



(12)发明专利申请

(10)申请公布号 CN 107580069 A

(43)申请公布日 2018.01.12

(21)申请号 201710865243.5

(22)申请日 2017.09.22

(71)申请人 挖财网络技术有限公司

地址 310012 浙江省杭州市西湖区华星路
96号第18层

(72)发明人 尤志强 潘琪 车曦

(74)专利代理机构 北京博思佳知识产权代理有
限公司 11415

代理人 林祥 李威

(51)Int.Cl.

H04L 29/08(2006.01)

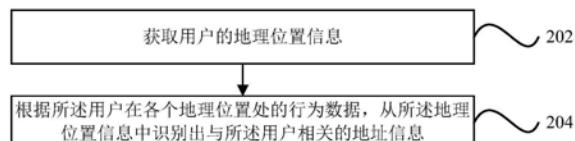
权利要求书6页 说明书15页 附图3页

(54)发明名称

用户地址的确定方法及装置

(57)摘要

本说明书提供一种用户地址的确定方法及装置,该方法可以包括:获取用户的地理位置信息;根据所述用户在各个地理位置处的行为数据,从所述地理位置信息中识别出与所述用户相关的地址信息。



1. 一种用户地址的确定方法,其特征在于,包括:
 - 获取用户的地理位置信息;
 - 根据所述用户在各个地理位置处的行为数据,从所述地理位置信息中识别出与所述用户相关的地址信息。
2. 根据权利要求1所述的方法,其特征在于,所述获取用户的地理位置信息,包括:
 - 获取所述用户所处位置的经纬度数据;
 - 将所述经纬度数据逆地址解析为相应的兴趣点信息,以将所述兴趣点信息作为所述地理位置信息。
3. 根据权利要求2所述的方法,其特征在于,所述将所述经纬度数据逆地址解析为相应的兴趣点,包括:
 - 向地图服务提供商的服务器发起逆地址解析请求;
 - 接收所述服务器返回的解析结果,所述解析结果中包含所述经纬度数据对应的兴趣点信息。
4. 根据权利要求1所述的方法,其特征在于,所述地理位置信息包括:所述用户在预设时间窗口内的地理位置信息。
5. 根据权利要求1所述的方法,其特征在于,所述根据所述用户在各个地理位置处的行为数据,从所述地理位置信息中识别出与所述用户相关的地址信息,包括:
 - 根据所述用户在各个地理位置处的行为数据,计算各个地理位置对应的地址重要性参数,以表征各个地理位置对所述用户的重要程度;
 - 根据所述地址重要性参数的取值,计算各个地理位置信息属于与所述用户相关的地址信息的概率;
 - 选取概率最大的预设数量的地理位置信息,以作为与所述用户相关的地址信息。
6. 根据权利要求5所述的方法,其特征在于,所述地址重要性参数与下述至少一种类型的行为数据相关:
 - 地理位置出现频次、地理位置出现天数、地理位置连续出现的时间间隔。
7. 根据权利要求5所述的方法,其特征在于,任一地理位置信息对应的地址重要性参数被按照下述公式进行计算:
$$W_G = \frac{TFD}{D} \times TFD \times \log_2\left(\frac{TF}{TFD} + 1\right) \times \frac{1}{\Delta t + 1}$$
其中,W_G为所述任一地理位置对应的地址重要性参数、TF为所述任一地理位置出现频次、TFD为所述任一地理位置出现天数、Δt为所述任一地理位置连续出现的时间间隔、D为所有地理位置信息对应的总天数。
8. 根据权利要求7所述的方法,其特征在于,当所述任一地理位置存在多个连续出现的时间间隔时,若多个连续出现的时间间隔的数量不大于预设数值,Δt的取值为所有时间间隔的几何平均数,若多个连续出现的时间间隔的数量大于所述预设数值,Δt的取值为所有时间间隔的中位数。
9. 根据权利要求5所述的方法,其特征在于,还包括:
 - 根据各个地理位置对应的地址稳定性参数,对相应的地址重要性参数进行修正,所述地址稳定性参数用于表征所述用户的轨迹稳定程度;

其中,修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。

10.根据权利要求9所述的方法,其特征在于,各个地理位置在各个时间段对应的地址稳定性参数采用各个地理位置所处时间段的信息熵进行表征。

11.根据权利要求10所述的方法,其特征在于,所述信息熵包括平衡熵;所述用户在时间段T对应的平衡熵被按照下述公式进行计算:

$$H_T = \frac{1}{N+2} \sum_{i=1}^M \left[(n_i + 1) \sum_{j=n_i+2}^{N+2} \frac{1}{j} \right]$$

其中,H_T为所述任一时间段对应的平衡熵、M为所述地理位置信息的总数量、N为所述用户在所述任一时间段所在的所有地点的累计出现次数、n_i为地理位置G_i的出现次数。

12.根据权利要求5所述的方法,其特征在于,还包括:

根据各个时间段对应的时间重要性参数,对相应的地址重要性参数进行修正,所述时间重要性参数用于表征各个时间段对应的重要程度;

其中,修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。

13.根据权利要求5所述的方法,其特征在于,还包括:

根据各个地理位置对应的自环情况参数,对相应的地址重要性参数进行修正,所述自环情况参数用于表征所述用户在同一周期的起始时间段与终止时间段处于同一地理位置的状况;

其中,修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。

14.根据权利要求13所述的方法,其特征在于,任一地理位置对应的自环情况参数被按照下述公式进行计算:

$$P_G = \lg \left[\frac{2 \times A^2}{B \times 0.5 + C \times 0.5} + 1 \right] + 1$$

其中,P_G为所述任一地理位置对应的自环情况参数、A为一天内针对所述任一地理位置发生自环的次数、B为一天内从所述任一地理位置出发的次数、C为一天内到达所述任一地理位置的次数。

15.根据权利要求1所述的方法,其特征在于,通过下述公式计算任一地理位置信息G属于与所述用户相关的地址信息的概率:

$$I_G = \sum_{T=0}^Q W_T \times (1 + a^{H_T}) \times P_G \times W_G \times \chi_{GT}$$

其中,I_G为任一地理位置信息G属于与所述用户相关的地址信息的概率、所述Q为时间段的数量、W_T为所述任一地理位置信息G在时间段T对应的时间重要性参数、H_T为所述任一地理位置信息G在时间段T对应的地址稳定性、P_G为所述任一地理位置信息G对应的自环情况参数、W_G为所述任一地理位置对应的地址重要性参数、a为可调系数且a ∈ (0, 1)、χ_{GT}为指示函数,其中当所述任一地理位置信息G出现在时间段T时χ_{GT}=1,当所述任一地理位置信息G未出现在时间段T时χ_{GT}=0。

16. 根据权利要求1所述的方法,其特征在于,与所述用户相关的地址信息包括一对地址信息,所述一对地址信息包括属于第一类型的第一地址信息、属于第二类型的第二地址信息;其中,所述一对地址信息符合下述可靠性验证规则中至少之一:

间隔距离验证规则,用于验证所述一对地址信息之间的间隔距离;其中,所述一对地址信息之间的间隔距离不大于预设距离;

跳转概率验证规则,用于验证所述一对地址信息之间的跳转概率;其中,所述一对地址信息之间的跳转概率大于预设概率。

17. 根据权利要求16所述的方法,其特征在于,所述一对地址信息被采用下述公式计算可靠性得分,以表征对所述可靠性验证规则的符合程度:

$$\left\{ \begin{array}{l} S_{pair} = (1 + \beta^2) \times \frac{I_{Gw} \times I_{Gf}}{(\beta^2 \times I_{Gw}) + I_{Gf}} \times \frac{D_{pair}}{100} \times (1 + P_{trans}) \\ D_{pair} = -0.01 \times d^2 + 100 \\ P_{trans} = \frac{K_{A-B}}{K_{A-ALL}} \end{array} \right.$$

其中, S_{pair} 为一对地理位置对应的可靠性得分、 I_{Gw} 为所述第一地址信息和所述第二地址信息中的任一地址被识别为与所述用户相关的地址信息的概率分值、 I_{Gf} 为所述第一地址信息和所述第二地址信息中的另一地址被识别为与所述用户相关的地址信息的概率分值、 β 为可调系数且 $\beta \in (0, 1)$ 、 D_{pair} 为应用于所述间隔距离验证规则的距离衰减函数、 d 为所述第一地址信息和所述第二地址信息之间的间隔距离、 P_{trans} 为所述第一地址信息对所述第二地址信息的跳转概率、 K_{A-B} 为所述第一地址信息对所述第二地址信息的跳转次数、 K_{A-ALL} 为所述第一地址信息对所有地理位置的跳转次数。

18. 根据权利要求1所述的方法,其特征在于,还包括:

根据与所述用户相关的地址信息,以及与所述地址信息相关的特定信息,生成所述用户的画像数据。

19. 一种用户地址的确定装置,其特征在于,包括:

位置获取单元,获取用户的地理位置信息;

地址识别单元,根据所述用户在各个地理位置处的行为数据,从所述地理位置信息中识别出与所述用户相关的地址信息。

20. 根据权利要求19所述的装置,其特征在于,所述位置获取单元具体用于:

获取所述用户所处位置的经纬度数据;

将所述经纬度数据逆地址解析为相应的兴趣点信息,以将所述兴趣点信息作为所述地理位置信息。

21. 根据权利要求20所述的装置,其特征在于,所述位置获取单元通过下述方式将所述经纬度数据逆地址解析为相应的兴趣点:

向地图服务提供商的服务器发起逆地址解析请求;

