

(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
2016 年 12 月 29 日 (29.12.2016)



(10) 国际公布号
WO 2016/206196 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/089823
- (22) 国际申请日: 2015 年 9 月 17 日 (17.09.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510363062.3 2015 年 6 月 26 日 (26.06.2015) CN
- (71) 申请人: 百度在线网络技术(北京)有限公司
(BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。
- (72) 发明人: 吴海山 (WU, Haishan); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。汪天一 (WANG, Tianyi); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。武政伟 (WU, Zhengwei); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。李正学 (LI, Zhengxue); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。张潼 (ZHANG,

Tong); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。

(74) 代理人: 北京英赛嘉华知识产权代理有限公司 (INSIGHT INTELLECTUAL PROPERTY LIMITED); 中国北京市海淀区知春路甲 48 号盈都大厦 A 座 19A, Beijing 100098 (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,

[见续页]

(54) Title: METHOD AND DEVICE FOR OBTAINING USER ATTRIBUTE INFORMATION, AND SERVER

(54) 发明名称: 用于获取用户属性信息的方法、装置及服务器

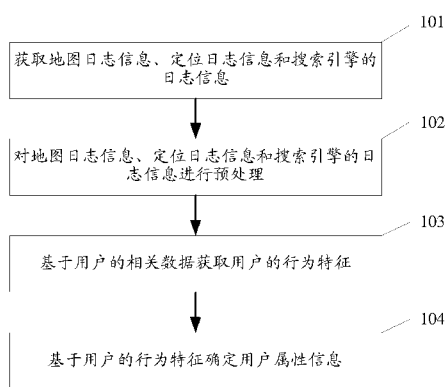


图 1

- 101 Obtain map log information, positioning log information and search engine log information
102 Pre-process the map log information, the positioning log information and the search engine log information
103 Obtain, on the basis of user-related data, user behavioral characteristics
104 Determine, on the basis of the user behavioral characteristics, user attribute information

(57) Abstract: A method and device for obtaining user attribute information, and server. The method comprises: obtaining map log information, positioning log information and search engine log information (101); pre-processing the map log information, the positioning log information and the search engine log information to obtain user-related data (102); obtaining, on the basis of the user-related data, user behavioral characteristics (103); and determining, on the basis of the user behavioral characteristics, user attribute information (104). Information such as user positioning and map search is fully utilized to analyze the user attribute information, thus improving comprehensiveness and accuracy of the obtained user attribute information.

(57) 摘要: 用于获取用户属性信息的方法、装置及服务器。所述方法包括: 获取地图日志信息、定位日志信息和搜索引擎的日志信息(101); 对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理, 以获取用户的相关数据(102); 基于用户的相关数据获取用户的行为特征(103); 以及基于用户的行为特征确定用户属性信息(104)。充分利用了

用户的定位、地图搜索等信息分析用户属性信息, 提升了所获取的用户属性信息的全面性和准确性。



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))。

说明书

用于获取用户属性信息的方法、装置及服务器

相关申请的交叉引用

本申请要求百度在线网络技术（北京）有限公司于 2015 年 6 月 26 日提交的，发明名称为“用于获取用户、属性信息的方法、装置及服务器”的、中国专利申请号“201510363062.3”的优先权，其全部内容作为整体并入本申请中。

技术领域

本申请涉及计算机技术领域，具体涉及终端技术领域，尤其涉及用于获取用户属性信息的方法、装置及服务器。

背景技术

用户画像可以是用户属性信息的集合，可以用一个模型来描述用户的特征。现有技术中，用户画像构建的主要方法是基于用户的线上搜索行为分析用户的属性信息。在这种方法中，由于用户的线上搜索行为中可能存在因恶意用户伪造的搜索信息等虚拟的用户搜索行为而造成的噪音，导致用户画像的构建结果不准确。此外，基于用户的线上搜索行为的画像构建还可能存在文本语义分歧的问题，同一搜索词可能指向不同的用户特征，例如用户搜索“庐山”，可能是关注旅游信息，也可能是喜欢与庐山相关的电影。

另外，现有技术中还有基于用户实际交易数据的用户画像构建方法，基于用户的线上交易数据进行构建。用户的线上交易在用户的行为中为低频行为，因此无法据其得出全面、完整、准确的用户属性信息。

发明内容

鉴于上述现有技术中的缺陷或不足，期望能够提供一种全面、准

确的用户属性信息的获取方法。本申请提供了用于获取用户属性信息的方法、装置及服务器。

第一方面，本申请提供了一种用于获取用户属性信息的方法，包括：获取地图日志信息、定位日志信息和搜索引擎的日志信息；对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，以获取用户的相关数据；基于用户的相关数据获取用户的行为特征；以及基于用户的行为特征确定用户属性信息。

在某些实现方式中，对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，包括：对地图日志信息、定位日志信息和搜索引擎的日志信息所包含的数据进行分析；提取地图日志信息、定位日志信息和搜索引擎的日志信息中与地理位置和用户行为相关的数据，作为用户的相关数据。

在进一步的实现方式中，对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，还包括：通过网络查找出与地图日志信息、定位日志信息和搜索引擎的日志信息包含的数据相关的信息作为用户的相关数据。

在某些实现方式中，用户的相关数据至少包括位置检索数据和/或定位数据。其中，位置检索数据包括以下至少一项：目标位置搜索数据、路线搜索数据和对应的线路信息；以及目标位置的周边数据。

在进一步的实现方式中，目标位置搜索数据包括以下至少一项：搜索的目的地、搜索的时刻、用户当前地理位置；路线搜索数据包括以下至少一项：用户检索路线的时刻、起始地理位置、目标地理位置、轨迹数据以及对应的出行方式；目标位置的周边数据包括以下至少一项：目标位置周边的建筑物数据、交通站点数据、停车场数据。

在某些实现方式中，基于用户的相关数据获取用户的行为特征，包括以下至少一项：基于定位数据对用户停留的地理位置的分布进行统计和分析，以确定用户固定活动的地点；基于位置检索数据获取用户的兴趣点信息；基于位置检索数据对用户的出行方式进行统计和分析，以确定用户偏好的出行方式；基于定位数据计算用户之间的相关度，以确定多个用户的亲密程度。

