

(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 3 月 19 日 (19.03.2020)



(10) 国际公布号
WO 2020/052338 A1

- (51) 国际专利分类号:
G06K 9/62 (2006.01) **G06F 16/29** (2019.01)
- (21) 国际申请号: PCT/CN2019/095106
- (22) 国际申请日: 2019 年 7 月 8 日 (08.07.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201811064084.X 2018年9月12日 (12.09.2018) CN
- (71) 申请人: 北京三快在线科技有限公司 (**BEIJING SANKUAI ONLINE TECHNOLOGY CO., LTD**) [CN/CN]; 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。
- (72) 发明人: 朱静雅 (**ZHU, Jingya**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。朱青祥 (**ZHU, Qingxiang**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。李诚 (**LI, Cheng**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。
- (74) 代理人: 北京润泽恒知识产权代理有限公司 (**BEIJING RUN ZEHENG INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区中关村南大街 31 号神舟大厦 702, Beijing 100081 (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:** ADDRESS IDENTIFIER AND LONGITUDE AND LATITUDE THEREOF MINING

(54) 发明名称: 地址标识及其经纬度挖掘

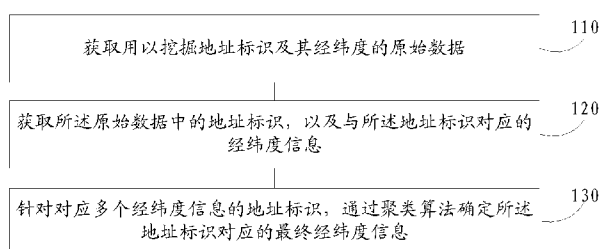


图 1

- 110 Obtain original data for mining an address identifier and the longitude and latitude thereof
- 120 Obtain the address identifier in the original data, and longitude and latitude information corresponding to the address identifier
- 130 For the address identifier corresponding to multiple pieces of longitude and latitude information, determine the final longitude and latitude information corresponding to the address identifier by means of a clustering algorithm

(57) **Abstract:** A method for mining an address identifier and the longitude and latitude thereof, relating to the technical field of navigation and positioning. The method comprises: obtaining original data for mining an address identifier and the longitude and latitude thereof (110); obtaining the address identifier in the original data, and longitude and latitude information corresponding to the address identifier (120); and for the address identifier corresponding to multiple pieces of longitude and latitude information, determining the final longitude and latitude information corresponding to the address identifier by means of a clustering algorithm (130).

(57) **摘要:** 一种地址标识及其经纬度的挖掘方法, 涉及导航定位技术领域。所述方法包括: 获取用以挖掘地址标识及其经纬度的原始数据 (110); 获取所述原始数据中的地址标识, 以及与所述地址标识对应的经纬度信息 (120); 针对对应多个经纬度信息的地址标识, 通过聚类算法确定所述地址标识对应的最终经纬度信息 (130)。

WO 2020/052338 A1

地址标识及其经纬度挖掘

本申请要求在 2018 年 9 月 12 日提交中国专利局、申请号为 201811064084.X、发明名称为“一种地址标识及其经纬度的挖掘方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5

技术领域

本申请涉及导航定位技术，具体涉及地址标识及其经纬度的挖掘。

背景技术

10 随着移动互联网技术的发展和智能手机应用的普及，人们对于导航定位的速度、精度和场景适应性的要求也越来越高，而且大多公司业务中存在大量的定位需求。在实际应用中，一般可以采用道路名称、门牌号等等地址标识对实体场所进行标号区分，而通过获取地址标识的位置信息，则可以辅助定位能力提升到地址标识级别。在现有的技术中，在线
15 地图或者电子地图等商家一般采用人力实采的方式获取地址标识及其经纬度，定期地花费人力去各个街道录入。

发明内容

依据本申请的第一方面，提供了一种地址标识及其经纬度的挖掘方
20 法，包括：

获取用以挖掘地址标识及其经纬度的原始数据；所述原始数据包括兴趣点数据和/或用户原创内容行为数据；

获取所述原始数据中的地址标识，以及与所述地址标识对应的经纬度信息；

25 针对对应多个经纬度信息的地址标识，通过聚类算法确定所述地址标识对应的最终经纬度信息。

根据本申请的第二方面，提供了一种地址标识及其经纬度的挖掘装置，包括：

原始数据获取模块，用于获取用以挖掘地址标识及其经纬度的原始数
30 据；所述原始数据包括兴趣点数据和/或用户原创内容行为数据；

数据挖掘模块，用于获取所述原始数据中的地址标识，以及与所述地址标识对应的经纬度信息；

最终经纬度信息确认模块，用于针对对应多个经纬度信息的地址标识，通过聚类算法确定所述地址标识对应的最终经纬度信息。

根据本申请的第三方面，提供了一种电子设备，包括：处理器、存储器以及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现前述的地址标识及其经纬度的挖掘方法。

根据本申请的第四方面，提供了一种可读存储介质，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行如前述的地址标识及其经纬度的挖掘方法。

10 根据本申请实施例公开的一种地址标识及其经纬度的挖掘方法，可以获取用以挖掘地址标识及其经纬度的原始数据；获取所述原始数据中的地址标识，以及与所述地址标识对应的经纬度信息；针对对应多个经纬度信息的地址标识，通过聚类算法确定所述地址标识对应的最终经纬度信息。由此取得了降低地址标识及其经纬度的挖掘人力成本，同时提高
15 挖掘得到的地址标识及其最终经纬度信息的时效性以及准确性的有益效果。

上述说明仅是本申请技术方案的概述，为了能够更清楚了解本申请的技术手段，而可依照说明书的内容予以实施，并且为了让本申请的上述和其它目的、特征和优点能够更明显易懂，以下特举本申请的具体实施方式。
20 方式。

附图说明

通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本申请的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：
25

图 1 示出了根据本申请一个实施例的一种地址标识及其经纬度的挖掘方法的步骤流程图；

图 2 示出了根据本申请一个实施例的一种地址标识及其经纬度的挖掘方法的步骤流程图；
30

图 3 示出了根据本申请一个实施例的一种密度聚类示意图；

图 4 示出了根据本申请一个实施例的一种地址标识及其经纬度的挖掘

装置的结构示意图；以及

图 5 示出了根据本申请一个实施例的一种地址标识及其经纬度的挖掘装置的结构示意图；

图 6 示意性地示出了用于执行根据本申请的方法的电子设备的框图；

5 以及

图 7 示意性地示出了用于保持或者携带实现根据本申请的方法的程序代码的存储单元。

具体实施例

10 下面将参照附图更详细地描述本申请的示例性实施例。虽然附图中显示了本申请的示例性实施例，然而应当理解，可以以各种形式实现本申请而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本申请，并且能够将本申请的范围完整的传达给本领域的技术人员。

15 实施例一

详细介绍本申请实施例提供的一种地址标识及其经纬度的挖掘方法。

参照图 1，示出了本申请实施例中的一种地址标识及其经纬度的挖掘方法的步骤流程图。

步骤 110，获取用以挖掘地址标识及其经纬度的原始数据。

20 在实际应用中，外卖平台、在线评价网站、在线购物平台等等用户或商家可以进行输入信息、评价等的平台可以记录用户的兴趣点(POI, Point of Interest)数据和/或用户原创内容(UGC, User Generated Content)行为数据。其中，POI 数据也可以叫做"Point of Information"，即"信息点"。POI 数据可以包括但不限于名称、地址、经度、纬度、类别等。在电子地
25 图上一般用气泡图标来表示 POI，像电子地图上的景点、政府机构、公司、商场、饭馆等，都是 POI。UGC 的概念最早起源于互联网领域，即用户将自己原创的内容通过互联网平台进行展示或者提供给其他用户。UGC 并不是某一种具体的业务，而是一种用户使用互联网的新方式，即由原来的以下载为主变成下载和上传并重。随着互联网运用的发展，网络用
30 户的交互作用得以体现，用户既是网络内容的浏览者，也是网络内容的创造者。在本申请实施例中的 UGC 行为数据可以包括但不限于用户原创内容对应的任何数据，例如更改的状态、发表的日志、发布的照片、发布的评价、报错的行为数据、新增的行为数据，等等。而且报错的行为

数据具体可以包括但不限于针对地址，和/或针对某一地址所对应的经纬度信息的报错，而且新增的行为数据也可以包括但不限于新增地址，和/或新增针对某一地址的经纬度信息，等等。

因此，在本申请实施例中，可以通过任何可用方式获取用以挖掘地址标识及其经纬度的原始数据，而且原始数据具体包含的内容也可以根据需求进行预先设置，对此本申请实施例不加以限定。

可选地，在本申请实施例中，原始数据可以包括但不限于兴趣点数据和/或用户原创内容行为数据。

例如，以美团点评平台为例，美团点评平台包含有大量的用户，因此具有大量 POI 数据包含的地址和经纬度，及大量用户新增和报错等行为中带来的地址和经纬度。因此则可以从美团点评平台中大量用户的 POI 数据和 UGC 行为数据中挖掘出地址与经纬度。那么此时则可以该平台中的 POI 数据和/或 UGC 行为数据作为原始数据。当然，在本申请实施例中，可以根据需求设置原始数据的具体来源以及获取渠道，对此本申请实施例不加以限定。