接收所述服务器返回的解析结果,所述解析结果中包含所述经纬度数据对应的兴趣点信息。

22. 根据权利要求19所述的装置,其特征在于,所述地理位置信息包括:所述用户在预

设时间窗口内的地理位置信息。

23. 根据权利要求19所述的装置,其特征在于,所述地址识别单元具体用于:

根据所述用户在各个地理位置处的行为数据,计算各个地理位置对应的地址重要性参数,以表征各个地理位置对所述用户的重要程度;

根据所述地址重要性参数的取值,计算各个地理位置信息属于与所述用户相关的地址信息的概率;

选取概率最大的预设数量的地理位置信息,以作为与所述用户相关的地址信息。

24. 根据权利要求23所述的装置,其特征在于,所述地址重要性参数与下述至少一种类型的行为数据相关:

地理位置出现频次、地理位置出现天数、地理位置连续出现的时间间隔。

25. 根据权利要求23所述的装置,其特征在于,任一地理位置信息对应的地址重要性参数被按照下述公式进行计算:

$$W_G = \frac{TFD}{D} \times TFD \times \log_2 \left(\frac{TF}{TFD} + 1 \right) \times \frac{1}{\Delta t + 1}$$

其中,W_G为所述任一地理位置对应的地址重要性参数、TF为所述任一地理位置出现频次、TFD为所述任一地理位置出现天数、△t为所述任一地理位置连续出现的时间间隔、D为所有地理位置信息对应的总天数。

26. 根据权利要求25所述的装置,其特征在于,当所述任一地理位置存在多个连续出现的时间间隔时,若多个连续出现的时间间隔的数量不大于预设数值,△t的取值为所有时间间隔的几何平均数,若多个连续出现的时间间隔的数量大于所述预设数值,△t的取值为所有时间间隔的中位数。

27. 根据权利要求23所述的装置,其特征在于,还包括:

第一修正单元,根据各个地理位置对应的地址稳定性参数,对相应的地址重要性参数进行修正,所述地址稳定性参数用于表征所述用户的轨迹稳定程度;

其中,修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。

28. 根据权利要求27所述的装置,其特征在于,各个地理位置在各个时间段对应的地址稳定性参数采用各个地理位置所处时间段的信息熵进行表征。

29. 根据权利要求28所述的装置,其特征在于,所述信息熵包括平衡熵;所述用户在时间段T对应的平衡熵被按照下述公式进行计算:

$$H_T = \frac{1}{N+2} \sum_{i=1}^M \left[(n_i + 1) \sum_{j=n_i+2}^{N+2} \frac{1}{j} \right]$$

其中,H_T为所述任一时间段对应的平衡熵,M为所述地理位置信息的总数量,N为所述用户在所述任一时间段所在的所有地点的累计出现次数,n_i为地理位置G_i的出现次数。

30. 根据权利要求23所述的装置,其特征在于,还包括:

第二修正单元,根据各个时间段对应的时间重要性参数,对相应的地址重要性参数进行修正,所述时间重要性参数用于表征各个时间段对应的重要程度;

其中,修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。

31. 根据权利要求23所述的装置,其特征在于,还包括:

第三修正单元,根据各个地理位置对应的自环情况参数,对相应的地址重要性参数进行修正,所述自环情况参数用于表征所述用户在同一周期的起始时间段与终止时间段处于同一地理位置的状况;

其中,修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。

32. 根据权利要求31所述的装置,其特征在于,任一地理位置对应的自环情况参数被按照下述公式进行计算:

$$P_G = \lg \left[\frac{2 \times A^2}{B \times 0.5 + C \times 0.5} + 1 \right] + 1$$

其中, P_G 为所述任一地理位置对应的自环情况参数、A为一天内针对所述任一地理位置发生自环的次数、B为一天内从所述任一地理位置出发的次数、C为一天内到达所述任一地理位置的次数。

33. 根据权利要求19所述的装置,其特征在于,通过下述公式计算任一地理位置信息G属于与所述用户相关的地址信息的概率:

$$I_G = \sum_{T=0}^Q W_T \times (1 + a^{H_T}) \times P_G \times W_G \times \chi_{GT}$$

其中, I_G 为任一地理位置信息G属于与所述用户相关的地址信息的概率、所述Q为时间段的数量、 W_T 为所述任一地理位置信息G在时间段T对应的时间重要性参数、 H_T 为所述任一地理位置信息G在时间段T对应的地址稳定性、 P_G 为所述任一地理位置信息G对应的自环情况参数、 W_G 为所述任一地理位置对应的地址重要性参数、a为可调系数且 $a \in (0, 1)$ 、 χ_{GT} 为指示函数,其中当所述任一地理位置信息G出现在时间段T时 $\chi_{GT}=1$,当所述任一地理位置信息G未出现在时间段T时 $\chi_{GT}=0$ 。

34. 根据权利要求19所述的装置,其特征在于,与所述用户相关的地址信息包括一对地址信息,所述一对地址信息包括属于第一类型的第一地址信息、属于第二类型的第二地址信息;其中,所述一对地址信息符合下述可靠性验证规则中至少之一:

间隔距离验证规则,用于验证所述一对地址信息之间的间隔距离;其中,所述一对地址信息之间的间隔距离不大于预设距离;

跳转概率验证规则,用于验证所述一对地址信息之间的跳转概率;其中,所述一对地址信息之间的跳转概率大于预设概率。

35. 根据权利要求34所述的装置,其特征在于,所述一对地址信息被采用下述公式计算可靠性得分,以表征对所述可靠性验证规则的符合程度:

$$\begin{cases} S_{pair} = (1 + \beta^2) \times \frac{I_{Gw} \times I_{Gf}}{(\beta^2 \times I_{Gw}) + I_{Gf}} \times \frac{D_{pair}}{100} \times (1 + P_{trans}) \\ D_{pair} = -0.01 \times d^2 + 100 \\ P_{trans} = \frac{K_{A-B}}{K_{A-ALL}} \end{cases}$$

其中, S_{pair} 为一对地理位置对应的可靠性得分、 I_{Gw} 为所述第一地址信息和所述第二地

址信息中的任一地址被识别为与所述用户相关的地址信息的概率分值、 I_{Gf} 为所述第一地址信息和所述第二地址信息中的另一地址被识别为与所述用户相关的地址信息的概率分值、 β 为可调系数且 $\beta \in (0, 1)$ 、 D_{pair} 为应用于所述间隔距离验证规则的距离衰减函数、 d 为所述第一地址信息和所述第二地址信息之间的间隔距离、 P_{trans} 为所述第一地址信息对所述第二地址信息的跳转概率、 K_{A-B} 为所述第一地址信息对所述第二地址信息的跳转次数、 K_{A-ALL} 为所述第一地址信息对所有地理位置的跳转次数。

36. 根据权利要求19所述的装置，其特征在于，还包括：

画像生成单元，根据与所述用户相关的地址信息，以及与所述地址信息相关的特定信息，生成所述用户的画像数据。

37. 一种用户地址的确定装置，其特征在于，包括：

处理器；

用于存储处理器可执行指令的存储器；

其中，所述处理器被配置为实现如权利要求1-18中任一项所述的用户地址的确定方法。

用户地址的确定方法及装置

技术领域

[0001] 本说明书涉及终端技术领域,尤其涉及一种用户地址的确定方法及装置。

背景技术

[0002] 随着快速的科技进步,智能化的发展使得用户生活的便利性不断提升,而用户对智能化生活的需求也越来越高。为了能够适应于用户需求,需要获知更多的用户信息,比如通过获知用户的家庭住址、工作地址等,以智能化地推送附近的餐厅信息、交通状况等。

发明内容

[0003] 有鉴于此,本说明书提供一种用户地址的确定方法及装置。

[0004] 为实现上述目的,本说明书提供技术方案如下:

[0005] 根据本说明书的第一方面,提出了一种用户地址的确定方法,包括:

[0006] 获取用户的地理位置信息;

[0007] 根据所述用户在各个地理位置处的行为数据,从所述地理位置信息中识别出与所述用户相关的地址信息。

[0008] 根据本说明书的第二方面,提出了一种用户地址的确定装置,包括:

[0009] 位置获取单元,获取用户的地理位置信息;

[0010] 地址识别单元,根据所述用户在各个地理位置处的行为数据,从所述地理位置信息中识别出与所述用户相关的地址信息。

[0011] 根据本说明书的第三方面,提出了一种用户地址的确定装置,包括:

[0012] 处理器;

[0013] 用于存储处理器可执行指令的存储器;

[0014] 其中,所述处理器被配置为实现如权利要求1-18中任一项所述的用户地址的确定方法。

附图说明

[0015] 图1是本申请一示例性实施例提供的一种用户地址的确定系统的架构示意图。

[0016] 图2是本说明书一示例性实施例提供的一种用户地址的确定方法的流程图。

[0017] 图3是本说明书一示例性实施例提供的一种通过地理位置生成用户画像的流程图。

[0018] 图4是本说明书一示例性实施例提供的一种通过地址识别模型实施地址识别操作的流程图。

[0019] 图5是本说明书一示例性实施例提供的一种电子设备的结构示意图。

[0020] 图6是本说明书一示例性实施例提供的一种用户地址的确定装置的框图。

具体实施方式

[0021] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本说明书一个或多个实施例相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本说明书一个或多个实施例的一些方面相一致的装置和方法的例子。