在某些实现方式中，基于用户的行为特征确定用户属性信息，包括：基于用户的行为特征，采用已训练的模型确定用户属性信息。

在某些实现方式中，用户属性信息包括以下至少一项：用户的年龄段、性别、职业、兴趣、收入水平、消费习惯、健康状况、社会关系以及固定资产状况。

第二方面，本申请提供了一种用于获取用户属性信息的装置，包括：第一获取单元，用于获取地图日志信息、定位日志信息和搜索引擎的日志信息；预处理单元，用于对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，以获取用户的相关数据；第二获取单元，用于基于用户的相关数据获取用户的行为特征；以及确定单元，用于基于用户的行为特征确定用户属性信息。

在某些实现方式中，预处理单元用于按如下方式对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理：对地图日志信息、定位日志信息和搜索引擎的日志信息所包含的数据进行分析；提取地图日志信息、定位日志信息和搜索引擎的日志信息中与地理位置和用户行为相关的数据，作为用户的相关数据。

在进一步的实现方式中，预处理单元还用于按如下方式对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理：通过网络查找出与地图日志信息、定位日志信息和搜索引擎的日志信息包含的数据相关的信息作为用户的相关数据。

在某些实现方式中，用户的相关数据至少包括位置检索数据和/或定位数据。其中，位置检索数据包括以下至少一项：目标位置搜索数据、路线搜索数据和对应的线路信息；以及目标位置的周边数据。

在进一步的实现方式中，目标位置搜索数据包括以下至少一项：搜索的目的地、搜索的时刻、用户当前地理位置；路线搜索数据包括以下至少一项：用户检索路线的时刻、起始地理位置、目标地理位置、轨迹数据以及对应的出行方式；目标位置的周边数据包括以下至少一项：目标位置周边的建筑物数据、交通站点数据、停车场数据。

在进一步的实现方式中，第二获取单元用于按如下至少一种方式获取用户的行为特征：基于定位数据对用户停留的地理位置的分布进

行统计和分析，以确定用户固定活动的地点；基于位置检索数据获取用户的兴趣点信息；基于位置检索数据对用户的出行方式进行统计和分析，以确定用户偏好的出行方式；基于定位数据计算用户之间的相关度，以确定多个用户的亲密程度。

在某些实现方式中，确定单元基于用户的行为特征，采用已训练的模型确定用户属性信息。

在某些实现方式中，用户属性信息包括以下至少一项：用户的年龄段、性别、职业、兴趣、收入水平、消费习惯、健康状况、社会关系以及固定资产状况。

第三方面，本申请提供了一种服务器，包括本申请第二方面所提供的用于获取用户属性信息的装置。

本申请提供的用于获取用户属性信息的方法、装置及服务器，通过获取地图日志信息、定位日志信息和搜索引擎的日志信息，随后对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，以获取用户的相关数据，之后基于用户的相关数据获取用户的行为特征，最后基于用户的行为特征确定用户属性信息，充分利用了用户的定位、地图搜索等信息分析用户属性信息，提升了所获取的用户属性信息的全面性和准确性。

附图说明

通过阅读参照以下附图所作的对非限制性实施例详细描述，本申请的其它特征、目的和优点将会变得更明显：

图 1 是本申请实施例提供的用于获取用户属性信息的方法的一个实施例的流程图；

图 2 是本申请实施例提供的对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理的方法的一个实施例的流程图；

图 3 是本申请实施例提供的用于获取用户属性信息的装置的一个实施例的结构示意图；

图 4 是可以应用本申请实施例的示例性系统架构示意图；

图 5 是适于用来实现本申请实施例的服务器的计算机系统的结构

示意图。

具体实施方式

下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外，为了便于描述，附图中仅示出了与有关发明相关的部分。

需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

本申请所涉及终端可以包括但不限于智能手机、平板电脑、个人数字助理、智能穿戴设备及膝上型便携计算机等等。出于示例描述目的以及为了简洁起见，在接下来的讨论中，结合安装有电子地图、浏览器、具有定位功能的终端来描述本申请的示例性实施例。

请参考图 1，其示出了根据本申请的用于获取用户属性信息方法的一个实施例的流程。该方法可以由服务器来执行。

如图 1 所示，在步骤 101 中，获取地图日志信息、定位日志信息和搜索引擎的日志信息。

一般来说，当用户通过地图查询目标地址、路线信息时，终端可以与地图服务器建立通信连接，以获取用户所查询的地图数据；用户在通过终端的浏览器访问网页或进行检索时，终端可以与网页服务器建立通信连接，以获取用户所要访问的网页信息；用户打开终端的定位功能时，终端可以与定位服务器建立通信连接，以获取当前的定位数据。终端在向用户反馈查询结果、访问结果和定位结果的同时，可以将查询、访问和定位相关的信息保存至终端的存储器中，也可以将查询、访问和定位相关的信息上传至对应的服务器，生成日志信息并保存。具体地，用户通过地图查询的相关信息（例如查询时间、查询内容）可以保存为地图日志信息；终端通过 GPS 等方式定位的相关信息（例如终端的三维地理位置数据以及定位时间）可以保存为定位日志信息；用户通过浏览器上的搜索引擎进行检索的相关信息（例如检

索内容、检索的时间)可以保存为搜索引擎的日志信息。

在本实施例中,如果上述日志信息保存在终端的存储器中,则服务器可以通过网络从终端获取地图日志信息、定位日志信息和搜索引擎的日志信息。如果上述日志信息保存在对应的服务器中,则服务器可以分别从地图服务器获取地图日志信息,从定位服务器获取定位日志信息,从搜索引擎的服务器获取搜索引擎的日志信息。在一些实现方式中,服务器可以从终端和对应的地图服务器、定位服务器以及搜索引擎的服务器获取日志信息,并基于用户的标识信息(如IP地址、用户ID等)识别并提取同一用户的上述日志信息。

