所述方法的计算机可读指令，所述计算机可读指令被处理器执行时，执行所述方法的各个步骤。

10 由上，将理解，为了说明的目的，这里已描述了本申请的具体实施例，但是，可作出各个修改，而不会背离本申请的范围。本领域的技术人员将理解，流程图步骤中所绘出或这里描述的操作和例程可以多种方式变化。更具体地，可重新安排步骤的次序，可并行执行步骤，可省略步骤，可包括其它步骤，可作出例程的各种组合或省略。因而，本申请仅由所附权利要求限制。

权利要求书

1、一种基于数据处理的用户画像自动生成方法，其特征在于包括如下步骤：

步骤 1、收集用户信息，所述用户信息包括用户的静态信息和行为信息；

5 步骤 2、根据所收集的用户信息，生成所述用户的特定标签；

步骤 3、对所生成的标签进行量化，计算得到所述用户的属性值；

步骤 4、根据计算出的所述用户的属性值，生成用户画像。

2、根据权利要求 1 所述的用户画像自动生成方法，其特征在于，所述步骤 3 包括：

步骤 3-1、将所述用户的各个标签量化为多个标准值；

10 步骤 3-2、针对于所述用户的某个属性，对所述用户的部分标签的量化后的标准值对进行加权求和，生成所述用户的所述属性的属性值。

3、根据权利要求 2 所述的用户画像自动生成方法，其特征在于，所述标签包括性别标签、学历标签、消费标签、运动标签、职位标签、购物标签、上网查阅标签、兴趣爱好标签、以及活动地方标签；

15 其中，所述用户画像显示量化值较高的标签以及属性值。

4、根据权利要求 3 所述的用户画像自动生成方法，其特征在于，所述属性值包括智慧值、财力值、能力值，所述智慧值、所述财力值、所述能力值计算公式为：

智慧值=学历标签×80%+购物标签×15%+其他类的标签×5%；

财力值=消费标签×60%+购物标签×30%+其他类标签×10%；

20 能力值=职位标签×40%+学历标签×30%+消费标签×20%+其他标签×10%；

体力值=运动标签×50%+购物标签×20%+其他标签×30%；

其中，上述公式中的各个标签代表其量化后的标准值。

5、根据权利要求 3 所述的用户画像自动生成方法，其特征在于，所述静态信息包括公共数据、以及机构内部的客户信息，所述客户信息包括人口属性、商业属性数据；

25 其中，人口属性包括：地域、年龄、性别、文化、职业、收入、生活习惯、消费习惯。

6、根据权利要求 5 所述的用户画像自动生成方法，其特征在于，所述行为信息包括所述用户在网络或应用程序上的行为信息、用户的网站浏览行为信息、用户交易行为信息，其中，所述标签的量化值取决于所述行为信息；

通过以下公式计算用户的行为信息所对应的行为的发生概率：

30 $P(A) = m/n$ ，

其中，“A”表示行为 A，“m”表示行为 A 发生的总数，“n”是全部行为发生的总数；

之后，若 P 大于特定阈值，则确定行为 A 为大概率事件，之后，并将行为 A 关联到所

述标签。

7、根据权利要求1所述的用户画像自动生成方法，其特征在于，所述步骤1包括：

步骤1-1、对静态信息和行为信息进行预处理，根据预处理后的行为信息获取用户在每个预设行为类别中的行为数据，使获取的同一类别的行为数据具有相同的格式。

5 8、根据权利要求1所述的用户画像自动生成方法，其特征在于，在所述步骤4之后还包括：

步骤5、在社交平台上，根据各个用户之间的评价权限，各个用户可给其它用户的属性进行打分，然后，后台系统根据打分情况，来对用户的属性值进行修正。

10 9、一种用于执行根据权利要求1至8中的任一个所述的方法的基于数据处理的用户画像自动生成系统，其特征在于包括：

用户信息收集模块，用于收集用户信息，所述用户信息包括用户的静态信息和行为信息；

用户信息分析模块，用于对用户信息进行分析，生成所述用户的特定标签；

用户属性计算模块，用于对所生成的标签进行量化，计算得到所述用户的属性值；

15 用户画像生成模块，用于根据用户的标签和属性值，生成用户画像。

10、根据权利要求9所述的基于数据处理的用户画像自动生成系统，其特征在于，所述标签可包括性别标签、学历标签、消费标签、运动标签、职位标签、购物标签、上网查阅标签、兴趣爱好标签以及活动地方标签。