[0022] 需要说明的是:在其他实施例中并不一定按照本说明书示出和描述的顺序来执行相应方法的步骤。在一些其他实施例中,其方法所包括的步骤可以比本说明书所描述的更多或更少。此外,本说明书中所描述的单个步骤,在其他实施例中可能被分解为多个步骤进行描述;而本说明书中所描述的多个步骤,在其他实施例中也可能被合并为单个步骤进行描述。

[0023] 本说明书通过获取用户的地理位置信息,可以基于用户在各个地理位置处的行为数据,准确地分析出与用户相关的地址信息,比如家庭住址、工作地址等,而无需用户对自身的相关地址进行手动填写或标注,也无需获取用户的物流地址数据。

[0024] 图1是本申请一示例性实施例提供的一种用户地址的确定系统的架构示意图。如图1所示,该系统可以包括服务器11、网络12、若干电子设备,比如手机13、PC14等。

[0025] 服务器11可以为包含一独立主机的物理服务器,或者该服务器11可以为主机集群承载的虚拟服务器,或者该服务器11可以为云服务器。在运行过程中,服务器11可以运行某一应用的服务器侧的程序,以实现该应用的相关业务功能,比如确定用户地址等。

[0026] 手机13、PC14均为用户可以使用的一种类型的电子设备。实际上,用户显然还可以使用诸如下述类型的电子设备:平板设备、笔记本电脑、掌上电脑(PDAs, Personal Digital Assistants)、可穿戴设备(如智能眼镜、智能手表等)等,本申请并不对此进行限制。在运行过程中,该电子设备可以运行某一应用的客户端侧的程序,以实现该应用的相关业务功能,比如上传用户的地理位置信息等。

[0027] 而对于手机13、PC14与服务器11之间进行交互的网络12,可以包括多种类型的有线或无线网络。在一实施例中,该网络12可以包括公共交换电话网络(Public Switched Telephone Network, PSTN)和因特网。

[0028] 因此,本申请的用户地址的确定方案可以应用于图1所示的实施例中,手机13或PC14等提供用户的地理位置信息,并由服务器11实现相应的地址确定操作。当然,需要指出的是:在一些情况下,用户的地理位置信息也可能并不需要由手机13、PC14等提供,而由服务器11通过其他方式获得;以及,在一些情况下,还可能由手机13、PC14等对用户的地理位置信息实施地址确定操作,而无需服务器11的配合;或者,还可能存在其他情况,但这些情况显然都属于对图1所示实施例的合理调整或变形,均属于本领域技术人员能够理解的关联方案,应当被包含于本申请的保护范围内。

[0029] 下面结合实施例,对本申请的用户地址的确定方案进行说明。

[0030] 图2是本申请一示例性实施例提供的一种用户地址的确定方法的流程图。如图2所示,该方法可以应用于诸如图1所示的服务器11、手机13或PC14等各类设备上,可以包括以下步骤:

[0031] 步骤202,获取用户的地理位置信息。

[0032] 在一实施例中,可以获取所述用户所处位置的经纬度数据,并将所述经纬度数据

逆地址解析为相应的兴趣点信息,以将所述兴趣点信息作为所述地理位置信息。通过对经
纬度数据进行逆地址解析为兴趣点信息,对将该兴趣点信息的结构化难度低,甚至可以直
接生成为结构化的兴趣点信息,有助于在后续过程中提升确定出的与用户相关的地址信息
的准确度。

[0033] 例如,可以向地图服务提供商的服务器发起逆地址解析请求;然后接收所述服务
器返回的解析结果,所述解析结果中包含所述经纬度数据对应的兴趣点信息。在其他实施
例中,也可以通过其他方式对经纬度数据进行逆地址解析为兴趣点信息,本说明书并不对
此进行限制。

[0034] 在一实施例中,所述地理位置信息包括:所述用户在预设时间窗口内的地理位
置信息。例如,该预设时间窗口可以包括一天、一周、一个月、一季度、一年等,本说明
书并不对此进行限制;通常而言,通过扩大预设时间窗口的时间跨度,可以增加获得的地理
位置信息的数据量,从而提升确定出的与用户相关的地址信息的准确度,当然本说明
书并不对此进行限制。

[0035] 在一实施例中,当用户向使用的电子设备中的某一应用程序授予对地理位置信
息的获取权限时,该应用程序即可通过调用电子设备中的定位模块(如GPS定位模块、北斗定
位模块等)、调用基站定位数据等方式,确定该用户所处的地理位置信息。出于对个人隐私
信息的保护意识,用户通常不愿意提供自身的物流地址数据,但与物流地址数据等信息相
比,用户通常具有更大概率向应用程序授予对地理位置信息的获取权限,因而能够获得大
量的地理位置信息、以用于确定与用户相关的地址信息,以确保该地址信息的准确度和有
效性。

[0036] 步骤204,根据所述用户在各个地理位置处的行为数据,从所述地理位置信息中识
别出与所述用户相关的地址信息。

[0037] 在一实施例中,可以根据所述用户在各个地理位置处的行为数据,计算各个地理
位置对应的地址重要性参数,以表征各个地理位置对所述用户的重要程度;然后,根据所述
地址重要性参数的取值,计算各个地理位置信息属于与所述用户相关的地址信息的概率,
从而选取概率最大的预设数量的地理位置信息,以作为与所述用户相关的地址信息。虽然
用户可能处于众多地理位置处,家庭住址、工作地址等地理位置的重要性程度显然远高于
公交站、商场、餐厅等地理位置,而用户在这些重要性更高的地理位置往往具有特定的行
为数据(可由上述的地址重要性参数予以体现),可以据此识别出与用户相关的地址信息。

[0038] 在一实施例中,所述地址重要性参数与下述至少一种类型的行为数据相关:地理
位置出现频次、地理位置出现天数、地理位置连续出现的时间间隔等,本说明书并不对此进
行限制。

[0039] 在一实施例中,任一地理位置信息对应的地址重要性参数可以被按照下述公式进
行计算:
$$W_G = \frac{TFD}{D} \times TFD \times \log_2\left(\frac{TF}{TFD} + 1\right) \times \frac{1}{\Delta t + 1}$$
。其中,W_G为所述任一地理位置对应的地址
重要性参数、TF为所述任一地理位置出现频次、TFD为所述任一地理位置出现天数、△t为所
述任一地理位置连续出现的时间间隔、D为所有地理位置信息对应的总天数。

[0040] 在一实施例中,当所述任一地理位置存在多个连续出现的时间间隔时,若多个连
续出现的时间间隔的数量不大于预设数值,△t的取值为所有时间间隔的几何平均数,若多

个连续出现的时间间隔的数量大于所述预设数值, Δt 的取值为所有时间间隔的中位数。例如, 该预设数值可以取值为 10、15、18、22、25、30、32 等, 可以根据实际情况进行选择和调整, 本说明书并不对此进行限制。

[0041] 在一实施例中, 可以根据各个地理位置对应的地址稳定性参数, 对相应的地址重要性参数进行修正, 所述地址稳定性参数用于表征所述用户的轨迹稳定程度; 其中, 修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。例如, 用户可能由于交通状况、工作内容等原因, 在不同日期驾驶时采用不同的上班线路, 但用户总是从家庭住址出发、到达工作地址, 因而在家庭住址、工作地址附近的地址稳定性远高于路途中经过的公交站、餐厅、商场等地理位置, 所以可以据此对地址重要性参数进行修正, 以提升确定出的与用户相关的地址信息的准确度。

[0042] 在一实施例中, 各个地理位置在各个时间段对应的地址稳定性参数采用各个地理位置所处时间段的信息熵进行表征。

[0043] 在一实施例中, 上述的信息熵可以包括平衡熵。通过采用平衡熵来表征地址稳定性参数, 使得即便地理位置信息在某一时间段内的出现次数较少, 仍然能够准确地刻画出该地理位置信息的地址稳定性, 以提升确定出的与用户相关的地址信息的准确度。

[0044] 在一实施例中, 所述用户在时间段 T 对应的平衡熵被按照下述公式进行计算:

$$H_T = \frac{1}{N+2} \sum_{i=1}^M \left[(n_i + 1) \sum_{j=n_i+2}^{N+2} \frac{1}{j} \right] \text{。其中, } H_T \text{ 为所述任一时间段对应的平衡熵、} M \text{ 为所述地理位置}$$

信息的总数量、N 为所述用户在所述任一时间段所在的所有地点的累计出现次数、 n_i 为地理位置 G_i 的出现次数。

[0045] 在一实施例中, 可以根据各个时间段对应的时间重要性参数, 对相应的地址重要性参数进行修正, 所述时间重要性参数用于表征各个时间段对应的重要程度; 其中, 修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。用户所处的地理位置往往具有一定的时间规律, 比如用户通常在早上从家庭住址出发前往工作地址、晚上从工作地址出发返回家庭住址, 而中午通常不会从工作地址返回家中, 因而在识别用户的家庭住址、工作地址时, 早上、晚上的时间段的重要性通常高于中午的时间段, 可以据此对不同时间段对应的时间重要性参数赋予不同的取值, 即不同时间段对应的权重。

[0046] 在一实施例中, 可以根据各个地理位置对应的自环情况参数, 对相应的地址重要性参数进行修正, 所述自环情况参数用于表征所述用户在同一周期的起始时间段与终止时间段处于同一地理位置的状况; 其中, 修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。比如用户通常在早上从家庭住址出发前往工作地址、晚上从工作地址出发返回家庭住址, 使得形成了“家庭住址-家庭住址”的自环情况, 可以据此提升对家庭住址的识别准确度。