上述地图日志信息、定位日志信息和搜索引擎的日志信息中可以包括:时间信息、操作信息、状态信息、网络协议信息、查询、定位或搜索结果以及结果中所包含的内容占用的存储空间等。在一些实现中,上述日志信息以字段形式保存。

在步骤102中,对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理。

在本实施例中,服务器可以对获取的地图日志信息、定位日志信息以及搜索引擎的日志信息进行预处理,以获取用户的相关数据。用户的相关数据可以是与用户行为相关的数据,也可以是与用户的属性特征(例如性别、性格)相关的数据,还可以是与用户的地理位置相关的数据。如前面所述,上述日志信息中除了包含时间信息、操作信息、查询、定位或搜索结果之外,还包括与用户的行为、属性特征、地理位置无关的状态信息、网络协议信息。在本实施例中,服务器可以将日志信息中与用户行为无关的信息过滤,例如当日志信息为字段形式时,可以通过分析字段内容来滤除表示状态信息和网络协议的字段。

进一步参考图2,其示出了本申请实施例提供的对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理的方法的一个实施例的流程图。

如图2所示,在步骤201中,对地图日志信息、定位日志信息和搜索引擎的日志信息所包含的数据进行分析。

在本实施例中，服务器可以对获取的地图日志信息、定位日志信息和搜索引擎的日志信息进行分析，例如可以对字段形式的日志信息进行切分，并分析每一部分字段所代表的内容。举例而言，搜索引擎的一条日志信息中可以包含用户的 IP 地址、搜索时间、搜索内容、状态码、访问的 URL 等。可以分析该日志信息，从而确定用户的搜索时间和搜索内容。

在步骤 202 中，提取地图日志信息、定位日志信息和搜索引擎的日志信息中与地理位置和用户行为相关的数据，作为用户的相关数据。

在本实施例中，服务器可以分别判断步骤 201 中分析所得到的数据是否与用户行为相关以及是否与地理位置相关。当分析所得到的数据与地理位置或用户行为相关时，可以将相关的数据提取出来，作为用户的相关数据。用户的相关数据可以用来分析用户的属性特征，构建用户画像。

例如，在上述实施例中，可以从地图日志信息中提取地图查询时间、查询内容作为用户的相关数据；可以从定位日志信息中提取定位时间和定位的地理位置数据作为用户的相关数据；还可以从搜索引擎的日志信息中提取用户检索的关键词、检索时间、所点击的检索结果中的网页 URL 作为用户的相关数据。

在一些实现中，用户的相关数据至少包括位置检索数据和/或定位数据。其中位置检索数据可以从地图日志信息、搜索引擎的日志信息中提取得出，定位数据可以从定位日志信息中获取。位置检索数据可以包括以下至少一项：目标位置搜索数据、路线搜索数据和对应的线路信息，以及目标位置的周边数据。

具体地，目标位置搜索数据可以包括以下至少一项：搜索的目的地、搜索的时刻、用户当前地理位置。当用户通过终端上的电子地图搜索目标位置时，地图服务器可以根据用户的搜索请求查找到匹配的目的地地理位置，还可以获取用户当前地理位置，记录搜索的时刻，并保存至地图日志信息中。服务器可以从地图日志信息中获取这些数据，作为用户相关的数据。

路线搜索数据可以包括以下至少一项：用户检索路线的时刻、起

始地理位置、目标地理位置、轨迹数据以及对应的出行方式。其中轨迹数据可以为地图服务器查询到的路线轨迹，对应的出行方式可以为驾车、公共交通或步行。在一些实现中，路线搜索数据还可以包括起始地理位置与目标地理位置间的距离。

路线搜索数据的对应的线路信息可以为导航数据。导航数据可以包括详细的定位信息和在对应时刻的移动距离。

目标位置的周边数据可以包括以下至少一项：目标位置周边的建筑物数据、交通站点数据、停车场数据。例如当用户搜索某一旅游景点时，服务器可以获取旅游景点周边的酒店、餐饮场所的数据、该旅游景点附近的地铁站、公交站数量与距离或者该旅游景点周边的停车场数量和车位数量。

在步骤 203 中，通过网络查找出与地图日志信息、定位日志信息和搜索引擎的日志信息包含的数据相关的信息作为用户的相关数据。

除了从日志信息中提取上述与用户行为或地理位置相关的数据之外，服务器还可以通过网络查询与地图日志信息、定位日志信息和搜索引擎的日志信息包含的数据相关的信息作为用户的相关数据。例如当服务器从地图日志信息提取出用户搜索的目的地时，服务器可以查找目的地的类型、属性等信息。举例而言，当目的地为餐馆时，可以通过网络查找到餐馆的平均消费、菜系、车位、评价、其他分店地址等信息，作为用户的相关数据；当目的地为购物场所时，可以查找售卖的商品品牌、类型、价位等信息，作为用户的相关数据。

在上述实施例中，服务器可以分析日志信息、提取其中与用户行为或者地理位置相关的数据，还可以查询与步骤 101 中获取的日志信息包含的数据相关的信息，从而获取用户的相关数据。

返回图 1，在步骤 103 中，基于用户的相关数据获取用户的行为特征。

用户的行为特征可以通过对步骤 102 中获取的用户的相关数据进行分析得出。在本实施例中，服务器可以通过数据统计来推断用户的行为特征。用户的行为特征可以包括但不限于用户固定活动的地点、兴趣点信息、偏好的出行方式、社交关系。当用户的相关数据包括位

置检索数据和定位数据时，可以通过多种方式来分别获取用户的上述行为特征。

在一些实现中，服务器可以基于定位数据对用户停留的地理位置的分布进行统计和分析，以确定用户固定活动的地点。具体地，服务器可以统计停留的时间和频率，根据统计结果得出用户停留时间最长或者停留频率最高的几个地理位置，并结合停留的具体时刻信息推断用户的家庭地点、工作地点及其他规律性活动的地点。例如，可以将所有地理位置中停留时间最长且停留时刻为非工作时间的地理位置作为家庭地点；将停留时间最长且停留时刻为工作时间的地理位置作为用户的工作地点。在实际应用中，还可以将用户停留频率高且稳定的地理位置作为用户规律性活动的地点（例如健身房、用户喜欢的餐馆等）。用户固定活动的地点可以用于分析用户的职业、兴趣、住房价格等信息。