20 11、根据权利要求9所述的基于数据处理的用户画像自动生成系统，其特征在于，所述属性值包括智慧值、财力值、能力值，所述智慧值、所述财力值、所述能力值计算公式为：

智慧值=学历标签 $\times$ 80%+购物标签 $\times$ 15%+其他类的标签 $\times$ 5%；

财力值=消费标签 $\times$ 60%+购物标签 $\times$ 30%+其他类标签 $\times$ 10%；

能力值=职位标签 $\times$ 40%+学历标签 $\times$ 30%+消费标签 $\times$ 20%+其他标签 $\times$ 10%；

25 体力值=运动标签 $\times$ 50%+购物标签 $\times$ 20%+其他标签 $\times$ 30%；

其中，上述公式中的各个标签代表其量化后的标准值。

12、根据权利要求9所述的基于数据处理的用户画像自动生成系统，其特征在于，所述静态信息包括公共数据、以及机构内部的客户信息，所述客户信息包括人口属性、商业属性数据；

30 其中，人口属性包括：地域、年龄、性别、文化、职业、收入、生活习惯、消费习惯。

13、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有用于执行根据权利要求1至8中的任一个所述的方法的计算机可读指令，所述计算机可读指令

被处理器执行时，执行以下步骤：

步骤 1、收集用户信息，所述用户信息包括用户的静态信息和行为信息；

步骤 2、根据所收集的用户信息，生成所述用户的特定标签；

步骤 3、对所生成的标签进行量化，计算得到所述用户的属性值；

5 步骤 4、根据计算出的所述用户的属性值，生成用户画像。

14、根据权利要求 13 所述的计算机可读存储介质，其特征在于，所述步骤 3 包括：

步骤 3-1、将所述用户的各个标签量化为多个标准值；

步骤 3-2、针对于所述用户的某个属性，对所述用户的部分标签的量化后的标准值对进行加权求和，生成所述用户的所述属性的属性值。

10 15、根据权利要求 14 所述的计算机可读存储介质，其特征在于，所述标签包括性别标签、学历标签、消费标签、运动标签、职位标签、购物标签、上网查阅标签、兴趣爱好标签、以及活动地方标签；

其中，所述用户画像显示量化值较高的标签以及属性值。

15 16、根据权利要求 15 所述的计算机可读存储介质，其特征在于，所述属性值包括智慧值、财力值、能力值，所述智慧值、所述财力值、所述能力值计算公式为：

智慧值=学历标签×80%+购物标签×15%+其他类的标签×5%；

财力值=消费标签×60%+购物标签×30%+其他类标签×10%；

能力值=职位标签×40%+学历标签×30%+消费标签×20%+其他标签×10%；

体力值=运动标签×50%+购物标签×20%+其他标签×30%；

20 其中，上述公式中的各个标签代表其量化后的标准值。

17、根据权利要求 15 所述的计算机可读存储介质，其特征在于，所述静态信息包括公共数据、以及机构内部的客户信息，所述客户信息包括人口属性、商业属性数据；

其中，人口属性包括：地域、年龄、性别、文化、职业、收入、生活习惯、消费习惯。

25 18、根据权利要求 17 所述的计算机可读存储介质，其特征在于，所述行为信息包括所述用户在网络或应用程序上的行为信息、用户的网站浏览行为信息、用户交易行为信息，其中，所述标签的量化值取决于所述行为信息；