步骤 120，获取所述原始数据中的地址标识，以及与所述地址标识对应的经纬度信息。

如前述，获取的原始数据中可以包含地址、经度、纬度等数据，而在地址中一般可以具体到省、市、区/县、乡镇、村、街道以及门牌号，等等。因此，在本申请实施例中，在获取得到原始数据之后，则可以进一步获取所述原始数据中的地址标识，以及与所述地址标识对应的经纬度信息。具体的，可以通过任何可用方法获取原始数据中的地址标识，以及与所述地址标识对应的经纬度信息，对此本申请实施例不加以限定。

其中的地址标识可以包括道路名称以及门牌号，当然还可以包括国家名称、省份名称、市级名称、区级名称、县级名称、乡镇级名称、村级名称中的一个或多个，具体的，地址标识包括的信息可以根据需求进行预先设置，对此本申请实施例不加以限定。经纬度信息可以包括相应地址标识对应的经度和/或纬度。

步骤 130，针对对应多个经纬度信息的地址标识，通过聚类算法确定所述地址标识对应的最终经纬度信息。

在实际应用中，不同的用户或者是同一用户在不同时刻都可以对同样的地址标识设置经纬度信息。因此同一地址标识可能对应多个经纬度信息，而且多个经纬度信息可以是不完全相同，那么则无法准确确定该地

址标识对应的准确的经纬度信息。因此，在本申请实施例中，对于对应多个经纬度信息的地址标识，则可以通过聚类算法确定所述地址标识对应的最终经纬度信息。在本申请实施例中，可以通过任何一种聚类算法确定各地址标识对应的最终经纬度信息，具体的，可以根据需求预先设置聚类方法，对此本申请实施例不加以限定。例如，可以采用 K-Means(K 均值)聚类、均值漂移聚类、基于密度的聚类、用高斯混合模型的最大期望聚类、凝聚层次聚类以及图团体检测（Graph Community Detection）聚类，等等中的任意一种。

当然，在本申请实施例中，对于对应一个经纬度信息的地址标识，则可以直接以该经纬度信息作为相应地址标识的最终经纬度信息。

在本申请实施例中，通过获取用以挖掘地址标识及其经纬度的原始数据；获取所述原始数据中的地址标识，以及与所述地址标识对应的经纬度信息；针对对应多个经纬度信息的地址标识，通过聚类算法确定所述地址标识对应的最终经纬度信息。从而降低了获取地址标识及其经纬度的人力成本，同时提高了时效性。

实施例二

详细介绍本申请实施例提供的一种地址标识及其经纬度的挖掘方法。

参照图 2，示出了本申请实施例中的一种地址标识及其经纬度的挖掘方法的步骤流程图。

步骤 210，获取用以挖掘地址标识及其经纬度的原始数据。

步骤 220，获取所述原始数据中的地址经纬度数据；所述地址经纬度数据包括地址数据，经纬度数据，以及地址与经纬度的对应关系。

如前述，在实际应用中，原始数据中一般可以包括地址，经纬度等数据，而且经纬度数据一般是某一地点的经纬度数据，同时还可以基于地址描述各地点，因此同一地点的地址数据以及经纬度数据之间存在对应关系。因此，在本申请实施例中，可以获取所述原始数据中的地址经纬度数据。其中，地址经纬度数据包括地址数据，经纬度数据，以及地址与经纬度的对应关系。

步骤 230，基于地理知识库，对所述地址数据进行结构化处理，并取结构化处理后得到的道路名称以及门牌号作为与所述地址数据对应的地址标识；所述地理知识库中包括地理信息实体库，以及各所述地理信息实体之间的关系。

其中地理信息实体库中可以包括不同的国家名称、省份名称、城市名称、

县区名称、乡镇名称、道路名称，等等。例如上海市、北京市、成都市等的城市，长宁区、华阴县等的区县，安化路，长安街，朱雀大街等的道路名称。各地理信息实体之间的关系可以包括不同等级的地理信息实体之间的包含关系，或者是从属关系，等等。诸如“上海市”包含“长宁区”，“长宁区”包含“安化路”，“长宁区”从属于“上海市”等。

在本申请实施例中，则可以基于地理知识库，对获取得到的地址数据进行结构化处理，得到与所述地理知识库中的地理信息实体匹配的结构化的地址数据。例如，对于地址数据“长宁区安化路 492 号德必易园”，经结构化处理后得到：“长宁区”、“安化路”、“492 号”、“德必易园”，且分别对应的类型为“区县”、“道路”、“门牌号”、“地标”的地址数据。

对于结构化处理后的地址数据，则可以准确获取地址数据中的道路名称以及门牌号，那么则可以取结构化处理后地址数据中的道路名称以及门牌号作为相应地址数据对应的地址标识。例如，对于上述的地址数据“长宁区安化路 492 号德必易园”，取结构化处理后得到的地址数据中的道路名称以及门牌号作为相应地址数据对应的地址标识，则可以得到其地址标识为“安化路 492 号”。

步骤 240，基于所述地址与经纬度的对应关系，将与所述地址数据对应的经纬度数据作为与所述地址标识对应的经纬度信息。

在确定了地址数据对应的地址标识之后，则可以进一步基于地址与经纬度的对应关系，将与相应的地址数据对应的经纬度数据作为与相应的地址标识对应的经纬度信息。

例如，假设基于地址与经纬度的对应关系，上述的地址数据“长宁区安化路 492 号德必易园”对应的经纬度数据包括（经度 1，纬度 1）、（经度 2，纬度 2），那么则可以得到该地址数据的地址标识“安化路 492 号”所对应的经纬度信息为（经度 1，纬度 1）、（经度 2，纬度 2）。

步骤 250，基于对应多个经纬度信息的地址标识，则根据第一距离阈值以及第一样本阈值，采用密度聚类算法对所述多个经纬度信息进行聚类，得到至少一个聚类簇。

在实际应用中，由于层次聚类算法和划分式聚类算法往往只能发现凸形的聚类簇。为了弥补这一缺陷，以发现任意形状的聚类簇，因此开发出基于密度的聚类算法（即密度聚类算法）。这类算法认为，在整个样本空间点中，各聚类簇是由一群稠密样本点组成的，而这些稠密样本点被低密度区域（噪

声)分割,而算法的目的就是要过滤低密度区域,发现稠密样本点。那么,在本申请实施例中,由于可以通过经纬度坐标点表征各经纬度信息,而且通过聚类簇中包含的经纬度坐标点的多少即可以确定地址标识所对应的各个经纬度信息的准确度。因此,在本申请实施例中,考虑到密度聚类算法的准确性以及可操作性更高,优选地可以针对对应多个经纬度信息的地址标识,通过密度聚类算法确定所述地址标识对应的最终经纬度信息。

其中的密度聚类算法可以为 DBSCAN (Densit-based Spatial Clustering of Application with Noise, 具有噪声的基于密度的聚类方法), OPTICS (Ordering Pointing To Identify the Cluster Structure), DENCLUE 等任意一种密度聚类算法。

例如,可以通过密度聚类算法对同一地址标识对应的多个经纬度信息进行聚类,得到至少一个聚类簇,然后可以从各个聚类簇中选择出最大簇,进而可以将最大簇中包含的经纬度信息的平均值作为相应地址标识的最终经纬度信息,所述最终经纬度信息为更加准确的经纬度信息。

基于密度的聚类算法,也即密度聚类算法一般假定类别可以通过样本分布的紧密程度决定。同一类别的样本,他们之间的紧密相连的,也就是说,在该类别任意样本周围不远处一定有同类别的样本存在。通过将紧密相连的样本划为一类,这样就得到了一个聚类类别。通过将所有各组紧密相连的样本划为各个不同的类别,则我们就得到了最终的所有聚类类别结果。

以 DBSCAN 密度聚类算法为例, DBSCAN 密度聚类算法是基于一组邻域来描述样本集的紧密程度的,参数(ϵ , MinPts)用来描述邻域的样本分布紧密程度。其中, ϵ 描述了某一样本的邻域距离阈值,也即第一距离阈值,MinPts 描述了某一样本的距离为 ϵ 的邻域中样本个数的阈值,也即第一样本阈值。

假设样本集是 $D=(x_1, x_2, \dots, x_m)$, 则 DBSCAN 具体的密度描述定义如下:

1) ϵ 邻域: 对于 $x_j \in D$, 其 ϵ 邻域包含样本集 D 中与 x_j 的距离不大于 ϵ 的子样本集, 即 $N_\epsilon(x_j)=\{x_i \in D | \text{distance}(x_i, x_j) \leq \epsilon\}$, 这个子样本集的个数记为 $|N_\epsilon(x_j)|$ 。