在一些实现中，服务器可以基于位置检索数据获取用户的兴趣点信息。用户的兴趣点信息可以是用户可能感兴趣的信息。具体地，服务器可以获取用户检索的目的地类型等属性信息，根据目的地类型等属性信息获取相关的信息作为用户的兴趣点信息。进一步地，服务器还可以获取用户检索同一属性信息的目的地的频率信息，作为用户的兴趣点信息。例如，当位置检索数据中包含用户检索某机场的位置信息时，服务器可以获取该机场所在地的旅游景点信息、酒店信息等，作为用户的兴趣点信息，以分析用户的兴趣、性格等属性信息。又例如，当位置检索数据中包含用户检索酒店的位置时，服务器可以获取酒店的星级、消费水平等信息，作为用户的兴趣点信息，以分析用户的资产状况。位置检索数据中还可以包含诸如加油站、停车场、洗车地点相关的信息，服务器可以获取这些信息来分析用户是否有车等资产信息。

在一些实现中，服务器可以基于位置检索数据对用户的出行方式进行统计和分析，以确定用户偏好的出行方式。位置检索数据中可以包含用户在检索路线时选择的路线信息及对应的出行方式，服务器可以对用户选择的出行方式进行统计，分析用户常用的出行方式（驾车、

公交、地铁、客车、火车、飞机等),从而推断用户对交通成本的敏感程度,进而分析用户的资产状况。

在一些实现中,服务器可以基于所述定位数据计算用户之间的相关度,以确定多个用户的亲密程度。例如服务器可以根据用户共同停留在临近地理位置等定位数据的交叉特征,分析用户之间社交属性的强弱,即多个用户的亲密程度。服务器还可以基于用户共同停留在临近地理位置的时间信息进一步分析用户的社交关系的类型,例如家人、朋友、工作伙伴等等。举例而言,当多个用户在非工作时间停留在同一地理位置的时间超过一个阈值时,可以认为多个用户的亲密程度较高,可能是家人或朋友关系。

可选地,服务器还可以获取用户所使用的移动终端的类型信息,例如安卓系统或 ios 系统,来推断用户的资产情况。

以上示例性地描述了几种用户行为特征的获取方法。可以理解,步骤 102 中得到的用户的相关数据可以用于分析并获取多种其他类型的用户行为特征,例如可以基于用户的定位数据和位置检索数据分析用户经常停留的场所类型;可以基于用户的位置检索数据对用户的出行规律进行统计等。

在步骤 104 中,基于用户的行为特征确定用户属性信息。

在本实施例中,服务器可以对用户的行为特征进行数据挖掘,基于步骤 103 的统计结果分析用户的属性信息。在一些实施例中,用户属性信息可以包括以下至少一项:用户的年龄段、性别、职业、兴趣、收入水平、消费习惯、健康状况、社会关系以及固定资产状况。

在一些实现中,服务器对用户的行为特征进行数据挖掘后,可以推断出多个备选用户属性信息及其概率,例如得出多个不同的用户的年龄段及每个年龄段对应的概率。之后将概率最大的备选用户属性信息作为用户属性信息。

具体地,用户的年龄段可以根据用户固定活动的地点来判断,例如用户固定活动的地点为大学时,可以推断用户的年龄较大概率为 18-24 岁,用户固定活动的地点为写字楼时,可以推断用户的年龄较大概率为 24-50 岁,用户固定活动的地点为老年中心时,可以推断用户

的年龄较大概率为大于 50 岁。用户的年龄段也可以根据用户的兴趣点信息进行推断。例如，用户通过地图检索游乐场时，可以推断用户的年龄为 15-30 岁的概率较大。

用户的性别可以根据其兴趣点信息来判断，例如搜索美容院等地址的用户较大概率为女性，而搜索篮球场等地址的用户较大概率为男性。

用户的职业可以通过用户的定位数据来判断。例如当用户的行为特征中用户固定活动的地点包括某 IT 企业的大厦时，则可以确定用户从事 IT 行业的概率较大；当用户的行为特征中用户固定活动的地点包括政府机关单位时，则可以确定用户的职业为公务员的概率较大。

用户的兴趣可以通过用户的行为特征中的兴趣点信息来确定。例如当用户的兴趣点信息包含的运动场所数量超过一定值时，可以认为用户喜欢运动。

用户的社交关系可以通过用户的位置检索数据来判断。例如当用户的位置检索数据中包含儿童游乐场、小学、少年宫等目标地点，则可以确定用户有小孩。进一步地，还可以通过有由位置检索数据获取的用户的兴趣点信息来判断用户是否有小孩。例如当用户的兴趣点信息中包含大量的类似“童”、“幼”、“婴”的关键词时，可以确定用户有小孩。

用户的社交关系还可以通过由多个用户的定位数据获取的用户间的亲密程度来确定。例如当多个用户的亲密程度较高时，服务器可以确定多个用户是家人或朋友关系的概率较高。

用户的收入水平、消费水平和固定资产状况可以从用户的多种行为特征来推断。例如，可以基于用户对交通成本的敏感度推断用户的收入水平和固定资产状况，可以基于用户的家庭住址推断住宅价格，从而判断用户的固定资产状况；也可以根据用户检索的消费场所（如酒店、餐馆等）的消费价位来判断用户的收入水平和消费水平；还可以根据用户检索高消费场所（例如高尔夫球场等）的频率来确定用户的收入水平和消费水平。另外还可以根据用户出行时是否经常驾车以及定位数据中是否包含停车场、加油站地址来判断用户是否有车。

用户的健康状况可以根据用户的定位数据中用户在医院和药房等医疗场所停留的频率来推断。例如当用户在医院停留的频率为每周2~3次时，可以推断用户健康状况差的概率较大。

以上示例性地描述了几种用户属性信息的确定方法。需要说明的是，用户属性信息可以不限于以上所描述的几种信息，本申请所提供的方法也不限于用于确定以上描述的几种用户属性信息，还可以用于确定其他类型的用户属性信息，例如可以基于用户经常停留的服装店的统计数据确定用户喜欢的穿着风格；可以基于用户出行频率的统计分析用户的性格（外向或内向）等。