通过以下公式计算用户的行为信息所对应的行为的发生概率：

$$P(A)=m/n,$$

其中，“A”表示行为 A，“m”表示行为 A 发生的总数，“n”是全部行为发生的总数；

30 之后，若 P 大于特定阈值，则确定行为 A 为大概率事件，之后，并将行为 A 关联到所述标签。

19、根据权利要求 13 所述的计算机可读存储介质，其特征在于，所述步骤 1 包括：

步骤 1-1、对静态信息和行为信息进行预处理，根据预处理后的行为信息获取用户在每个预设行为类别中的行为数据，使获取的同一类别的行为数据具有相同的格式。

20、根据权利要求 13 所述的计算机可读存储介质，其特征在于，在所述步骤 4 之后还包括：

- 5 步骤 5、在社交平台上，根据各个用户之间的评价权限，各个用户可给其它用户的属性进行打分，然后，后台系统根据打分情况，来对用户的属性值进行修正。

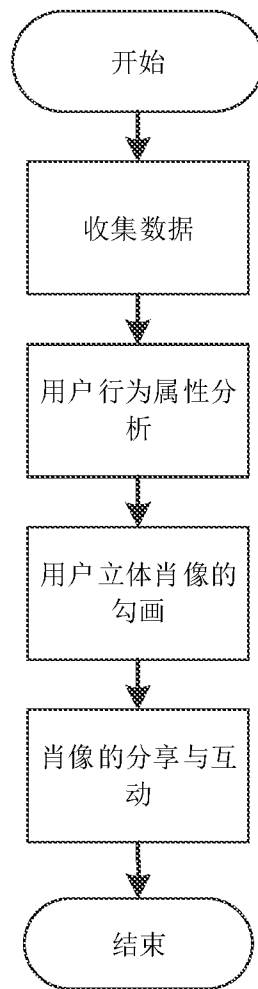