2) 核心对象: 对于任一样本 $x_j \in D$, 如果其 ϵ 邻域对应的 $N_\epsilon(x_j)$ 至少包含 MinPts 个样本, 即如果 $|N_\epsilon(x_j)| \geq \text{MinPts}$, 则 x_j 是核心对象。

3) 密度直达: 如果 x_i 位于 x_j 的 ϵ 邻域中, 且 x_j 是核心对象, 则称 x_i 由 x_j 密度直达。反之不一定成立, 即此时不能说 x_j 由 x_i 密度直达, 除非 x_i 也是

核心对象。

4) 密度可达: 对于 x_i 和 x_j , 如果存在样本序列 $p_1, p_2, \dots, p_T$, 满足 $p_1 = x_i$, $p_T = x_j$, 且 p_{t+1} 由 p_t 密度直达, 则称 x_j 由 x_i 密度可达。也就是说, 密度可达满足传递性。此时序列中的传递样本 $p_1, p_2, \dots, p_{T-1}$ 均为核心对象, 因为只有核心对象才能使其他样本密度直达。密度可达也不满足对称性, 这个可以由密度直达的不对称性得出。

5) 密度相连: 对于 x_i 和 x_j , 如果存在核心对象样本 x_k , 使 x_i 和 x_j 均由 x_k 密度可达, 则称 x_i 和 x_j 密度相连。密度相连关系是满足对称性的。

如图 3 可以很容易理解上述定义, 假设图中的 $\text{MinPts}=5$, 其中每个带有箭头的线段起点所在的样本都是核心对象, 因为其 ϵ 邻域至少有 5 个样本。其他的样本则是非核心对象。所有核心对象密度直达的样本在以相应核心对象为中心的超球体内, 如果不在超球体内, 则不能密度直达。图中用带有箭头的线段连起来的核心对象组成了密度可达的样本序列。在这些密度可达的样本序列的 ϵ 邻域内所有的样本相互都是密度相连的。

15 如果基于对应多个经纬度信息的地址标识, 则根据第一距离阈值以及第一样本阈值, 采用 DBSCAN 密度聚类算法对所述多个经纬度信息进行聚类, 得到至少一个聚类簇。此时上述的样本集 $D=(x_1, x_2, \dots, x_m)$ 中包含的样本则为一个地址标识对应的多个经纬度信息。其中, 第一距离阈值以及第一样本阈值的具体取值可以根据需求进行预先设置, 对此本申请实施例不加以限定。

20 对于每个地址标识所关联的经纬度信息, 其在地图上均可显示, 并可计算各个经纬度信息之间的距离。如图 3 所示, 一个地址标识对应多个经纬度信息, 图 3 中的一个样本点对应一个经纬度信息, 经 DBSCAN 密度聚类后, 一个地址标识对应的多个经纬度点可能聚成至少一个聚类簇。如图 3 则得到两个聚类簇, 分别为左右两侧两个连续的有向线段所连接的核心对象对应的

25 超球体构成的聚类簇。

步骤 260, 从所述至少一个聚类簇中选定最大簇。

很明显, 最大簇所包含的经纬度信息对于相应的地址标识更为准确, 因此在本申请的一些实施例中, 可以从经密度聚类得到的至少一个聚类簇中选定最大簇, 以确定相应地址标识的最终经纬度信息。具体的, 可以采用任何

30 可用方法选定出最大簇, 对此本申请实施例不加以限定。

例如, 可以包含的核心对象的数量最多的聚类簇作为最大簇, 或者是以包含的样本点数量最多, 也即包含的经纬度信息最多的聚类簇作为最大簇,

等等。

可选地，在本申请实施例中，所述步骤 260 进一步可以包括：

子步骤 261，以所述聚类簇中包含的经纬度信息数量最多的一个聚类簇作为所述最大簇。

- 5 优选地，在本申请实施例中，可以一地址标识对应的至少一个聚类簇中包含的经纬度信息数量最多的一个聚类簇作为相应地址标识对应的最大簇。

例如，对于图 3 所示的一个地址标识的经纬度信息的聚类效果图。其中的每个样本点对应一个经纬度信息，可以看出其中左侧的聚类簇中包含的经纬度信息数量最多，那么则可以该聚类簇作为相应地址标识对应的最大簇。

- 10 步骤 270，对所述最大簇中的经纬度信息取平均，得到平均后的经纬度数值作为所述地址标识对应的最终经纬度信息。

- 在确认了地址标识对应的最大簇之后，则可以进一步对相应的最大簇中的经纬度信息取平均，得到平均后的经纬度数值作为相应地址标识对应的最终经纬度信息。如前述，聚类簇中的每个样本点对应一个样本，而本申请实
15 施例中的样本即为经纬度信息，因此聚类簇中的一个样本点对应一个经纬度信息。那么此时可以将最大簇中的每个样本点对应的经纬度信息取平均，得到平均后的经纬度数值作为所述地址标识对应的最终经纬度信息。具体的，可以将最大簇中的各个样本点对应的经纬度信息中经度值取平均，同时将各个样本点对应的经纬度信息中纬度值取平均，进而以平均后的经度值以及纬
20 度值，构建得到相应地址标识对应的最终经纬度信息。

步骤 280，获取所述地址标识对应的最大簇中包含的经纬度信息的第一数量，以及所述地址标识对应的最终经纬度信息与所述地址标识对应的道路之间的第一距离。

- 在实际应用中，经上述步骤可以批量得到的大量地址标识的最终经纬度
25 信息。但是各个最终经纬度信息与相应地址标识的匹配程度并不能确定。而且在定位服务等需要用到地址标识以及相应的最终经纬度信息时，如果定位服务的精度要求较高，而返回的地址标识以及相应的最终经纬度信息可能无法达到相应的精度要求。因此，在本申请实施例中，为了避免上述问题，可以进一步确定各个地址标识以及相应的最终经纬度信息的置信度。具体的，
30 可以基于地址标识对应的最大簇中包含的经纬度信息的第一数量，以及相应地址标识对应的最终经纬度信息与相应地址标识对应的道路之间的第一距离，确定相应地址标识及其最终经纬度信息的置信度。

那么首先需要获取所述地址标识对应的最大簇中包含的经纬度信息的第一数量，以及所述地址标识对应的最终经纬度信息与所述地址标识对应的道路之间的第一距离。

其中，在本申请实施例中可以通过任何可用方法获取地址标识对应的最大簇中包含的经纬度信息的第一数量，以及地址标识对应的最终经纬度信息与相应的地址标识对应的道路之间的第一距离。例如，最大簇中每个样本点对应一个经纬度信息，因此可以统计最大簇中样本点的数量，进而得到最大簇中包含的经纬度信息的第一数量。在电子地图中获取地址标识对应的最终经纬度信息所对应的坐标点到相应的地址标识对应的道路之间的垂直距离作为第一距离；等等。

步骤 290，根据所述第一数量以及所述第一距离，确定所述地址标识与所述最终经纬度信息的置信度。

在获取得到第一数量以及第一距离之后，则可以根据第一数量以及第一距离，确定相应的地址标识与其最终经纬度信息的置信度。其中，置信度可以用于表征相应的地址标识与其最终经纬度信息的准确度。而且，第一数量以及第一距离与置信度之间的对应关系可以根据需求进行预先设置，对此本申请实施例不加以限定。例如，可以设置第一数量与置信度成正比关系，且第一距离与置信度成反比关系，此时置信度的取值越大，则说明相应地址标识及其最终经纬度信息的准确度越高；当然也可以设置第一数量与置信度成反比关系，且第一距离与置信度成正比关系，此时置信度的取值越小，则说明相应地址标识及其最终经纬度信息的准确度越高；等等。

可选地，在本申请实施例中，所述步骤 290 进一步可以包括：

子步骤 291，根据所述第一距离以及预设基础分值和预设距离阈值，确定所述地址标识与所述最终经纬度信息的初始分值。

其中的预设基础分值以及预设距离阈值都可以根据需求进行预先设置，对此本申请实施例不加以限定。而且，第一距离以及预设基础分值和预设距离阈值与初始分值的映射关系也可以根据需求或经验等进行预先设置，对此本申请实施例也不加以限定。

子步骤 292，根据所述第一数量确定所述地址标识与所述最终经纬度信息的惩罚损失分值。

在实际应用中，如果地址标识对应的最大簇中包含经纬度信息越多，则说明由该最大簇中包含的经纬度信息确认得到的相应地址标识的最终经纬

度信息的准确度越高，反之则越低。因此，在本申请一些实施例中，为了提高最终得到的置信度的准确性，还可以进一步基于相应地址标识对应的最大簇中所包含的经纬度信息的第一数量，确认相应地址标识及其最终经纬度信息的惩罚损失分值。其中，第一数量与惩罚损失分值之间的对应关系可以根据需求或者是经验等进行预先设置，对此本申请实施例不加以限定。例如，