在本实施例中，服务器可以根据预设的规则来计算每一用户属性信息的概率。例如可以预设一个频率阈值，当用户停留在某一地理位置的频率超过该频率阈值时，将用户家庭住址为该地理位置的概率设定为80%。又例如可以预设高档消费场所列表，当用户的定位数据中包含预设的高消费场所列表中的场所时可以将用户为高收入水平的概率设为大于50%。

在一些可选的实现方式中，服务器可以基于用户的行为特征，采用已训练的模型确定用户属性信息。服务器可以将通过已知用户属性的用户的位置检索数据和定位数据获取的用户行为特征作为训练集来训练分类模型。该分类模型可以包括多个子模型，每个子模型用于对用户的一种属性信息进行分类。该分类模型也可以为一个综合模型，用于对用户的多种属性信息进行分类。可选地，还可以基于通过与训练集中不同的用户的位置检索数据和定位数据获取的用户行为特征建立测试集，对分类模型进行优化。在应用时，可以利用已训练并优化的模型对步骤103中获取的用户的行为特征分析，得出用户的属性信息。

上述实施例所提供的用于获取用户属性信息的方法，通过对获取的地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，以获取用户的相关数据，之后基于用户的相关数据获取用户的行为特征，最后基于用户的行为特征确定用户属性信息，充分利用了用户的定位、地图搜索等信息分析用户属性信息，提升了所获取的用户属性

信息的全面性和准确性。

本申请上述实施例所提供的用于获取用户属性信息的方法，可以用于构建用户画像。进一步地，可以基于用户画像向用户推荐内容，例如地图服务器可以根据用户画像分析用户饮食偏好，在用户搜索美食时向用户推荐更符合用户口味的个性化美食。还可以基于用户画像为选址提供依据，例如选址服务器可以基于用户画像确定目标群体的分布、偏好等特征，对选址结果进行优化。

进一步参考图 3，其示出了本申请实施例提供的用于获取用户属性信息的装置的一个实施例的结构示意图。如图 3 所示，用于获取用户属性信息的装置 300 可以包括第一获取单元 301、预处理单元 302、第二获取单元 303 以及确定单元 304。其中，第一获取单元 301 可以用于获取地图日志信息、定位日志信息和搜索引擎的日志信息。预处理单元 302 可以用于对地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，以获取用户的相关数据。第二获取单元 303 可以用于基于用户的相关数据获取用户的行为特征。确定单元 304 可以用于基于用户的行为特征确定用户属性信息。

在本实施例中，第一获取单元 301 可以从终端或地图服务器、定位服务器以及搜索引擎的服务器中获取同一用户的地图日志信息、定位日志信息和搜索引擎的日志信息。预处理单元 302 可以对第一获取单元 301 所获取的日志信息中包含的数据进行分析，提取其中与用户行为或地理位置相关的数据作为用户的相关数据。预处理单元 302 还可以通过网络查找出与地图日志信息、定位日志信息和搜索引擎的日志信息包含的数据相关的信息作为所述用户的相关数据。

在一些实现中，用户的相关数据可以至少包括位置检索数据和/或定位数据。其中，位置检索数据可以包括以下至少一项：目标位置搜索数据、路线搜索数据和对应的线路信息；以及目标位置的周边数据。

进一步地，目标位置搜索数据可以包括以下至少一项：搜索的目的地、搜索的时刻、用户当前地理位置。路线搜索数据可以包括以下至少一项：用户检索路线的时刻、起始地理位置、目标地理位置、轨

迹数据以及对应的出行方式。目标位置的周边数据可以包括以下至少一项：目标位置周边的建筑物数据、交通站点数据、停车场数据。

进一步地，第二获取单元 303 可以用于按如下至少一种方式获取用户的行为特征：基于预处理单元 302 得出的定位数据对用户停留的地理位置的分布进行统计和分析，以确定用户固定活动的地点；基于预处理单元 302 得出的位置检索数据获取用户的兴趣点信息；基于预处理单元 302 得出的位置检索数据对用户的出行方式进行统计和分析，以确定用户偏好的出行方式；基于预处理单元 302 得出的定位数据计算用户之间的相关度，以确定多个用户的亲密程度。

确定单元 304 可以基于第二获取单元 303 所获取的用户的行为特征，采用已训练的模型确定用户属性信息。其中，用户属性信息可以包括以下至少一项：用户的年龄段、性别、职业、兴趣、收入水平、消费习惯、健康状况、社会关系以及固定资产状况。

应当理解，用于获取用户属性信息的装置 300 中记载的诸单元参考图 1-2 描述的方法中的各个步骤相对应。由此，上文针对方法描述的操作和特征同样适用于用于获取用户属性信息的装置 300 及其中包含的单元，在此不再赘述。

进一步参考图 4，其示出了可以应用本申请实施例的示例性系统架构示意图。如图 4 所示，系统 400 可以包括终端 401、402、地图服务器 403、定位服务器 404、搜索引擎的服务器 405 以及用于实现本申请所提供的用于获取用户属性信息的服务器 407。服务器 407 可以包括上述实施例中用于获取用户属性信息的装置 300。在系统架构中还可以包括网络 406。网络 406 用以在终端 401、402、服务器 403、404、405 和 407 之间提供通信链路的介质。网络 406 可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

终端 401、402 可以通过网络 406 与服务器 403、404、405、407 交互，以接收或发送消息等。终端 401、402 具有定位功能，可以安装地图应用和浏览器，可以将地图日志信息通过网络 406 发送给地图服务器 403，将定位日志信息通过网络发送到定位服务器 404，将搜索引擎的日志信息通过网络 406 发送给搜索引擎的服务器 405。服务器 407

可以通过网络 406 从终端 401、402 以及服务器 403、404、405 获取日志信息，对日志信息进行预处理，提取用户行为特征，确定用户属性信息。在实际应用中，服务器 407 还可以将确定的用户属性信息通过网络 406 发送给地图服务器 403 和搜索引擎的服务器 405。地图服务器 403 可以在用户使用地图搜索目标地点时基于用户属性信息向用户推荐相关信息；搜索引擎的服务器 405 可以基于用户属性信息对网页进行重新排序，以使用户更快地搜索到满足需求的信息。