图 1

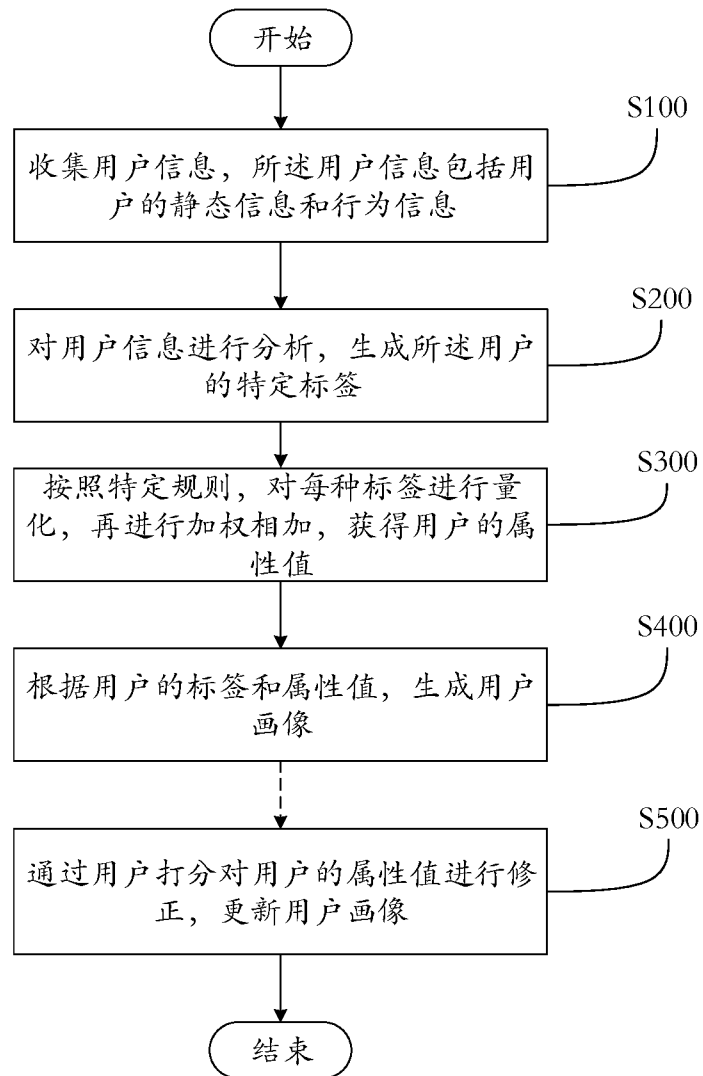


图 2

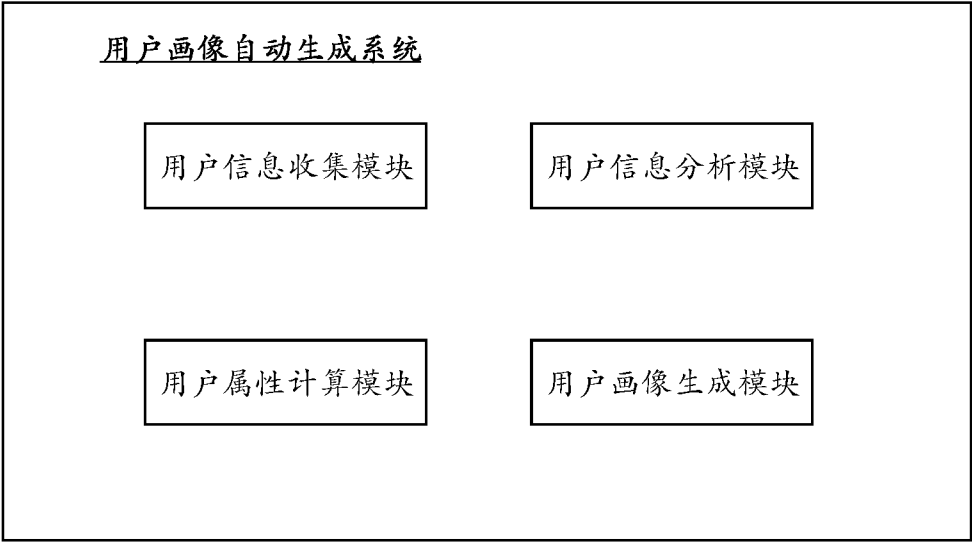


图 3

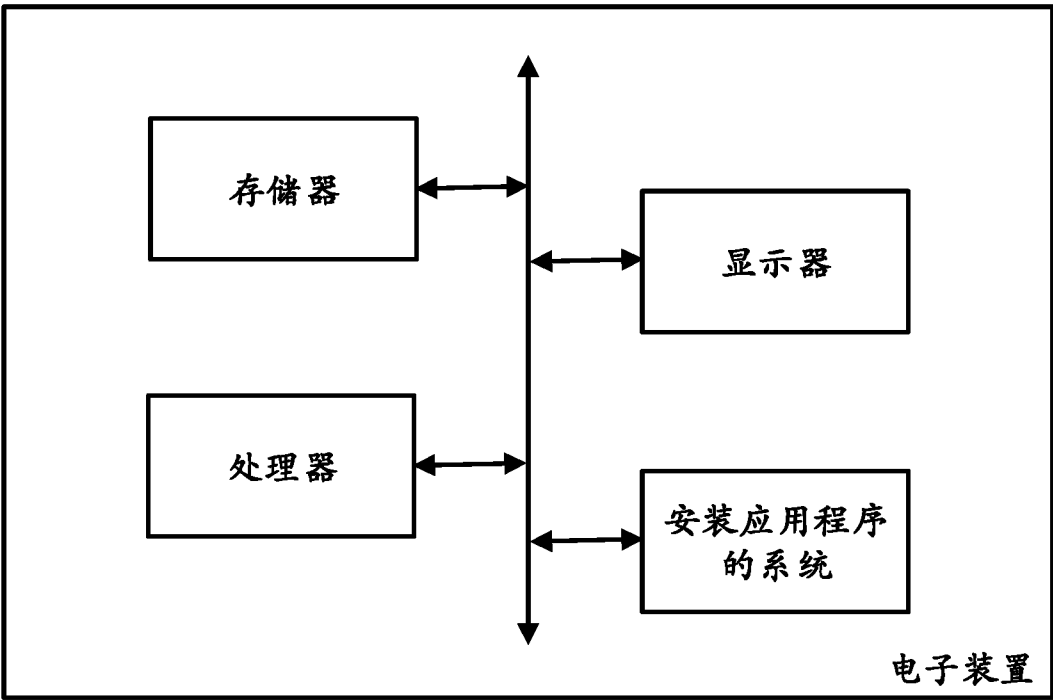


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/124669

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/9535(2019.01)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,G06K,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)