5 可以设置第一数量与惩罚损失分值呈线性反比关系，等等。

子步骤 293，基于所述预设基础分值、所述初始分值以及所述惩罚损失分值，确定所述地址标识与所述最终经纬度信息的置信度。

在确定了惩罚损失分值之后，则可以基于预设基础分值、所述初始分值

10 以及所述惩罚损失分值，确定相应地址标识及其最终经纬度信息的置信度。其中，置信度与预设基础分值、初始分值以及惩罚损失分值之间的对应关系可以根据需求或者是经验等进行预先设置，对此本申请实施例不加以限定。

可选地，在本申请实施例中，可以根据公式

$C = \text{baseScore} + (1 - D / \text{threshold})^2 \times (100 - \text{baseScore}) - \text{cntLossScore}$ ，确定所述

15 地址标识与所述最终经纬度信息的置信度；其中，C 表示所述地址标识与所述最终经纬度信息的置信度，baseScore 表示预设基础分值，D 表示所述第一距离，threshold 表示预设距离阈值，cntLossScore 表示基于所述第一数量的惩罚损失分值；当所述第一距离大于预设距离阈值时，所述 $(1 - D / \text{threshold})$ 的取值为 0。

20 在本申请实施例中，可以设置置信度的取值为百分制，且取值越大则说明相应地址标识及其最终经纬度信息的准确度越高。根据经验则可以根据公式 $C = \text{baseScore} + (1 - D / \text{threshold})^2 \times (100 - \text{baseScore}) - \text{cntLossScore}$ ，确定所述地址标识与所述最终经纬度信息的置信度。其中，baseScore、threshold，以及 cntLossScore 的具体取值均可以根据需求进行预设设置，对此本申请实施

25 例不加以限定。而且，如果地址标识对应的第一距离大于预设距离阈值，则说明该地址标识的最终经纬度信息距离其所属道路较远，偏差较大，此时则可以直接设置上述公式中的 $(1 - D / \text{threshold})$ 的取值为 0，从而降低该地址标识及其最终经纬度信息的置信度。

根据上述公式可以看出，此时初始分值与第一距离以及预设基础分值和

30 预设距离阈值之间的对应关系为：

$\text{InitialScore} = (1 - D / \text{threshold})^2 \times (100 - \text{baseScore})$ ，其中 InitialScore 为初始分值，此时是取 $(1 - D / \text{threshold})$ 的平方值计算初始分值，在实际应用中也可以

取 $(1-D/\text{threshold})$ 的 N 次方计算初始分值，其中的 N 可以根据需求或经验等进行预先设置，对此本申请实施例不加以限定。

可选地，根据实际生产中的经验，为了能够比较好的将置信度映射到 100 分的区间内，且能较好的反应地址标识及其最终经纬度的可信度，在本申请
5 一些实施例中，可以设置所述预设基础分值为 35；所述预设距离阈值为 500 米；当所述第一数量小于等于 3 时，所述惩罚损失分值为 20，当所述第一数量大于 3 且小于等于 6 时，所述惩罚损失分值为 10，当所述第一数量大于 6 且小于等于 10 时，所述惩罚损失分值为 5，当所述第一数量大于 10 时，所述惩罚损失分值为 0。

10 可选地，在本申请实施例中，还包括：

步骤 2110，当接收到定位服务请求时，根据所述定位服务请求的精度要求，以及所述地址标识与所述最终经纬度信息的置信度，对所述定位服务请求进行定位服务响应。

在本申请实施例中，在确定了各地址标识及其最终经纬度信息的置信度
15 之后，则可以基于置信度，将各地址标识及其最终经纬度信息进行应用。比如定位服务，当接收到定位服务请求时，则可以根据所述定位服务请求的精度要求，以及所述地址标识与所述最终经纬度信息的置信度，对所述定位服务请求进行定位服务响应。例如，如果定位服务请求的精度要求很高的话，就只使用置信度非常高的地址标识及其最终经纬度信息，对所述定位服务请
20 求进行定位服务响应；而如果定位服务请求的精度要求不高，则可以宽泛一些。

其中，定位服务请求的精度要求与置信度之间的对应关系可以根据需求进行预先设置，对此本申请实施例不加以限定。例如，可以在设置精度要求时，直接设置精度要求的内容为要求置信度在预设分数以上；或者设置不同
25 的精度要求等级，而不同的精度要求等级又分别对应于不同的置信度取值范围，等等。

步骤 2120，根据所述地址标识与所述最终经纬度信息的置信度，对所述最终经纬度信息对应商户的地址进行调整。

另外，在本申请实施例中，还可以基于置信度校准各商户的地址。例如，
30 对于某一经纬度信息对应的商户，可以根据以该经纬度作为最终经纬度信息的地址标识的置信度取值大小，考虑是否将商户的地址更改为此经纬度信息对应的地址标识。具体的，可以根据需求预先设置一置信度阈值，当地址标

识及其最终经纬度信息的置信度的取值大于等于该置信度阈值，则可以将相应商户的地址中的地址标识部分更改为此最终经纬度信息对应的地址标识，或者可以提示商户将相应商户的地址中的地址标识部分更改为此最终经纬度信息对应的地址标识，并由商户决定最终是否更改；否则不更改或者不

5 行提示。

步骤 2130，根据调整后的商户地址以及所述商户地址与道路的关系，更新所述地理知识库。

在本申请实施例中，在基于置信度对商户的地址进行调整之后，为了避免后续在确定该商户的地址时出现错误，可以根据调整后的商户地址以及所述商户地址与道路的关系，更新所述地理知识库。具体的，可以将地理知识库中记录的相应商户与其调整前的地址之间对应关系取消或者删除，而将调整后的商户地址与相应商户的对应关系，以及商户地址与道路的关系，记录至地理知识库中。

例如，假设在地理知识库中记录有商户 A 所在地址为“长宁区安化路 492 号”，可以看出此时商户 A 属于“安化路”，而基于上述的置信度调整后的商户 A 所在地址为“长宁区安贞路 500 号”，那么此时则可以将地理知识库中的商户 A 与“长宁区安化路 492 号”的对应关系取消或删除，而将商户 A 地址更新为“长宁区安贞路 500 号”。

在本申请一些实施例中，通过获取用以挖掘地址标识及其经纬度的原始数据；所述原始数据包括兴趣点数据和/或用户原创内容行为数据；获取所述原始数据中的地址标识，以及与所述地址标识对应的经纬度信息；针对对应多个经纬度信息的地址标识，通过聚类算法确定所述地址标识对应的最终经纬度信息。从而降低了获取地址标识及其经纬度的人力成本，同时提高了时效性。

其次，在本申请另一些实施例中，还可以获取所述原始数据中的地址经纬度数据；所述地址经纬度数据包括地址数据，经纬度数据，以及地址与经纬度的对应关系；基于地理知识库，对所述地址数据进行结构化处理，并取结构化处理后得到的道路名称以及门牌号作为与所述地址数据对应的地址标识；所述地理知识库中包括地理信息实体库，以及各所述地理信息实体之间的关系；基于所述地址与经纬度的对应关系，将与所述地址数据对应的经纬度数据作为与所述地址标识对应的经纬度信息。从而进一步提高了地址标识及其经纬度的挖掘效率。

再次，在本申请再一些实施例中，还可以基于对应多个经纬度信息的地址标识，则根据第一距离阈值以及第一样本阈值，采用密度聚类算法对所述多个经纬度信息进行聚类，得到至少一个聚类簇；从所述至少一个聚类簇中选定最大簇；对所述最大簇中的经纬度信息取平均，得到平均后的经纬度数值作为所述地址标识对应的最终经纬度信息。并且，以所述聚类簇中包含的经纬度信息数量最多的一个聚类簇作为所述最大簇。从而提高确定的最终经纬度信息的准确度。

而且，在本申请一些实施例中，还可以获取所述地址标识对应的最大簇中包含的经纬度信息的第一数量，以及所述地址标识对应的最终经纬度信息与所述地址标识对应的道路之间的第一距离；根据所述第一数量以及所述第一距离，确定所述地址标识与所述最终经纬度信息的置信度。并且，根据所述第一距离以及预设基础分值和预设距离阈值，确定所述地址标识与所述最终经纬度信息的初始分值；根据所述第一数量确定所述地址标识与所述最终经纬度信息的惩罚损失分值；基于所述预设基础分值、所述初始分值以及所述惩罚损失分值，确定所述地址标识与所述最终经纬度信息的置信度，从而可以进一步确定各地址标识及其最终经纬度信息的准确性。

另外，在本申请一些实施例中，当接收到定位服务请求时，根据所述定位服务请求的精度要求，以及所述地址标识与所述最终经纬度信息的置信度，对所述定位服务请求进行定位服务响应；根据所述地址标识与所述最终经纬度信息的置信度，对所述最终经纬度信息对应商户的地址进行调整；根据调整后的商户地址以及所述商户地址与道路的关系，更新所述地理知识库。从而基于置信度将挖掘得到的地址标识与所述最终经纬度信息进行应用，同时提高地理知识库的时效性以及准确性。