终端 401、402 可以是各种电子设备，包括但不限于个人电脑、智能手机、智能手表、平板电脑、个人数字助理等等。服务器 407 可以对接收到的数据进行存储、分析等处理，并将处理结果反馈给终端和服务器 403、404 以及 405。

应当理解，图 4 中的终端、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端、网络和服务器的。

进一步参考图 5，其示出了是适于用来实现本申请实施例的服务器的计算机系统的结构示意图。如图 5 所示，计算机系统 500 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 RAM 503 中，还存储有系统 500 操作所需的各种程序和数据。CPU 501、ROM 502 以及 RAM 503 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。

以下部件连接至 I/O 接口 505：输入部分 506；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。

附图中的流程图和框图，图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码

的一部分，所述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本申请实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中，例如，可以描述为：一种处理器包括第一获取单元，预处理单元、第二获取单元和确定单元。其中，这些单元的名称在某种情况下并不构成对该单元本身的限定，例如，预处理单元还可以被描述为“用于预处理的单元”。

作为另一方面，本申请还提供了一种计算机可读存储介质，该计算机可读存储介质可以是上述实施例中所述装置中所包含的计算机可读存储介质；也可以是单独存在，未装配入终端中的计算机可读存储介质。所述计算机可读存储介质存储有一个或者一个以上程序，所述程序被一个或者一个以上的处理器用来执行描述于本申请的用于获取用户属性信息的方法。

以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本申请中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离所述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

权 利 要 求 书

1、一种用于获取用户属性信息的方法，其特征在于，所述方法包括：

获取地图日志信息、定位日志信息和搜索引擎的日志信息；

对所述地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，以获取用户的相关数据；

基于所述用户的相关数据获取用户的行为特征；以及

基于所述用户的行为特征确定用户属性信息。

2、根据权利要求1所述的方法，其特征在于，所述对所述地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，包括：

对所述地图日志信息、定位日志信息和搜索引擎的日志信息所包含的数据进行分析；

提取所述地图日志信息、定位日志信息和搜索引擎的日志信息中与地理位置和用户行为相关的数据，作为所述用户的相关数据。

3、根据权利要求2所述的方法，其特征在于，所述对所述地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，还包括：

通过网络查找出与所述地图日志信息、定位日志信息和搜索引擎的日志信息包含的数据相关的信息作为所述用户的相关数据。

4、根据权利要求1-3任意一项所述的方法，其特征在于，所述用户的相关数据至少包括位置检索数据和/或定位数据；

其中，所述位置检索数据包括以下至少一项：目标位置搜索数据、路线搜索数据和对应的线路信息、以及目标位置的周边数据。

5、根据权利要求4所述的方法，其特征在于，

所述目标位置搜索数据包括以下至少一项：搜索的目的地、搜索的时刻、用户当前地理位置；

所述路线搜索数据包括以下至少一项：用户检索路线的时刻、起始地理位置、目标地理位置、轨迹数据以及对应的出行方式；

所述目标位置的周边数据包括以下至少一项：目标位置周边的建筑物数据、交通站点数据、停车场数据。

6、根据权利要求 4 或 5 所述的方法，其特征在于，所述基于所述用户的相关数据获取用户的行为特征，包括以下至少一项：

基于所述定位数据对用户停留的地理位置的分布进行统计和分析，以确定用户固定活动的地点；

基于所述位置检索数据获取用户的兴趣点信息；

基于所述位置检索数据对用户的出行方式进行统计和分析，以确定用户偏好的出行方式；

基于所述定位数据计算用户之间的相关度，以确定多个用户的亲密程度。

7、根据权利要求 1-6 任意一项所述的方法，其特征在于，所述基于所述用户的行为特征确定用户属性信息，包括：

基于所述用户的行为特征，采用已训练的模型确定所述用户属性信息。

8、根据权利要求 1-7 任意一项所述的方法，其特征在于，所述用户属性信息包括以下至少一项：用户的年龄段、性别、职业、兴趣、收入水平、消费习惯、健康状况、社会关系以及固定资产状况。

9、一种用于获取用户属性信息的装置，其特征在于，所述装置包括：

第一获取单元，用于获取地图日志信息、定位日志信息和搜索引擎的日志信息；

预处理单元，用于对所述地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，以获取用户的相关数据；

第二获取单元，用于基于所述用户的相关数据获取用户的行为特征；以及

确定单元，用于基于所述用户的行为特征确定用户属性信息。

10、根据权利要求 9 所述的装置，其特征在于，所述预处理单元用于按如下方式对所述地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理：

对所述地图日志信息、定位日志信息和搜索引擎的日志信息所包含的数据进行分析；

提取所述地图日志信息、定位日志信息和搜索引擎的日志信息中与地理位置和用户行为相关的数据，作为所述用户的相关数据。

11、根据权利要求 10 所述的装置，其特征在于，所述预处理单元还用于按如下方式对所述地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理：

通过网络查找出与所述地图日志信息、定位日志信息和搜索引擎的日志信息包含的数据相关的信息作为所述用户的相关数据。

12、根据权利要求 9-11 任意一项所述的装置，其特征在于，所述用户的相关数据至少包括位置检索数据和/或定位数据；

其中，所述位置检索数据包括以下至少一项：目标位置搜索数据、路线搜索数据和对应的线路信息；以及目标位置的周边数据。

13、根据权利要求 12 所述的装置，其特征在于，

所述目标位置搜索数据包括以下至少一项：搜索的目的地、搜索的时刻、用户当前地理位置；

所述路线搜索数据包括以下至少一项：用户检索路线的时刻、起始地理位置、目标地理位置、轨迹数据以及对应的出行方式；

所述目标位置的周边数据包括以下至少一项：目标位置周边的建筑物数据、交通站点数据、停车场数据。

14、根据权利要求 12 或 13 所述的装置，其特征在于，所述第二获取单元用于按如下至少一种方式获取用户的行为特征：

基于所述定位数据对用户停留的地理位置的分布进行统计和分析，以确定用户固定活动的地点；

基于所述位置检索数据获取用户的兴趣点信息；

基于所述位置检索数据对用户的出行方式进行统计和分析，以确定用户偏好的出行方式；

基于所述定位数据计算用户之间的相关度，以确定多个用户的亲密程度。

15、根据权利要求 9-14 任意一项所述的装置，其特征在于，所述确定单元基于所述用户的行为特征，采用已训练的模型确定所述用户属性信息。

16、根据权利要求 9-15 任意一项所述的装置，其特征在于，所述用户属性信息包括以下至少一项：用户的年龄段、性别、职业、兴趣、收入水平、消费习惯、健康状况、社会关系以及固定资产状况。