[0047] 在一实施例中, 任一地理位置对应的自环情况参数被按照下述公式进行计算:

$$P_G = \lg \left[\frac{2 \times A^2}{B \times 0.5 + C \times 0.5} + 1 \right] + 1 \text{。其中, } P_G \text{ 为所述任一地理位置对应的自环情况参数、} A \text{ 为一天}$$

内针对所述任一地理位置发生自环的次数、B 为一天内从所述任一地理位置出发的次数、C

为一天内到达所述任一地理位置的次数。

[0048] 在一实施例中,可以通过下述公式计算任一地理位置信息G属于与所述用户相关的地址信息的概率: $I_G = \sum_{T=0}^Q W_T \times (1+a^{H_T}) \times P_G \times W_G \times \chi_{GT}$ 。其中, I_G 为任一地理位置信息G属于与所述用户相关的地址信息的概率、所述Q为时间段的数量、 W_T 为所述任一地理位置信息G在时间段T对应的时间重要性参数、 H_T 为所述任一地理位置信息G在时间段T对应的地址稳定性、 P_G 为所述任一地理位置信息G对应的自环情况参数、 W_G 为所述任一地理位置对应的地址重要性参数、 a 为可调系数且 $a \in (0, 1)$ 、 χ_{GT} 为指示函数,其中当所述任一地理位置信息G出现在时间段T时 $\chi_{GT}=1$,当所述任一地理位置信息G未出现在时间段T时 $\chi_{GT}=0$ 。

[0049] 在一实施例中,与所述用户相关的地址信息包括一对地址信息,所述一对地址信息包括属于第一类型的第一地址信息、属于第二类型的第二地址信息;其中,所述一对地址信息符合下述可靠性验证规则中至少之一:

[0050] 间隔距离验证规则,用于验证所述一对地址信息之间的间隔距离;其中,所述一对地址信息之间的间隔距离不大于预设距离。例如,在验证一对地址信息是否为用户的家庭住址和工作地址时,由于用户的工作地址通常在家庭住址的一定范围内,比如该范围通常不会大于30km,因而如果这一对地址信息的间隔距离过大,那么这一对地址信息属于用户的家庭住址和工作地址的概率会随之降低。

[0051] 跳转概率验证规则,用于验证所述一对地址信息之间的跳转概率;其中,所述一对地址信息之间的跳转概率大于预设概率。一对地址信息之间的跳转概率即用户从第一地址(对应于第一地址信息)前往第二地址(对应于第二地址信息)或从第二地址前往第一地址的概率,比如家庭住址和工作地址对应的跳转概率通常较大,因而可以据此提升对家庭住址和工作地址的识别准确度。

[0052] 在一实施例中,所述一对地址信息被采用下述公式计算可靠性得分,以表征对所述可靠性验证规则的符合程度:

$$[0053] \left\{ \begin{array}{l} S_{pair} = (1+\beta^2) \times \frac{I_{Gw} \times I_{Gf}}{(\beta^2 \times I_{Gw}) + I_{Gf}} \times \frac{D_{pair}}{100} \times (1 + P_{trans}) \\ D_{pair} = -0.01 \times d^2 + 100 \\ P_{trans} = \frac{K_{A-B}}{K_{A-ALL}} \end{array} \right.$$

[0054] 其中, S_{pair} 为一对地理位置对应的可靠性得分、 I_{Gw} 为所述第一地址信息和所述第二地址信息中的任一地址被识别为与所述用户相关的地址信息的概率分值、 I_{Gf} 为所述第一地址信息和所述第二地址信息中的另一地址被识别为与所述用户相关的地址信息的概率分值、 β 为可调系数且 $\beta \in (0, 1)$ 、 D_{pair} 为应用于所述间隔距离验证规则的距离衰减函数、 d 为所述第一地址信息和所述第二地址信息之间的间隔距离、 P_{trans} 为所述第一地址信息对所述第二地址信息的跳转概率、 K_{A-B} 为所述第一地址信息对所述第二地址信息的跳转次数、 K_{A-ALL} 为所述第一地址信息对所有地理位置的跳转次数。

[0055] 在一实施例中,可以根据与所述用户相关的地址信息,以及与所述地址信息相关的特定信息,生成所述用户的画像数据。其中,与地址信息相关的特定信息可以包括:地址

信息对应楼盘的价格、地址信息对应地点的繁华程度、地址信息对应地点的企业类型等，本说明书并不对此进行限制。例如，当采用楼盘价格时，如果楼盘价格较高，说明用户可能具有较高收入，可以将用户划分至高净值类型，以便于针对该用户实施精准的营销推送；或者，当采用企业类型时，如果企业类型属于医院、公司等，说明用户的职业可能为医生、白领等，可以据此评估用户的信贷风险等。

[0056] 为了便于理解，下面以记账应用为例，对本说明书的技术方案进行说明。假定手机13或PC14上运行有记账应用客户端、服务器11上运行有记账应用服务端，其中手机13或PC14上的记账应用客户端登录有用户的注册账号，使得手机13或PC14可以在接收用户输入的记账信息时，获取用户对应的地理位置信息，并基于该注册账号与服务器11进行数据交互，使得服务器11可以获取该用户的地理位置信息，并据此分析并确定出与该用户相关的地址信息。例如，下文通过获取用户的地理位置信息，分析确定出该用户的家庭住址和工作地址。

[0057] 图3是本说明书一示例性实施例提供的一种通过地理位置生成用户画像的流程图。假定该流程应用于图1所示的服务器11，如图3所示，该流程可以包括以下步骤：

[0058] 步骤302，获取用户的历史经纬度信息。

[0059] 在一实施例中，通过本说明书的技术方案生成任一用户的用户画像时，可以通过获取该任一用户的历史经纬度信息，以经由后续处理得到该任一用户的用户画像。其中，历史经纬度信息包括该任一用户在一历史时间窗口内所处的历史地理位置的经纬度信息，比如该历史时间窗口可以包括最近3个月、最近半年、去年2月份至8月份等，本说明书并不限制其具体形式。

[0060] 在一实施例中，通过运行于用户使用的电子设备上的应用程序，可以在该应用程序具有对经纬度信息的获取权限时，获取该电子设备所处的经纬度信息，并上传至服务器11。由于用户通常随身携带该电子设备，因而可以将该电子设备所处的经纬度信息作为该用户所处的经纬度信息。例如，电子设备可以内置有GPS模块或北斗模块等定位模块，应用程序可以调用该定位模块，以获取该定位模块检测到的经纬度信息。

[0061] 举例而言，单条经纬度信息可以如下表1所示，可以通过用户ID、创建时刻、经度、纬度等多个维度对该经纬度信息进行定义。

[0062]	用户 ID	172139127
	创建时刻	2017-06-19 12:30:10
[0063]	经度	109.19602
	纬度	27.71329

[0064] 表1

[0065] 对于每个用户而言，历史经纬度信息可以包括若干条如表1所示的经纬度信息。仍以该用户ID为172139127的用户为例，其历史经纬度信息的一部分数据可以如下表2所示：

[0066]

用户ID	创建时刻	经度	纬度
172139127	2017-04-16 10:51:13	111.332257	30.67374
172139127	2017-05-20 15:10:45	119.773359	32.418112

172139127	2017-05-30 19:11:14	119.124364	34.907897
172139127	2017-06-12 00:02:38	120.175071	32.348272
172139127	2017-06-15 00:02:51	115.632143	39.022728
172139127	2017-06-19 12:30:10	109.19602	27.71329
...

[0067] 表2

[0068] 步骤304,将历史经纬度信息转换为历史POI (Point of Interest) 信息。

[0069] 在一实施例中,服务器11可以对历史经纬度信息进行逆地址解析,以确定出每一历史经纬度信息分别对应的历史POI信息。一种情况下,服务器11可以自主维护地图数据,以通过该地图数据实现对历史经纬度信息的转换操作,得到相应的历史POI信息。另一种情况下,服务器11可以通过第三方实现对历史经纬度信息的转换操作,例如该第三方可以包括地图服务提供商的服务器,服务器11可以通过向第三方发起针对历史经纬度信息的转换请求,并由第三方返回转换后的历史POI信息。其中,服务器11可以在收到电子设备上传的经纬度信息后,就将其转换为相应的POI信息,或者服务器11可以在需要对相应的用户实施地址识别时,才将存储的历史经纬度信息转换为相应的历史POI信息,本说明书并不限制实施该转换操作的时刻。

[0070] 举例而言,通过对上述表1所示的经纬度信息进行逆地址解析后,可以得到如下表3所示的POI信息,可以通过经度、纬度、地址、国家、省份、城市、区、离经纬度最近的建筑物名称、建筑物类别、建筑物离经纬度的距离等多个维度对该POI信息进行定义。

[0071]

经度	109.19602
纬度	27.71329
地址	碧江区人民路西铜仁市第一人民医院东
国家	中国
省份	贵州省
城市	铜仁市
区	碧江区
离经纬度最近的建筑物名称	铜仁市第一人民医院
建筑物类别	医疗保健-综合医院
建筑物离经纬度的距离	53米

[0072] 表3

[0073] 在其他实施例中,用户使用的电子设备可以直接向服务器11上传POI信息,而无需服务器11实施上述的逆地址解析操作。

[0074] 在一实施例中,通过将历史经纬度信息通过逆地址解析转换为相应的历史POI信息,可以确保该历史POI信息的完整性和结构化,便于后续操作中基于该历史POI信息实施地址识别操作,从而既不需要用户手动填写地址信息、简化用户操作,又可以避免用户手动填写时存在填写错误、填写内容不规范等情况而导致的难以结构化的问题。