对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请实施例并不受所描述的动作顺序的限制，因为依据本申请实施例，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作并不一定是本申请实施例所必须的。

实施例三

详细介绍本申请实施例提供的一种地址标识及其经纬度的挖掘装置。

参照图 4，示出了本申请实施例中一种地址标识及其经纬度的挖掘装置的结构示意图。具体包括：

原始数据获取模块 310, 用于获取用以挖掘地址标识及其经纬度的原始数据;

数据挖掘模块 320, 用于获取所述原始数据中的地址标识, 以及与所述地址标识对应的经纬度信息;

- 5 最终经纬度信息确认模块 330, 用于针对对应多个经纬度信息的地址标识, 通过密度聚类算法确定所述地址标识对应的最终经纬度信息。

可选地, 在本申请实施例中, 所述原始数据包括兴趣点数据和/或用户原创内容行为数据。

在本申请实施例中, 通过获取用以挖掘地址标识及其经纬度的原始数据;

- 10 获取所述原始数据中的地址标识, 以及与所述地址标识对应的经纬度信息; 针对对应多个经纬度信息的地址标识, 通过密度聚类算法确定所述地址标识对应的最终经纬度信息。从而降低了获取地址标识及其经纬度的人力成本, 同时提高了时效性。

实施例四

- 15 下面结合图 5 详细介绍本申请另一实施例提供的一种地址标识及其经纬度的挖掘装置。

参照图 5, 示出了本申请实施例中一种地址标识及其经纬度的挖掘装置的结构示意图。具体包括:

- 20 原始数据获取模块 410, 用于获取用以挖掘地址标识及其经纬度的原始数据;

数据挖掘模块 420, 用于获取所述原始数据中的地址标识, 以及与所述地址标识对应的经纬度信息。

其中, 所述数据挖掘模块 420, 进一步可以包括:

- 25 数据挖掘子模块 421, 用于获取所述原始数据中的地址经纬度数据; 所述地址经纬度数据包括地址数据, 经纬度数据, 以及地址与经纬度的对应关系;

结构化处理子模块 422, 用于基于地理知识库, 对所述地址数据进行结构化处理, 并取结构化处理后得到的道路名称以及门牌号作为与所述地址数据对应的地址标识; 所述地理知识库中包括地理信息实体库, 以及各所述地理信息实体之间的关系; 以及,

- 30 经纬度信息确认子模块 423, 用于基于所述地址与经纬度的对应关系, 将与所述地址数据对应的经纬度数据作为与所述地址标识对应的经纬度信息。

参照图 5, 所述装置还包括最终经纬度信息确认模块 430, 用于针对对应

多个经纬度信息的地址标识,通过聚类算法确定所述地址标识对应的最终经纬度信息。

其中,所述最终经纬度信息确认模块 430,进一步可以包括:

密度聚类子模块 431,用于基于对应多个经纬度信息的地址标识,则根据
5 第一距离阈值以及第一样本阈值,采用密度聚类算法对所述多个经纬度信息进行聚类,得到至少一个聚类簇;

最大簇确定子模块 432,用于从所述至少一个聚类簇中选定最大簇。

可选地,在本申请实施例中,所述最大簇确定子模块 432,还用于以所述聚类簇中包含的经纬度信息数量最多的一个聚类簇作为所述最大簇。

10 所述最终经纬度信息确认模块 430,进一步还可以包括最终经纬度信息获取子模块 433,用于对所述最大簇中的经纬度信息取平均,得到平均后的经纬度数值作为所述地址标识对应的最终经纬度信息;

置信度数据获取模块 440,用于获取所述地址标识对应的最大簇中包含的经纬度信息的第一数量,以及所述地址标识对应的最终经纬度信息与所述地
15 址标识对应的道路之间的第一距离;以及,

置信度确定模块 450,用于根据所述第一数量以及所述第一距离,确定所述地址标识与所述最终经纬度信息的置信度。

可选地,在本申请实施例中,所述置信度确定模块 450,进一步可以包括:

初始分值确定子模块,用于根据所述第一距离以及预设基础分值和预设
20 距离阈值,确定所述地址标识与所述最终经纬度信息的初始分值;

损失分值确定子模块,用于根据所述第一数量确定所述地址标识与所述最终经纬度信息的惩罚损失分值;

置信度确定子模块,用于基于所述预设基础分值、所述初始分值以及所述惩罚损失分值,确定所述地址标识与所述最终经纬度信息的置信度。

25 可选地,在本申请实施例中,所述装置还可以包括:

定位服务响应模块,用于当接收到定位服务请求时,根据所述定位服务请求的精度要求,以及所述地址标识与所述最终经纬度信息的置信度,对所述定位服务请求进行定位服务响应。

30 商户地址调整模块,用于根据所述地址标识与所述最终经纬度信息的置信度,对所述最终经纬度信息对应商户的地址进行调整。

地理知识库更新模块,用于根据调整后的商户地址以及所述商户地址与道路的关系,更新所述地理知识库。

本申请实施例还提供了一种电子设备，包括：

处理器、存储器以及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现如上述的地址标识及其经纬度的挖掘方法。

- 5 本申请实施例还提供了一种可读存储介质，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行如上述的地址标识及其经纬度的挖掘方法。

在本申请实施例中，通过获取用以挖掘地址标识及其经纬度的原始数据；所述原始数据包括兴趣点数据和/或用户原创内容行为数据；获取所述原始数据
10 据中的地址标识，以及与所述地址标识对应的经纬度信息；针对对应多个经纬度信息的地址标识，通过聚类算法确定所述地址标识对应的最终经纬度信息，从而降低了获取地址标识及其经纬度的人力成本，同时提高了时效性。

其次，在本申请另一些实施例中，还可以获取所述原始数据中的地址经纬度数据；所述地址经纬度数据包括地址数据，经纬度数据，以及地址与经
15 纬度的对应关系；基于地理知识库，对所述地址数据进行结构化处理，并取结构化处理后得到的道路名称以及门牌号作为与所述地址数据对应的地址标识；所述地理知识库中包括地理信息实体库，以及各所述地理信息实体之间的关系；基于所述地址与经纬度的对应关系，将与所述地址数据对应的经纬度数据作为与所述地址标识对应的经纬度信息。从而进一步提高了地址标
20 识及其经纬度的挖掘效率。

再次，在本申请又一些实施例中，还可以基于对应多个经纬度信息的地址标识，则根据第一距离阈值以及第一样本阈值，采用密度聚类算法对所述
25 多个经纬度信息进行聚类，得到至少一个聚类簇；从所述至少一个聚类簇中选定最大簇；对所述最大簇中的经纬度信息取平均，得到平均后的经纬度数值作为所述地址标识对应的最终经纬度信息。并且，以所述聚类簇中包含的经纬度信息数量最多的一个聚类簇作为所述最大簇。从而提高确定的最终经纬度信息的准确度。

而且，在本申请一些实施例中，还可以获取所述地址标识对应的最大簇中包含的经纬度信息的第一数量，以及所述地址标识对应的最终经纬度信息
30 与所述地址标识对应的道路之间的第一距离；根据所述第一数量以及所述第一距离，确定所述地址标识与所述最终经纬度信息的置信度。并且，根据所述第一距离以及预设基础分值和预设距离阈值，确定所述地址标识与所述最

终经纬度信息的初始分值；根据所述第一数量确定所述地址标识与所述最终经纬度信息的惩罚损失分值；基于所述预设基础分值、所述初始分值以及所述惩罚损失分值，确定所述地址标识与所述最终经纬度信息的置信度，从而可以进一步确定各地址标识及其最终经纬度信息的准确性。

- 5 另外，在本申请一些实施例中，当接收到定位服务请求时，根据所述定位服务请求的精度要求，以及所述地址标识与所述最终经纬度信息的置信度，对所述定位服务请求进行定位服务响应；根据所述地址标识与所述最终经纬度信息的置信度，对所述最终经纬度信息对应商户的地址进行调整；根据调整后的商户地址以及所述商户地址与道路的关系，更新所述地理知识
- 10 库。从而基于置信度将挖掘得到的地址标识与所述最终经纬度信息进行应用，同时提高地理知识库的时效性以及准确性。

对于装置实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

- 在此提供的算法和显示不与任何特定计算机、虚拟系统或者其它设备固
- 15 有相关。各种通用系统也可以与基于在此的示教一起使用。根据上面的描述，构造这类系统所要求的结构是显而易见的。此外，本申请也不针对任何特定编程语言。应当明白，可以利用各种编程语言实现在此描述的本申请的内容，并且上面对特定语言所做的描述是为了披露本申请的最佳实施方式。

- 在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本
- 20 申请的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

- 类似地，应当理解，为了精简本申请并帮理解各个发明方面中的一个或多个，在上面对本申请的示例性实施例的描述中，本申请的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该公开
- 25 的方法解释成反映如下意图：即所要求保护的本申请要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如下面的权利要求书所反映的那样，申请方面在于少于前面公开的单个实施例的所有特征。因此，遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本申请的单独实施例。