17、一种服务器，其特征在于，包括如权利要求 9-16 任一所述的装置。

18、一种设备，其特征在于，包括：

一个或者多个处理器；

存储器；

一个或者多个程序，所述一个或者多个程序存储在所述存储器中，当被所述一个或多个处理器执行时：

获取地图日志信息、定位日志信息和搜索引擎的日志信息；

对所述地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，以获取用户的相关数据；

基于所述用户的相关数据获取用户的行为特征；以及
基于所述用户的行为特征确定用户属性信息。

19、一种非易失性计算机存储介质，所述计算机存储介质存储有一个或多个程序，当所述一个或者多个程序被一个设备执行时，使得所述设备：

获取地图日志信息、定位日志信息和搜索引擎的日志信息；

对所述地图日志信息、定位日志信息和搜索引擎的日志信息进行预处理，以获取用户的相关数据；

基于所述用户的相关数据获取用户的行为特征；以及
基于所述用户的行为特征确定用户属性信息。

说明书附图

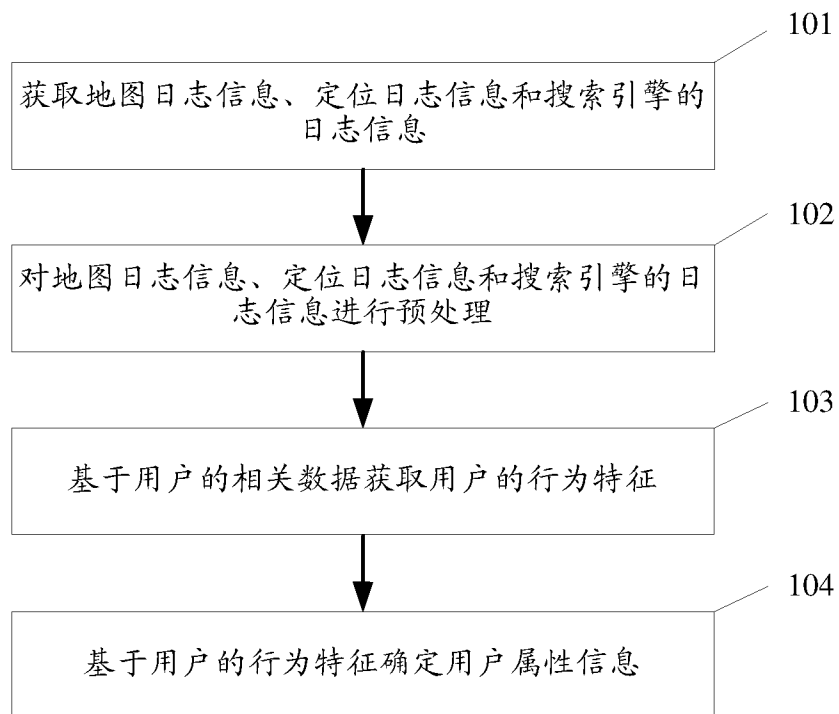


图 1

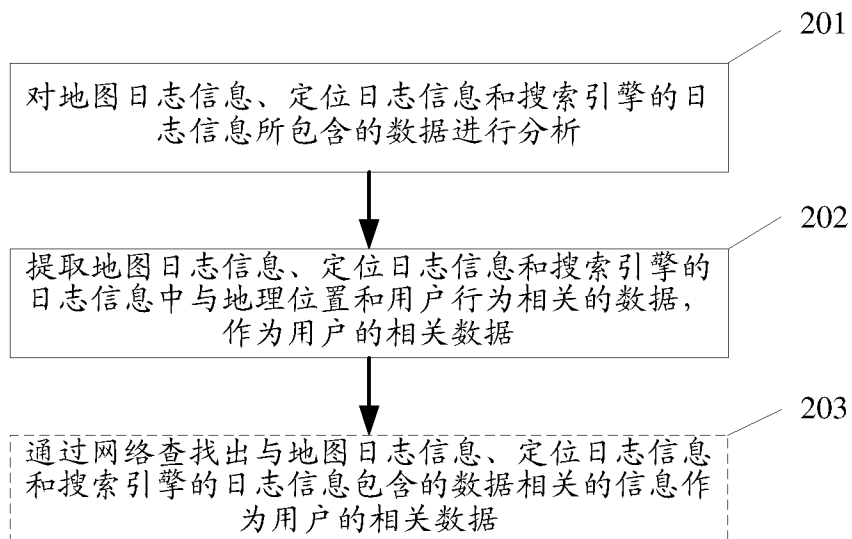


图 2



图 3

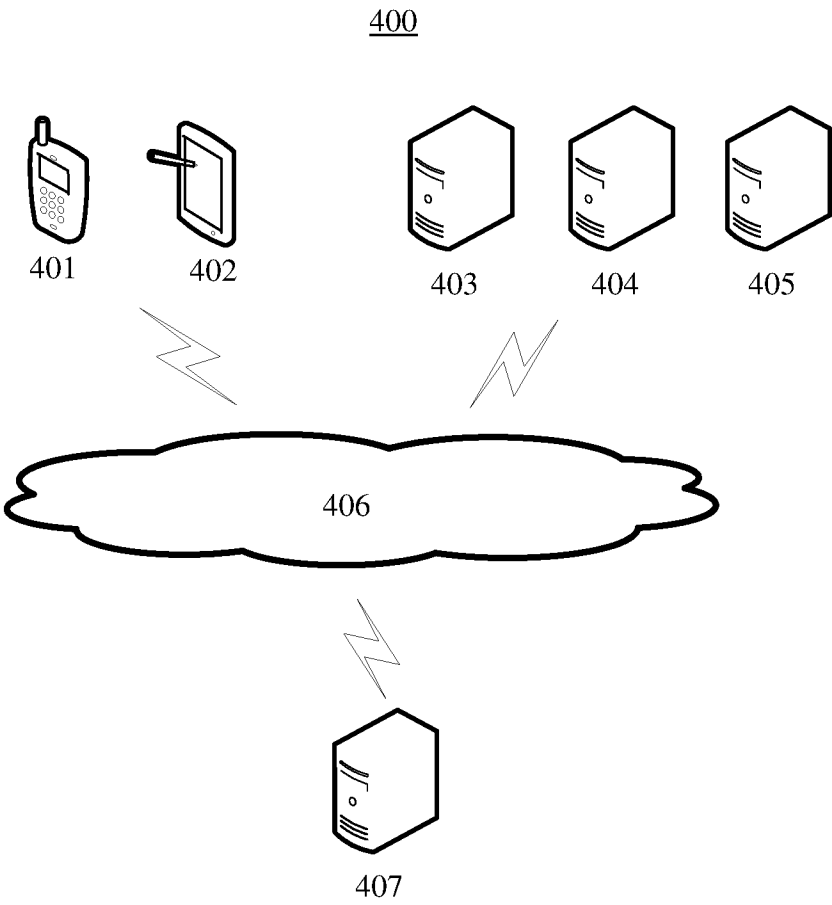


图 4

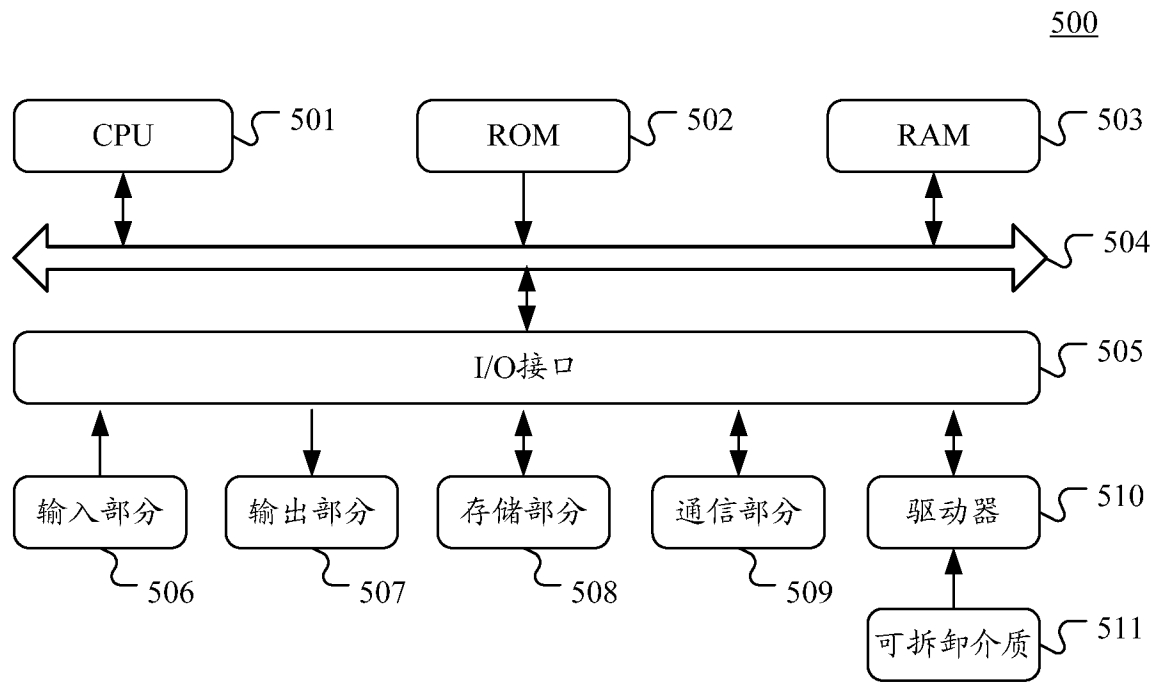


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/089823

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; H04W

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: preference, locate, map, point of interest, POI, user, location, position, search, log, behavior

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 104239526 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 24 December 2014 (24.12.2014), description, paragraphs [0028]-[0033], and figure 2	1-19
Y	CN 104391853 A (SHENZHEN UNIVERSITY et al.), 04 March 2015 (04.03.2015), description, paragraphs [0143]-[0164], and figure 2	1-19