[0075] 步骤306,通过预先定义的地址识别模型,对历史POI信息进行处理,以确定出用户的家庭住址和工作地址等地址信息。

[0076] 在一实施例中,可以参照下述结合图4描述的地址识别过程,将对基于该地址识别模型的地址识别操作进行详细描述,此处暂不赘述。

[0077] 步骤308,根据识别出的家庭住址和工作地址等,生成相应的用户画像。

[0078] 在一实施例中,通过获取家庭住址、工作地址等的相关信息,可以确定出用户的相关属性信息。例如,可以获取家庭住址对应小区的房价信息;再例如,可以获取工作地址对应建筑物的类别,并根据该建筑物类别推断用户的职业,以根据房价信息、职业等推断用户的经济水平、偿贷能力等,从而生成相应的用户画像。譬如,当用户的家庭住址属于一房价较高的高档小区、用户的职业属于收入较高的医生或企业白领时,可以推断该用户的经济水平较高、偿贷能力较好,从而将该用户划归至高净值人群。

[0079] 图4是本说明书一示例性实施例提供的一种通过地址识别模型实施地址识别操作的流程图。假定该流程应用于图1所示的服务器11,如图4所示,该流程可以包括以下步骤:

[0080] 步骤402,对历史POI信息按时序排列。

[0081] 在一实施例中,假定基于用户X的历史POI信息,对该用户X的家庭住址和工作地址进行识别。初始状态的历史POI信息可以为乱序或按照非时序顺序排列,比如该初始状态的历史POI信息可以包括:[uid:Gt₃,Gt₅,Gt₂,...,Gt_n,Gt_{n-1}],Gt_i表示在t_i时刻对应的历史POI信息,其中1≤i≤n。

[0082] 在对上述初始状态的历史POI信息按时序顺序进行排列后,可以得到:[uid:Gt₁,Gt₂,Gt₃,...,Gt_{n-1},Gt_n],以便于后续实施与时序相关的操作。

[0083] 步骤404,按天为粒度进行分组。

[0084] 在一实施例中,上述的历史POI信息可以来自于用户X在一历史时间窗口内的POI信息,比如该历史时间窗口可以为最近3个月或其他任意时间窗口。

[0085] 在一实施例中,通过获取Gt_i对应的创建时刻(比如上述表1所示的“创建时刻”信息),可以确定其对应的日期,从而将其划分至该日期对应的分组。例如,上述的历史POI信息可以被划分为:[uid:Day1(Gt₁,Gt₂),Day2(Gt₃,Gt₄,Gt₅)...DayM(Gt_{n-1},Gt_n)],其中Day1(Gt₁,Gt₂)表示第一天包括Gt₁、Gt₂对应的两条POI信息,Day2(Gt₃,Gt₄,Gt₅)表示第二天包括Gt₃、Gt₄和Gt₅对应的三条POI信息,DayM(Gt_{n-1},Gt_n)表示第M天包括Gt_{n-1}、Gt_n对应的两条POI信息。

[0086] 在其他实施例中,可以采用其他粒度进行分组,比如按小时、按周等,本说明书并不对此进行限制。

[0087] 步骤406,计算历史POI信息对应的各个地理位置的地址重要性W_G。

[0088] 在一实施例中,虽然服务器11获得了针对用户X的n条历史POI信息,但是这些历史POI信息中可能存在部分重复地理位置,比如总共可以统计出m个地理位置,且m≤n;然后,可以针对每一地理位置G_i(1≤i≤m)计算相应的地址重要性W_{Gi},以表征该地理位置G_i对用户X的重要程度。

[0089] 在一实施例中,可以根据上述的历史POI信息计算出与地址重要性W_G相关的用户行为数据,以根据该用户行为数据确定出地址重要性W_G的取值。例如,该用户行为数据可以包括:地理位置出现频次TF、地理位置出现天数TFD、地理位置连续出现的时间间隔△t等,本说明书并不对此进行限制。

[0090] 地理位置出现频次TF,用于表征每一地理位置在上述历史POI信息中出现的次数;

比如, TF_{Gi} 表示地理位置 Gi 在上述历史POI信息中出现的次数, 即该历史POI信息中与该地理位置 Gi 相关的POI信息的条数。因此, 可以分别计算出上述 m 个地理位置对应的地理位置出现频次, 得到: $TF_{G1}, TF_{G2}, TF_{G3} \dots TF_{Gm-1}, TF_{Gm}$ 。在不同实施例中, TF可以采用不同的时间粒度, 例如, 在一实施例中可以采用“全量”粒度, 则 TF_{Gi} 表示上述的“在上述历史POI信息中出现的次数”; 在另一实施例中, 可以采用“时间段”粒度, 则 TF_{Gi} 表示地理位置 Gi 在某一时间段(关于时间段的划分, 可参考下述步骤408)中出现的次数; 在又一实施例中, 可以采用“天”粒度, 则 TF_{Gi} 表示地理位置 Gi 在某天中出现的次数等, 此处不再一一赘述。此外, 同一实施例中还可以同时采用多个时间粒度, 比如同时采用“全量”粒度和“时间段”粒度, 并同时用于确定地址重要性 W_G 。

[0091] 地理位置出现天数 TFD , 用于表征每一地理位置在上述历史POI信息中出现的天数, 比如 TFD_{Gi} 表示地理位置 Gi 在上述历史POI信息中出现的天数, 其中每一地理位置在同一天中可能多次出现, 但在地理位置出现天数 TFD 中仅记录为一天。因此, 可以分别计算出上述 m 个地理位置对应的地理位置出现天数, 得到: $TFD_{G1}, TFD_{G2}, TFD_{G3} \dots TFD_{Gm-1}, TFD_{Gm}$ 。

[0092] 地理位置连续出现的时间间隔 Δt , 该指标以天为粒度确定每一地理位置出现的日期, 并用于表征每一地理位置在上述历史POI信息中连续出现的日期之间的时间间隔。例如, 假定用户X在星期一在地理位置 $G2$ 出现2次, 星期三在地理位置 $G2$ 出现3次, 星期四在地理位置 $G2$ 出现2次, 星期日在地理位置 $G2$ 出现1次, 星期二、星期五、星期六均没有出现在地理位置 $G2$, 那么该地理位置 $G2$ 连续在星期一和星期三出现、对应的时间间隔为1(即间隔了星期二), 该地理位置 $G2$ 连续在星期三和星期四出现、对应的时间间隔为0, 该地理位置 $G2$ 连续在星期四和星期日出现、对应的时间间隔为2(即间隔了星期五和星期六), 因而该地理位置 $G2$ 对应的时间间隔数组可以为 $G2[\Delta t_{1-G2}, \Delta t_{2-G2}, \Delta t_{3-G2}]$, 其中 $\Delta t_{1-G2}=1, \Delta t_{2-G2}=0, \Delta t_{3-G2}=2$ 。类似地, 可以分别计算出上述 m 个地理位置对应的时间间隔数组: $G1[\Delta t_{1-G1}, \Delta t_{2-G1}], G2[\Delta t_{1-G2}, \Delta t_{2-G2}, \Delta t_{3-G2}] \dots Gm[\Delta t_{1-Gm}, \Delta t_{2-Gm}]$ 。

[0093] 其中, 还需要进一步根据每一地理位置对应的时间间隔数组, 计算出相应的地理位置连续出现的时间间隔, 比如 Δt_{Gi} 表示地理位置 Gi 对应的地理位置连续出现的时间间隔。其中, 如果地理位置 Gi 对应的时间间隔数组仅包含单个时间间隔, 比如 $Gi[\Delta t_{1-Gi}]$, 可使 $\Delta t_{Gi}=\Delta t_{1-Gi}$; 如果地理位置 Gi 对应的时间间隔数组包含多个时间间隔, 可以按照下述方式计算: 若多个连续出现的时间间隔的数量不大于预设数值, 取值为所有时间间隔的几何平均数, 从而在数据量较小的情况下仍然能够确保取值的准确度; 若多个连续出现的时间间隔的数量大于所述预设数值, 取值为所有时间间隔的中位数。

[0094] 在一实施例中, 可以按照下述公式计算上述 m 个地理位置分别对应的地址重要性(其中, D 表示上述历史POI信息对应的总天数):

$$[0095] W_G = \frac{TFD}{D} \times TFD \times \log_2 \left(\frac{TF}{TFD} + 1 \right) \times \frac{1}{\Delta t + 1}$$

[0096] 例如, 对于地理位置 Gi , 相应的地址重要性的取值可以为:

$$W_{Gi} = \frac{TFD_{Gi}}{D} \times TFD_{Gi} \times \log_2 \left(\frac{TF_{Gi}}{TFD_{Gi}} + 1 \right) \times \frac{1}{\Delta t_{Gi} + 1}.$$

[0097] 步骤408, 设定不同时间段对应的时间重要性 W_T , 以用于对上述的地址重要性 W_G 进行修正。

[0098] 在一实施例中,可以将上述步骤404获得的按天为粒度进行分组得到的[uid:Day1(Gt₁,Gt₂),Day2(Gt₃,Gt₄,Gt₅)…DayM(Gt_{n-1},Gt_n)] ,进一步在每天按时间段对相应的地理位置进行划分。

[0099] 在一实施例中,可以预先将每天划分出若干时间段,比如上午(morn)[9:00,11:30]、中午(noon)[11:30,14:00]、下午(afno)[14:00,17:00]、晚餐(supp)[17:00,20:00]、夜间(nigh)[20:00,8:00]、上午过渡(motr)[8:00,9:00]等,然后针对步骤404中被划分至每一天的历史POI信息,分别将每一天的历史POI信息划分至相应的各个时间段中。