- 30 本领域那些技术人员可以理解，可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件，以及此外可

以把它们分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外，可以采用任何组合对本说明书（包括伴随的权利要求、摘要和附图）中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书（包括伴随的权利要求、摘要和附图）中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

此外，本领域的技术人员能够理解，尽管在此所述的一些实施例包括其它实施例中所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本申请的范围之内并且形成不同的实施例。例如，在下面的权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

本申请的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器（DSP）来实现根据本申请实施例的地址标识及其经纬度的挖掘设备中的一些或者全部部件的一些或者全部功能。本申请还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序（例如，计算机程序和计算机程序产品）。这样的实现本申请的程序可以存储在计算机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到，或者在载体信号上提供，或者以任何其他形式提供。

例如，图 6 示出了可以实现根据本申请的方法的电子设备，该电子设备传统上包括处理器 610 和以存储器 620 形式的计算机程序产品或者计算机可读存储介质。所述电子设备可以为 PC 机、移动终端、个人数字助理、平板电脑等。所述电子设备为计算处理设备。所述存储器 620 可以是诸如闪存、EEPROM（电可擦除可编程只读存储器）、EPROM、硬盘或者 ROM 之类的电子存储器。存储器 620 具有用于执行上述方法中的任何方法步骤的程序代码的存储空间 6201。例如，用于程序代码的存储空间 6201 可以包括分别用于实现上面的方法中的各种步骤的程序代码 6202。所述计算机程序产品或者计算机可读存储介质，其上存储有计算机程序的程序代码，该程序代码被处理器 610 执行时实现如本申请实施例一和实施例二所述的地址标识及其经纬度挖掘的方法。这些程序代码可以从一个或者多个计算机程序产品中读出或

者写入到这一个或者多个计算机程序产品中。这些计算机程序产品包括诸如硬盘，紧致盘（CD）、存储卡或者软盘之类的程序代码载体。这样的计算机程序产品或者计算机可读存储介质通常为如参考图 7 所述的便携式或者固定存储单元。该存储单元用于保持或者携带实现根据本申请的方法的程序代码，该存储单元可以具有与图 6 的电子设备中的存储器 620 类似布置的存储段、存储空间等。程序代码可以例如以适当形式进行压缩。通常，存储单元包括计算机可读代码 6202’，即可以由例如诸如 610 之类的处理器读取的代码，这些代码当电子设备运行时，导致该电子设备执行上面所描述的方法中的各个步骤。

10 本文中所述的“一个实施例”、“实施例”或者“一个或者多个实施例”意味着，结合实施例描述的特定特征、结构或者特性包括在本公开的至少一个实施例中。此外，请注意，这里“在一个实施例中”的词语例子不一定全指同一个实施例。

在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，15 本申请的实施例可以在没有这些具体细节的情况下被实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本申请可以借助于包20 括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中，这些装置中的若干个可以通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

应该注意的是上述实施例对本申请进行说明而不是对本申请进行限制，25 并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本申请可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。30 在列举了若干装置的单元权利要求中，这些装置中的若干个可以通过同一

个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

权 利 要 求

1、一种地址标识及其经纬度的挖掘方法，包括：

获取用以挖掘地址标识及其经纬度的原始数据；

获取所述原始数据中的地址标识，以及与所述地址标识对应的经纬度信

5 息；

针对对应多个经纬度信息的地址标识，通过聚类算法确定所述地址标识对应的最终经纬度信息。

2、根据权利要求 1 所述的方法，所述获取所述原始数据中的地址标识，以及与所述地址标识对应的经纬度信息的步骤，包括：

10 获取所述原始数据中的地址经纬度数据；所述地址经纬度数据包括地址数据，经纬度数据，以及地址与经纬度的对应关系；

基于地理知识库，对所述地址数据进行结构化处理，并取结构化处理后得到的道路名称以及门牌号作为与所述地址数据对应的地址标识；所述地理知识库中包括地理信息实体库，以及各所述地理信息实体之间的关系；

15 基于所述地址与经纬度的对应关系，将与所述地址数据对应的经纬度数据作为与所述地址标识对应的经纬度信息。

3、根据权利要求 1 所述的方法，所述针对对应多个经纬度信息的地址标识，通过聚类算法确定所述地址标识对应的最终经纬度信息的步骤，包括：

20 基于对应多个经纬度信息的地址标识，则根据第一距离阈值以及第一样本阈值，采用密度聚类算法对所述多个经纬度信息进行聚类，得到至少一个聚类簇；

从所述至少一个聚类簇中选定最大簇；

对所述最大簇中的经纬度信息取平均，得到平均后的经纬度数值作为所述地址标识对应的最终经纬度信息。

25 4、根据权利要求 3 所述的方法，所述从所述至少一个聚类簇中选定最大簇的步骤，包括：

以所述聚类簇中包含的经纬度信息数量最多的一个聚类簇作为所述最大簇。

30 5、根据权利要求 3 所述的方法，在所述对于对应多个经纬度信息的地址标识，基于密度聚类算法确定所述地址标识对应的最终经纬度信息的步骤之后，还包括：

获取所述地址标识对应的最大簇中包含的经纬度信息的第一数量，以及

所述地址标识对应的最终经纬度信息与所述地址标识对应的道路之间的第一距离；

根据所述第一数量以及所述第一距离，确定所述地址标识与所述最终经纬度信息的置信度。

- 5 6、根据权利要求 5 所述的方法，所述根据所述第一数量以及所述第一距离，确定所述地址标识与所述最终经纬度信息的置信度的步骤，包括：

根据所述第一距离以及预设基础分值和预设距离阈值，确定所述地址标识与所述最终经纬度信息的初始分值；

- 10 根据所述第一数量确定所述地址标识与所述最终经纬度信息的惩罚损失分值；

基于所述预设基础分值、所述初始分值以及所述惩罚损失分值，确定所述地址标识与所述最终经纬度信息的置信度。

- 15 7、根据权利要求 5 所述的方法，在所述根据所述第一数量以及所述第一距离，确定所述地址标识与所述最终经纬度信息的置信度的步骤之后，还包括：

当接收到定位服务请求时，根据所述定位服务请求的精度要求，以及所述地址标识与所述最终经纬度信息的置信度，对所述定位服务请求进行定位服务响应；

- 20 根据所述地址标识与所述最终经纬度信息的置信度，对所述最终经纬度信息对应商户的地址进行调整；

根据调整后的商户地址以及所述商户地址与道路的关系，更新所述地理知识库。

8、根据权利要求 1-7 任一项所述的方法，所述原始数据包括兴趣点数据和/或用户原创内容行为数据。

- 25 9、一种地址标识及其经纬度的挖掘装置，包括：

原始数据获取模块，用于获取用以挖掘地址标识及其经纬度的原始数据；

数据挖掘模块，用于获取所述原始数据中的地址标识，以及与所述地址标识对应的经纬度信息；

- 30 最终经纬度信息确认模块，用于针对对应多个经纬度信息的地址标识，通过聚类算法确定所述地址标识对应的最终经纬度信息。

10、根据权利要求 9 所述的装置，所述数据挖掘模块，包括：

数据掘子模块,用于获取所述原始数据中的地址经纬度数据;所述地址经纬度数据包括地址数据,经纬度数据,以及地址与经纬度的对应关系;

结构化处理子模块,用于基于地理知识库,对所述地址数据进行结构化处理,并取结构化处理后得到的道路名称以及门牌号作为与所述地址数据对应的地址标识;所述地理知识库中包括地理信息实体库,以及各所述地理信息实体之间的关系;

经纬度信息确认子模块,用于基于所述地址与经纬度的对应关系,将与所述地址数据对应的经纬度数据作为与所述地址标识对应的经纬度信息。

11、根据权利要求9所述的装置,所述最终经纬度信息确认模块,包括:
10 密度聚类子模块,用于基于对应多个经纬度信息的地址标识,则根据第一距离阈值以及第一样本阈值,采用密度聚类算法对所述多个经纬度信息进行聚类,得到至少一个聚类簇;

最大簇确定子模块,用于从所述至少一个聚类簇中选定最大簇;

最终经纬度信息获取子模块,用于对所述最大簇中的经纬度信息取平均,得到平均后的经纬度数值作为所述地址标识对应的最终经纬度信息。

12、根据权利要求11所述的装置,所述最大簇确定子模块,还用于以所述聚类簇中包含的经纬度信息数量最多的一个聚类簇作为所述最大簇。

13、根据权利要求11所述的装置,还包括:

置信度数据获取模块,用于获取所述地址标识对应的最大簇中包含的经纬度信息的第一数量,以及所述地址标识对应的最终经纬度信息与所述地址标识对应的道路之间的第一距离;

置信度确定模块,用于根据所述第一数量以及所述第一距离,确定所述地址标识与所述最终经纬度信息的置信度。

14、根据权利要求13所述的装置,所述置信度确定模块,包括:

25 初始分值确定子模块,用于根据所述第一距离以及预设基础分值和预设距离阈值,确定所述地址标识与所述最终经纬度信息的初始分值;

损失分值确定子模块,用于根据所述第一数量确定所述地址标识与所述最终经纬度信息的惩罚损失分值;

置信度确定子模块,用于基于所述预设基础分值、所述初始分值以及所述惩罚损失分值,确定所述地址标识与所述最终经纬度信息的置信度。

15、根据权利要求13所述的装置,还包括:

定位服务响应模块,用于当接收到定位服务请求时,根据所述定位服务

请求的精度要求,以及所述地址标识与所述最终经纬度信息的置信度,对所述定位服务请求进行定位服务响应;

商户地址调整模块,用于根据所述地址标识与所述最终经纬度信息的置信度,对所述最终经纬度信息对应商户的地址进行调整;

- 5 地理知识库更新模块,用于根据调整后的商户地址以及所述商户地址与道路的关系,更新所述地理知识库。

16、根据权利要求 9-15 任一项所述的装置,所述原始数据包括兴趣点数据和/或用户原创内容行为数据。

17、一种电子设备,包括:

- 10 处理器、存储器以及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如权利要求 1-8 中的任一项所述的地址标识及其经纬度的挖掘方法。

- 18、一种计算机程序,包括计算机可读代码,当所述计算机可读代码在电子设备上运行时,导致所述电子设备执行根据权利要求 1-8 中的任一项所述的地址标识及其经纬度的挖掘方法。
- 15

19、一种可读存储介质,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行如权利要求 1-8 中的任一项所述的地址标识及其经纬度的挖掘方法。