A	US 2010131335 A1 (ROH, D.H. et al.), 27 May 2010 (27.05.2010), the whole document	1-19

☐ 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 13 January 2016 (13.01.2016)	Date of mailing of the international search report 26 January 2016 (26.01.2016)
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 SU, Wen Telephone No.: (86-10) 010-62413821

International application No.
PCT/CN2015/089823

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2015/089823

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; H04W

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, IEEE:地图, 兴趣点, 喜好, 用户, 定位, 位置, 搜索, 日志, 行为, map, point of interest, POI, user, location, position, search, log, behavior

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
Y	CN 104239526 A (百度在线网络技术北京有限公司) 2014年 12月 24日 (2014 - 12 - 24) 说明书第[0028]-[0033]段、图2	1-19
Y	CN 104391853 A (深圳大学等) 2015年 3月 4日 (2015 - 03 - 04) 说明书第[0143]-[0164]段、图2	1-19
A	US 2010131335 A1 (ROH, DONG-HYUN等) 2010年 5月 27日 (2010 - 05 - 27) 全文	1-19

☐ 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 1月 13日

国际检索报告邮寄日期

2016年 1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

苏文

传真号 (86-10)62019451

电话号码 (86-10)010-62413821

PCT/CN2015/089823

表 PCT/ISA/210 (同族专利附件) (2009年7月)