[0100] 例如,假定第一天中的Gt₁属于上午、Gt₂属于下午,第二天中的Gt₃属于上午、Gt₄属于中午、Gt₅属于下午,第M天的Gt_{n-1}属于中午、Gt_n属于晚餐,可以将步骤404获得的分组进一步处理为:[uid:(mornGt₁,afnoGt₂),(mornGt₃,noonGt₄,nighGt₅)…(noonGt_{n-1},nighGt_n)]。

[0101] 在一实施例中,可以设定各个时间段对应的时间重要性W_T,即各个时间段对于各个地理位置的权重值,比如对应于“早上”时间段的权重值W_{T-morn}、对应于“中午”时间段的权重值W_{T-noon}、对应于“夜间”时间段的权重值W_{T-nigh}。比如在识别家庭住址时,由于用户更可能在早上、夜间等时间段位于家中,而在中午、下午等时间段通常不在家,因而可以为处于早上、夜间等时间段的地理位置设置更大的时间重要性(即权重值更大),而为处于中午、下午等时间段的地理位置设置更小的时间重要性(即权重值更小)。

[0102] 步骤410,计算各个时间段的地址稳定性H_T,以用于对上述的地址重要性W_G进行修正。

[0103] 在一实施例中,基于步骤408中获得的按时间段划分的历史POI信息,可以分别计算各个时间段对应的信息熵,从而用该信息熵用于表征处于相应时间段的地理位置的地址稳定性。例如,H_{morn}表示“早上”时间段对应的信息熵、H_{nigh}表示“夜间”时间段对应的信息熵,如果地理位置G1、G3出现在“早上”时间段、地理位置G2、G5出现在“夜间”时间段,可以采用H_{morn}表征地理位置G1、G3的地址稳定性、采用H_{nigh}表征地理位置G2、G5的地址稳定性。

[0104] 在一实施例中,由于各个地理位置在各个时间段出现的频次不同,对于一些地理位置的出现频次较小(如小于预设频次)的时间段而言,为了确保最终计算出的信息熵能够准确表达其地址稳定性,可以在步骤410中采用这些时间段的平衡熵、以用于表征处于该时间段的地理位置对应的地址稳定性;当然,地理位置的出现频次较大的时间段同样可以采用平衡熵来表征。例如,平衡熵可以采用下述公式进行计算:

$$[0105] H_T = \frac{1}{N+2} \sum_{i=1}^M \left[(n_i + 1) \sum_{j=n_i+2}^{N+2} \frac{1}{j} \right]$$

[0106] 其中,在通过该公式计算“早上”时间段的平衡熵时,M=m为地理位置的总数量,N为用户X在“早上”时间段所在的所有地点的累计出现次数(比如在地理位置G1出现10次、地理位置G2出现2次,则N=10+2=12)、n_i为用户X在“早上”时间段出现在地理位置G_i的次数(比如用户X在地理位置G1出现10次,则n₁=10)。

[0107] 步骤412,计算每一地理位置的地点回归性P_G,以用于对上述的地址重要性W_G进行修正。

[0108] 在一实施例中,可以将上述步骤404获得的按天为粒度进行分组得到的[uid:Day1(Gt₁,Gt₂),Day2(Gt₃,Gt₄,Gt₅)…DayM(Gt_{n-1},Gt_n)] ,进一步将前一天的截止位置(即用户X在第一天最后所在的地理位置)与第二天的起始位置(即用户X在第二天最先所在的地理位

置)进行组合,例如可以得到[uid:1(Gt₂,Gt₃),…,(Gt_{n-2},Gt_{n-1})]。

[0109] 对于家庭住址而言,很容易出现“自环”情况,即同一天的起始位置与截止位置相同,表示用户上午从家中出发、夜间从外面返回家中。因此,在需要识别的地址包括家庭住址时,可以根据自环情况予以识别。其中,任一地理位置Gi对应的自环情况可以被按照下述公式进行计算:

$$[0110] P_G = \lg \left[\frac{2 \times A^2}{B \times 0.5 + C \times 0.5} + 1 \right] + 1$$

[0111] 其中,A为一天内针对地理位置Gi发生自环的次数、B为一天内从地理位置Gi出发的次数、C为一天内到达地理位置Gi的次数。例如,假定用户X在第j天从地理位置Gi出发的次数为2、到达地理位置Gi的次数为3,可以确定B=2、C=3、A=2,从而计算出地理位置在这一天对应的自环情况P_{Day,j-Gi}。

[0112] 步骤414,根据时间重要性、地址稳定性、地点回归性等,对步骤406得到的地址重要性进行修正。

[0113] 在一实施例中,可以通过下述公式计算任一地理位置信息G属于家庭住址或工作地址的概率:

$$[0114] I_G = \sum_{T=0}^Q W_T \times (1 + a^{H_T}) \times P_G \times W_G \times \chi_{GT}$$

[0115] 其中,Q为时间段的数量,比如在上述实施例中共设定了6个时间段,则Q=6;a为可调系数且a∈(0,1);χ_{GT}为指示函数,其中当所述任一地理位置信息G出现在时间段T时χ_{GT}=1,当所述任一地理位置信息G未出现在时间段T时χ_{GT}=0。

[0116] 在一实施例中,根据计算出的概率I_G,可以确定出相应的候选家庭住址列表和候选工作地址列表,以统计出候选家庭住址、候选工作地址。

[0117] 步骤416,对候选家庭住址、候选工作地址进行验证。

[0118] 在一实施例中,可以将候选家庭住址、候选工作地址组成“候选家庭住址-候选工作地址”候选地址对,并对候选地址对实施交叉验证,以提升输出结果的可靠性。

[0119] 在一实施例中,可以采用距离验证。由于在通常情况下,用户的工作地址处于家庭住址的一定范围内,且出于交通因素,该范围应当处于一合理数值,比如30-60km。因此,分别计算各个候选地址对中候选家庭住址与候选工作地址之间的间隔距离,而候选地址对的可靠性与该间隔距离呈负相关,即间隔距离越大可靠性越低。

[0120] 例如,可以通过一距离衰减函数D_{pair}=-0.01×d²+100来表征可靠性;其中,当计算候选地址对Gi-Gj对应的可靠性D_{pair-Gi-Gj}时,d为地理位置Gi与Gj之间的间隔距离。

[0121] 在一实施例中,可以采用地址对转移概率验证。由于用户X在正常生活过程中,在家庭住址与工作地址之间能够保持连续、稳定的位置转移,即在上班过程中从家庭住址前往工作地址、在下班过程中从工作地址前往家庭住址,因而可以在家庭住址与工作地址之间保持较高的转移概率。

[0122] 例如,可以通过下述公式计算每一候选地址对中的候选家庭住址与候选工作地址之间的跳转概率:P_{trans}=K_{A-B}/K_{A-ALL},其中A、B构成一候选地址对,K_{A-B}表示由地理位置A向地理位置B的跳转次数、K_{A-ALL}表示由地理位置A向所有地理位置的跳转次数。

[0123] 在一实施例中,结合上述的距离衰减函数和跳转概率计算公式,可以通过下述公

式计算出每一候选地址对的得分：

$$[0124] S_{pair} = (1 + \beta^2) \times \frac{I_{Gw} \times I_{Gf}}{(\beta^2 \times I_{Gw}) + I_{Gf}} \times \frac{D_{pair}}{100} \times (1 + P_{trans})$$

[0125] 其中， I_{Gw} 为候选地址对中的候选工作地址Gw在步骤414中计算得到的预测分值、 I_{Gf} 为候选地址对中的候选工作地址Gf在步骤414中计算得到的预测分值、 β 为可调系数且 $\beta \in (0, 1)$ 。

[0126] 最终，通过对各个候选地址对的得分进行排序，可将最高得分的候选地址对选择为用户X对应的地址对，该地址对中的候选家庭住址为该用户X的家庭住址、该地址对中的工作地址为该用户X的工作地址。

[0127] 图5是本说明书一示例性实施例提供的一种电子设备的示意结构图。请参考图5，在硬件层面，该电子设备包括处理器502、内部总线504、网络接口506、内存508以及非易失性存储器510，当然还可能包括其他业务所需要的硬件。处理器502从非易失性存储器510中读取对应的计算机程序到内存508中然后运行，在逻辑层面上形成用户地址的确定装置。当然，除了软件实现方式之外，本说明书并不排除其他实现方式，比如逻辑器件抑或软硬件结合的方式等等，也就是说以下处理流程的执行主体并不限定于各个逻辑单元，也可以是硬件或逻辑器件。

[0128] 请参考图6，在软件实施方式中，该用户地址的确定装置可以包括：

[0129] 位置获取单元61，获取用户的地理位置信息；

[0130] 地址识别单元62，根据所述用户在各个地理位置处的行为数据，从所述地理位置信息中识别出与所述用户相关的地址信息。

[0131] 可选的，所述位置获取单元61具体用于：

[0132] 获取所述用户所处位置的经纬度数据；

[0133] 将所述经纬度数据逆地址解析为相应的兴趣点信息，以将所述兴趣点信息作为所述地理位置信息。

[0134] 可选的，所述位置获取单元61通过下述方式将所述经纬度数据逆地址解析为相应的兴趣点：

[0135] 向地图服务提供商的服务器发起逆地址解析请求；

[0136] 接收所述服务器返回的解析结果，所述解析结果中包含所述经纬度数据对应的兴趣点信息。

[0137] 可选的，所述地理位置信息包括：所述用户在预设时间窗口内的地理位置信息。

[0138] 可选的，所述地址识别单元62具体用于：

[0139] 根据所述用户在各个地理位置处的行为数据，计算各个地理位置对应的地址重要性参数，以表征各个地理位置对所述用户的重要程度；

[0140] 根据所述地址重要性参数的取值，计算各个地理位置信息属于与所述用户相关的地址信息的概率；

[0141] 选取概率最大的预设数量的地理位置信息，以作为与所述用户相关的地址信息。

[0142] 可选的，所述地址重要性参数与下述至少一种类型的行为数据相关：

[0143] 地理位置出现频次、地理位置出现天数、地理位置连续出现的时间间隔。

[0144] 可选的，任一地理位置信息对应的地址重要性参数被按照下述公式进行计算：

$$[0145] W_G = \frac{TFD}{D} \times TFD \times \log_2 \left(\frac{TF}{TFD} + 1 \right) \times \frac{1}{\Delta t + 1}$$