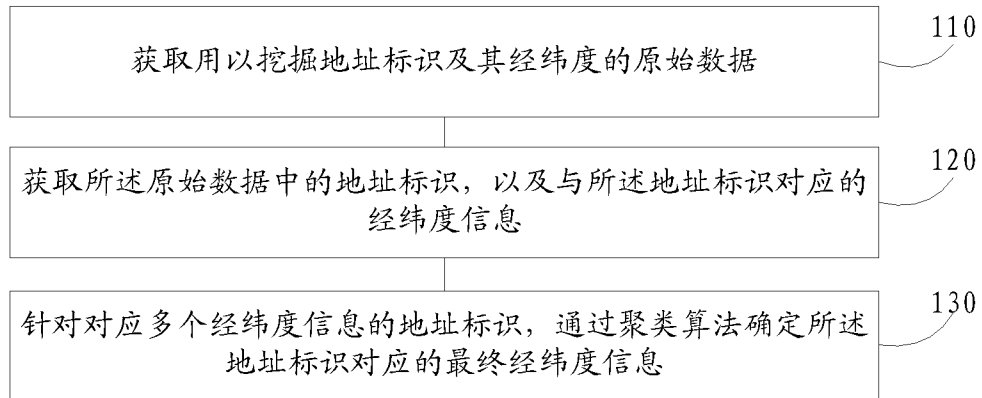


图 1

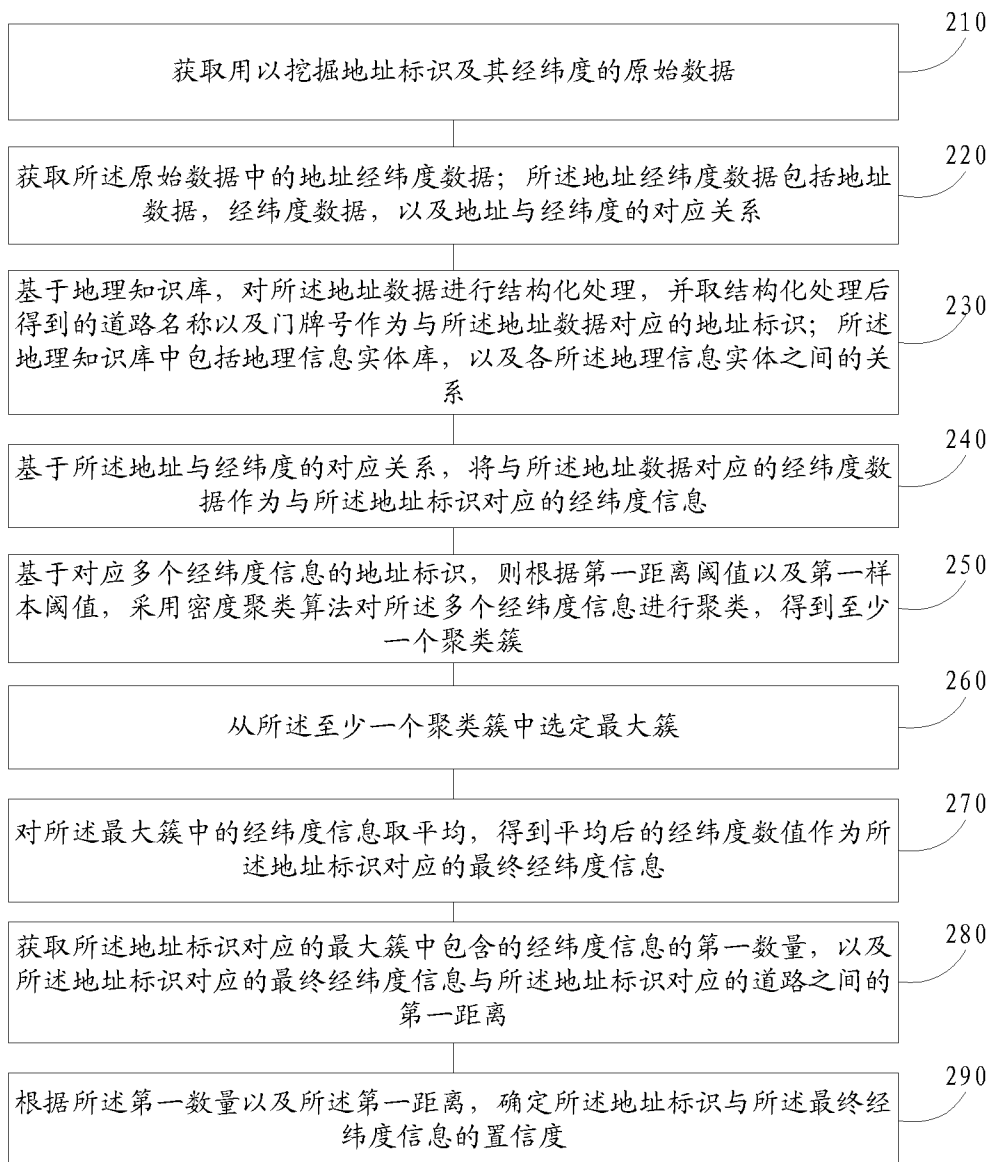


图 2

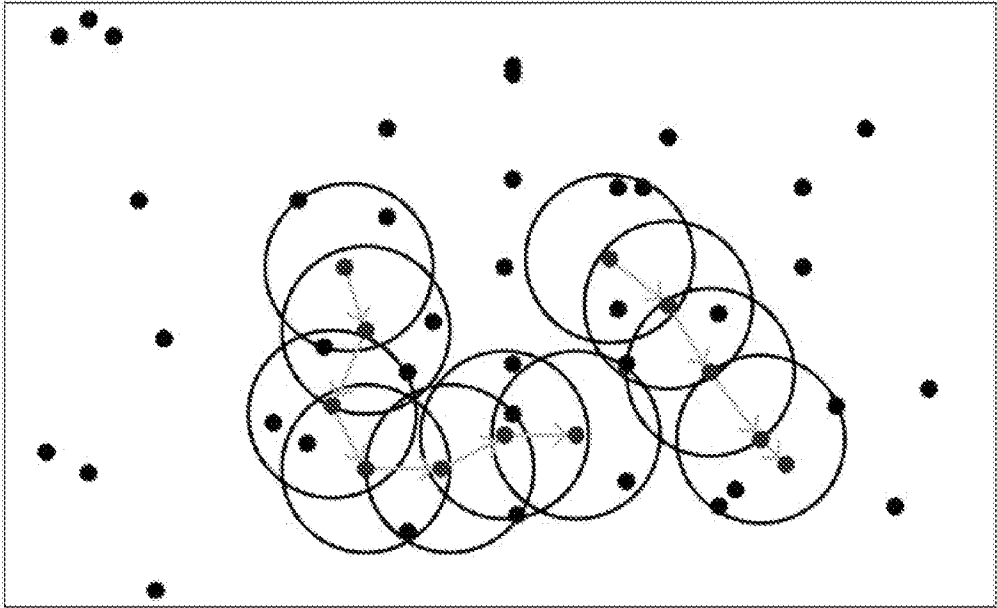


图 3

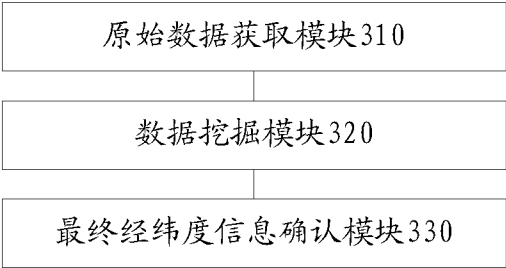


图 4

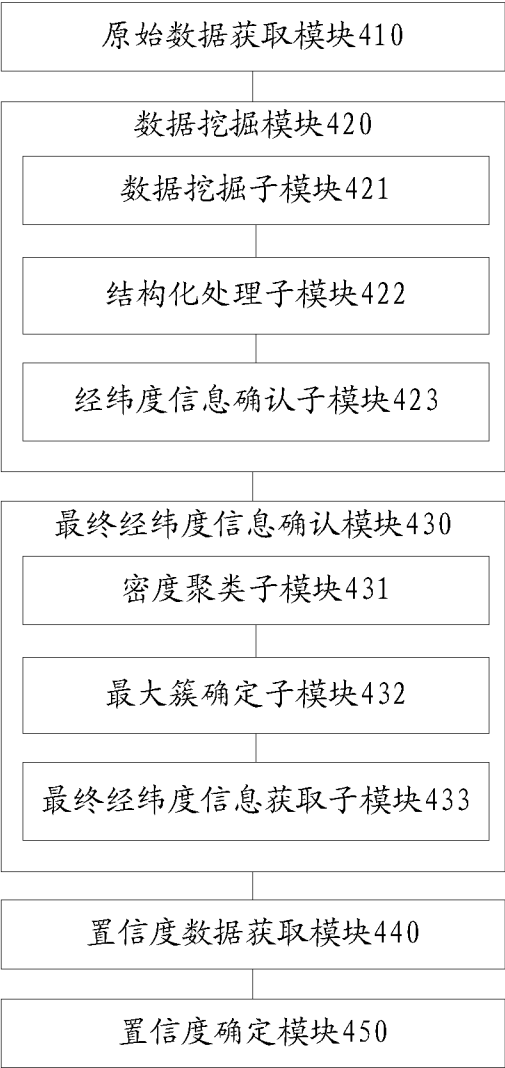


图 5

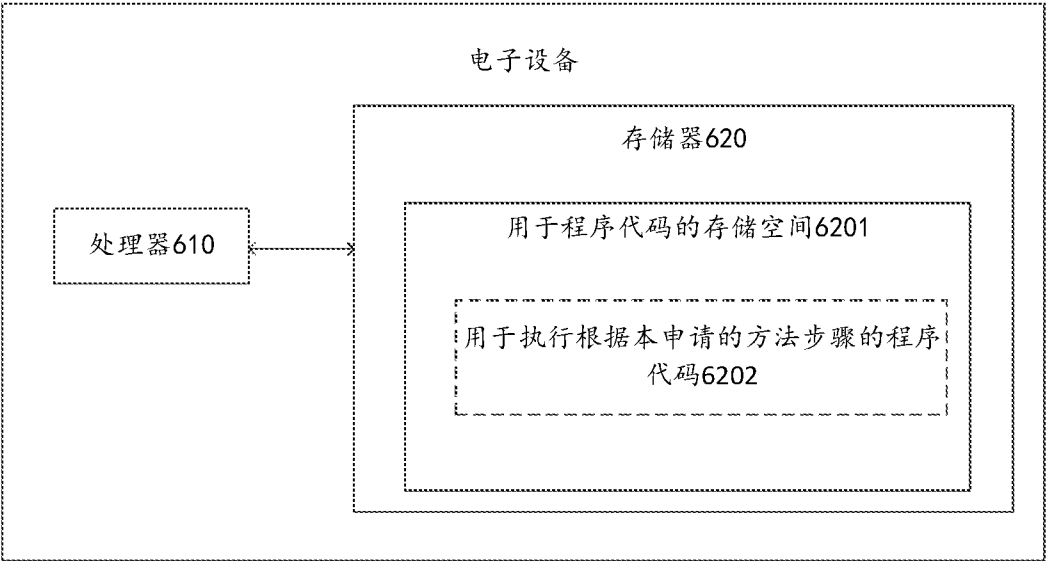


图 6

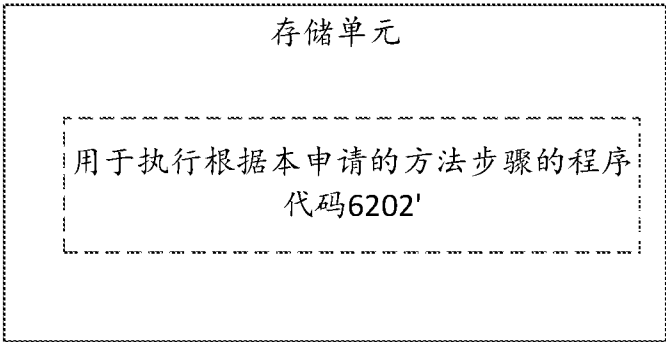


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/095106

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2006.01)i; G06F 16/29(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)

G06K, 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: 地址, 地理, 标识, 道路, 门牌, 经纬度, 经度, 纬度, 地理坐标, 位置, 校准, 校正, 正确, 有误, 纠错, 修正, 偏差, 多个, 对应, 地图, 导航, 聚类, address, identif+, road, longitude, latitude, coordinate, adjust+, calibrat+, wrong, correct+, modif+, multiple, map, navigat+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109376761 A (BEIJING SCIENCE AND TECHNOLOGY CO., THREE FAST ONLINE) 22 February 2019 (2019-02-22) claims 1-18, and description, paragraphs [0027]-[0157]	1-19
A	CN 107547633 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 05 January 2018 (2018-01-05) description, paragraphs [0039]-[0106]	1-19
A	CN 108271120 A (ALIBABA GROUP HOLDING LIMITED) 10 July 2018 (2018-07-10) entire document	1-19
A	CN 101441088 A (WU, YUXIAN) 27 May 2009 (2009-05-27) entire document	1-19
A	CN 104050196 A (ALIBABA GROUP HOLDING LIMITED) 17 September 2014 (2014-09-17) entire document	1-19
A	US 2016092456 A1 (UNITED STATES POSTAL SERVICE) 31 March 2016 (2016-03-31) entire document	1-19

☐ 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

17 September 2019

Date of mailing of the international search report

27 September 2019

Name and mailing address of the ISA/CN

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

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/095106

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109376761	A	22 February 2019	None			
CN	107547633	A	05 January 2018	None			
CN	108271120	A	10 July 2018	None			
CN	101441088	A	27 May 2009	None			
CN	104050196	A	17 September 2014	HK	1201601	A1	04 September 2015
US	2016092456	A1	31 March 2016	WO	2016049357	A1	31 March 2016
				EP	3198535	A1	02 August 2017
				CN	107004175	A	01 August 2017

国际检索报告

国际申请号

PCT/CN2019/095106

A. 主题的分类

G06K 9/62 (2006.01) i; G06F 16/29 (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)

G06K, 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, EPDOC: 地址, 地理, 标识, 道路, 门牌, 经纬度, 经度, 纬度, 地理坐标, 位置, 校准, 校正, 正确, 有误差, 纠错, 修正, 偏差, 多个, 对应, 地图, 导航, 聚类, address, identif+, road, longitude, latitude, coordinate, adjust+, calibrat+, wrong, correct+, modif+, multiple, map, navigat+

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
PX	CN 109376761 A (北京三快在线科技有限公司) 2019年 2月 22日 (2019 - 02 - 22) 权利要求1-18, 说明书第[0027]-[0157]段	1-19
A	CN 107547633 A (腾讯科技深圳有限公司) 2018年 1月 5日 (2018 - 01 - 05) 说明书第[0039]-[0106]段	1-19
A	CN 108271120 A (阿里巴巴集团控股有限公司) 2018年 7月 10日 (2018 - 07 - 10) 全文	1-19
A	CN 101441088 A (吴玉先) 2009年 5月 27日 (2009 - 05 - 27) 全文	1-19
A	CN 104050196 A (阿里巴巴集团控股有限公司) 2014年 9月 17日 (2014 - 09 - 17) 全文	1-19
A	US 2016092456 A1 (UNITED STATES POSTAL SERVICE) 2016年 3月 31日 (2016 - 03 - 31) 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

2019年 9月 17日

国际检索报告邮寄日期

2019年 9月 27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

受权官员

王伟

电话号码 86-(10)-53961525

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/095106

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	109376761	A	2019年 2月 22日	无			
CN	107547633	A	2018年 1月 5日	无			
CN	108271120	A	2018年 7月 10日	无			
CN	101441088	A	2009年 5月 27日	无			
CN	104050196	A	2014年 9月 17日	HK	1201601	A1	2015年 9月 4日
US	2016092456	A1	2016年 3月 31日	WO	2016049357	A1	2016年 3月 31日
				EP	3198535	A1	2017年 8月 2日
				CN	107004175	A	2017年 8月 1日

表 PCT/ISA/210 (同族专利附件) (2015年1月)