CNKI, CNPAT, WPI, EPODOC: 用户画像, 行为, 静态, 属性, 标签, user, profile, static, behavior, label, attribute

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 105893407 A (LETV CLOUD COMPUTING CO., LTD.) 24 August 2016 (2016-08-24) description, paragraphs [0007]-[0094]	1-20
X	CN 106909686 A (JILIN SHENGCHUANG TECHNOLOGY CO., LTD.) 30 June 2017 (2017-06-30) description, paragraphs [0005]-[0069]	1-20
A	CN 106503015 A (NATIONAL COMPUTER NETWORK AND INFORMATION SECURITY MANAGEMENT CENTER) 15 March 2017 (2017-03-15) entire document	1-20
A	CN 108038237 A (GUANGZHOU URUN BIG DATA SERVICE CO., LTD.) 15 May 2018 (2018-05-15) entire document	1-20
A	CN 107908606 A (ONECONNECT FINANCIAL TECHNOLOGY CO., LTD. (SHANGHAI)) 13 April 2018 (2018-04-13) entire document	1-20
A	CN 108363821 A (SHENZHEN ONECONNECT TECHNOLOGY CO., LTD.) 03 August 2018 (2018-08-03) 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

10 July 2019

Date of mailing of the international search report

26 July 2019

Name and mailing address of the ISA/CN

**National Intellectual Property Administration, PRC (ISA/
CN)**
**No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China**

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/124669

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 107818344 A (ONECONNECT FINANCIAL TECHNOLOGY CO., LTD. (SHANGHAI)) 20 March 2018 (2018-03-20) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/124669

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	105893407	A	24 August 2016	WO	2017080176	A1	18 May 2017
CN	106909686	A	30 June 2017	None			
CN	106503015	A	15 March 2017	None			
CN	108038237	A	15 May 2018	None			
CN	107908606	A	13 April 2018	None			
CN	108363821	A	03 August 2018	None			
CN	107818344	A	20 March 2018	None			

国际检索报告

国际申请号

PCT/CN2018/124669

A. 主题的分类

G06F 16/9535 (2019. 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, G06K, 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))

CNKI, CNPAT, WPI, EPDOC: 用户画像, 行为, 静态, 属性, 标签, user, profile, static, behavior, label, attribute

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 105893407 A (乐视云计算有限公司) 2016年 8月 24日 (2016 - 08 - 24) 说明书第[0007]-[0094]段	1-20
X	CN 106909686 A (吉林省盛创科技有限公司) 2017年 6月 30日 (2017 - 06 - 30) 说明书第[0005]-[0069]段	1-20
A	CN 106503015 A (国家计算机网络与信息安全管理中心) 2017年 3月 15日 (2017 - 03 - 15) 全文	1-20
A	CN 108038237 A (广州市云润大数据服务有限公司) 2018年 5月 15日 (2018 - 05 - 15) 全文	1-20
A	CN 107908606 A (上海壹账通金融科技有限公司) 2018年 4月 13日 (2018 - 04 - 13) 全文	1-20
A	CN 108363821 A (深圳壹账通智能科技有限公司) 2018年 8月 3日 (2018 - 08 - 03) 全文	1-20
A	CN 107818344 A (上海壹账通金融科技有限公司) 2018年 3月 20日 (2018 - 03 - 20) 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

2019年 7月 10日

国际检索报告邮寄日期

2019年 7月 26日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

王加新

电话号码 86-10-53961694

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/124669

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	105893407	A	2016年 8月 24日	WO 2017080176 A1	2017年 5月 18日
CN	106909686	A	2017年 6月 30日	无	
CN	106503015	A	2017年 3月 15日	无	
CN	108038237	A	2018年 5月 15日	无	
CN	107908606	A	2018年 4月 13日	无	
CN	108363821	A	2018年 8月 3日	无	
CN	107818344	A	2018年 3月 20日	无	