[0146] 其中, W_G 为所述任一地理位置对应的地址重要性参数、TF 为所述任一地理位置出现频次、TFD 为所述任一地理位置出现天数、 Δt 为所述任一地理位置连续出现的时间间隔、D 为所有地理位置信息对应的总天数。

[0147] 可选的, 当所述任一地理位置存在多个连续出现的时间间隔时, 若多个连续出现的时间间隔的数量不大于预设数值, Δt 的取值为所有时间间隔的几何平均数, 若多个连续出现的时间间隔的数量大于所述预设数值, Δt 的取值为所有时间间隔的中位数。

[0148] 可选的, 还包括:

[0149] 第一修正单元63, 根据各个地理位置对应的地址稳定性参数, 对相应的地址重要性参数进行修正, 所述地址稳定性参数用于表征所述用户的轨迹稳定程度;

[0150] 其中, 修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。

[0151] 可选的, 各个地理位置在各个时间段对应的地址稳定性参数采用各个地理位置所处时间段的信息熵进行表征。

[0152] 可选的, 所述信息熵包括平衡熵; 所述用户在时间段T对应的平衡熵被按照下述公式进行计算:

$$[0153] H_T = \frac{1}{N+2} \sum_{i=1}^M \left[(n_i + 1) \sum_{j=n_i+2}^{N+2} \frac{1}{j} \right]$$

[0154] 其中, H_T 为所述任一时间段对应的平衡熵、M 为所述地理位置信息的总数量、N 为所述用户在所述任一时间段所在的所有地点的累计出现次数、 n_i 为地理位置 G_i 的出现次数。

[0155] 可选的, 还包括:

[0156] 第二修正单元64, 根据各个时间段对应的时间重要性参数, 对相应的地址重要性参数进行修正, 所述时间重要性参数用于表征各个时间段对应的重要程度;

[0157] 其中, 修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。

[0158] 可选的, 还包括:

[0159] 第三修正单元65, 根据各个地理位置对应的自环情况参数, 对相应的地址重要性参数进行修正, 所述自环情况参数用于表征所述用户在同一周期的起始时间段与终止时间段处于同一地理位置的状况;

[0160] 其中, 修正后的地址重要性参数被用于计算各个地理位置信息属于与所述用户相关的地址信息的概率。

[0161] 可选的, 任一地理位置对应的自环情况参数被按照下述公式进行计算:

$$[0162] P_G = \lg \left[\frac{2 \times A^2}{B \times 0.5 + C \times 0.5} + 1 \right] + 1$$

[0163] 其中, P_G 为所述任一地理位置对应的自环情况参数、A 为一天内针对所述任一地理位置发生自环的次数、B 为一天内从所述任一地理位置出发的次数、C 为一天内到达所述任一地理位置的次数。

[0164] 可选的,通过下述公式计算任一地理位置信息G属于与所述用户相关的地址信息的概率:

$$[0165] I_G = \sum_{T=0}^Q W_T \times (1 + a^{H_T}) \times P_G \times W_G \times \chi_{GT}$$

[0166] 其中, I_G 为任一地理位置信息G属于与所述用户相关的地址信息的概率、所述Q为时间段的数量、 W_T 为所述任一地理位置信息G在时间段T对应的时间重要性参数、 H_T 为所述任一地理位置信息G在时间段T对应的地址稳定性、 P_G 为所述任一地理位置信息G对应的自环情况参数、 W_G 为所述任一地理位置对应的地址重要性参数、 a 为可调系数且 $a \in (0, 1)$ 、 χ_{GT} 为指示函数,其中当所述任一地理位置信息G出现在时间段T时 $\chi_{GT}=1$,当所述任一地理位置信息G未出现在时间段T时 $\chi_{GT}=0$ 。

[0167] 可选的,与所述用户相关的地址信息包括一对地址信息,所述一对地址信息包括属于第一类型的第一地址信息、属于第二类型的第二地址信息;其中,所述一对地址信息符合下述可靠性验证规则中至少之一:

[0168] 间隔距离验证规则,用于验证所述一对地址信息之间的间隔距离;其中,所述一对地址信息之间的间隔距离不大于预设距离;

[0169] 跳转概率验证规则,用于验证所述一对地址信息之间的跳转概率;其中,所述一对地址信息之间的跳转概率大于预设概率。

[0170] 可选的,所述一对地址信息被采用下述公式计算可靠性得分,以表征对所述可靠性验证规则的符合程度:

$$[0171] \left\{ \begin{array}{l} S_{pair} = (1 + \beta^2) \times \frac{I_{Gw} \times I_{Gf}}{(\beta^2 \times I_{Gw}) + I_{Gf}} \times \frac{D_{pair}}{100} \times (1 + P_{trans}) \\ D_{pair} = -0.01 \times d^2 + 100 \\ P_{trans} = \frac{K_{A-B}}{K_{A-ALL}} \end{array} \right.$$

[0172] 其中, S_{pair} 为一对地理位置对应的可靠性得分、 I_{Gw} 为所述第一地址信息和所述第二地址信息中的任一地址被识别为与所述用户相关的地址信息的概率分值、 I_{Gf} 为所述第一地址信息和所述第二地址信息中的另一地址被识别为与所述用户相关的地址信息的概率分值、 β 为可调系数且 $\beta \in (0, 1)$ 、 D_{pair} 为应用于所述间隔距离验证规则的距离衰减函数、 d 为所述第一地址信息和所述第二地址信息之间的间隔距离、 P_{trans} 为所述第一地址信息对所述第二地址信息的跳转概率、 K_{A-B} 为所述第一地址信息对所述第二地址信息的跳转次数、 K_{A-ALL} 为所述第一地址信息对所有地理位置的跳转次数。

[0173] 可选的,还包括:

[0174] 画像生成单元66,根据与所述用户相关的地址信息,以及与所述地址信息相关的特定信息,生成所述用户的画像数据。

[0175] 上述实施例阐明的系统、装置、模块或单元,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。一种典型的实现设备为计算机,计算机的具体形式可以是个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件收发设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的

任意几种设备的组合。

[0176] 在一个典型的配置中,计算机包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0177] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0178] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0179] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,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、商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“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0180] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本说明书相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本说明书的一些方面相一致的装置和方法的例子。

[0181] 在本说明书使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本说明书。在本说明书和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。还应当理解,本文中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[0182] 应当理解,尽管在本说明书可能采用术语第一、第二、第三等来描述各种信息,但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如,在不脱离本说明书范围的情况下,第一信息也可以被称为第二信息,类似地,第二信息也可以被称为第一信息。取决于语境,如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[0183] 以上所述仅为本说明书的较佳实施例而已,并不用以限制本说明书,凡在本说明书的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本说明书保护的范围之内。

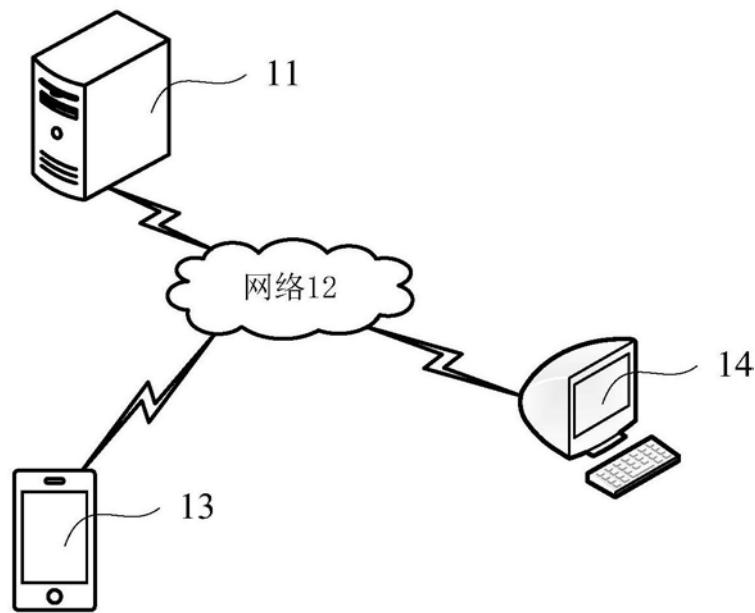


图1

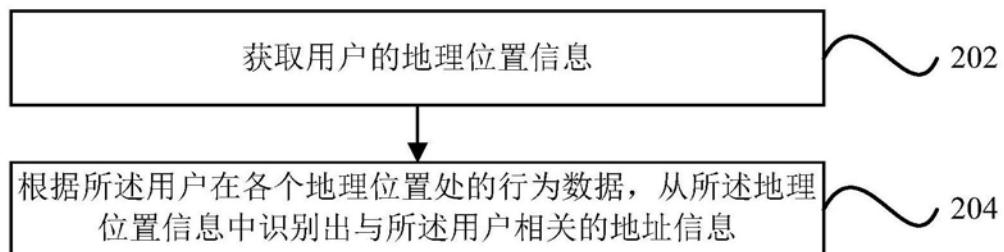


图2

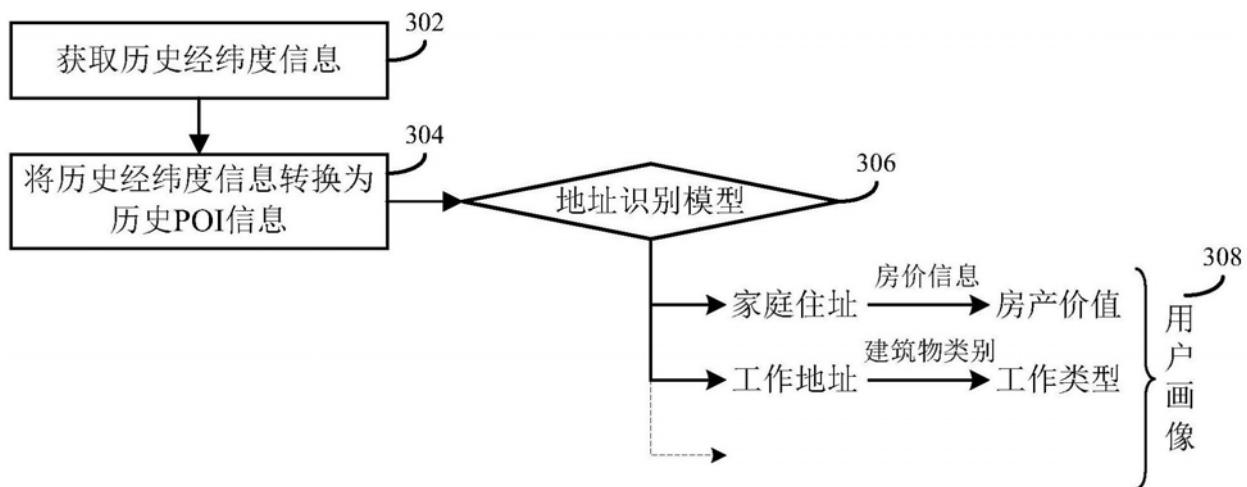


图3

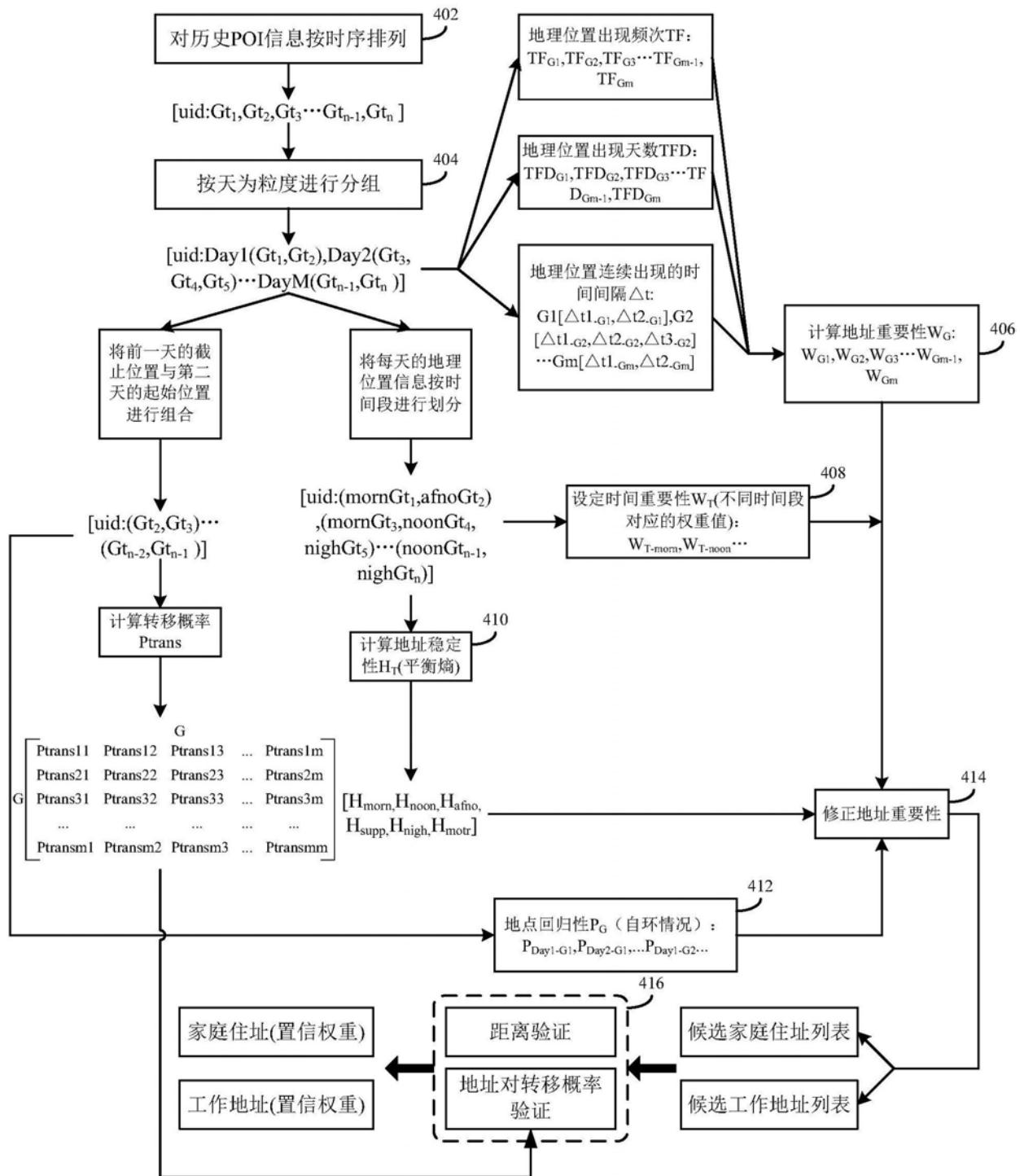


图4

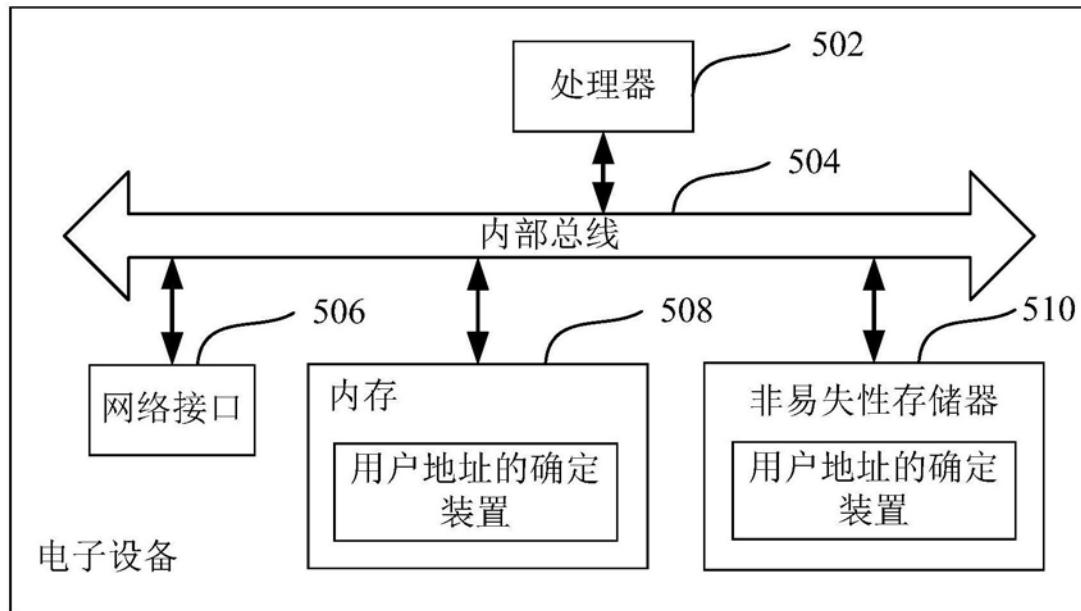


图5

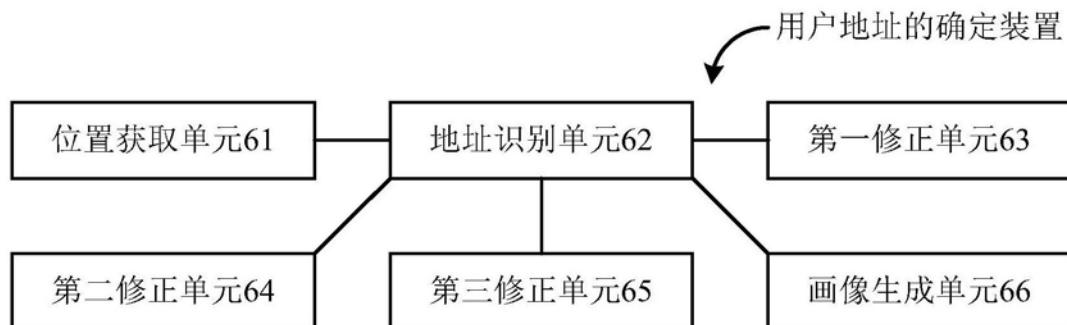


图6