



(12) 发明专利申请

(10) 申请公布号 CN 115714957 A

(43) 申请公布日 2023. 02. 24

(21) 申请号 202211359517.0

G06Q 50/26 (2012.01)

(22) 申请日 2022.11.02

G06F 16/29 (2019.01)

(71) 申请人 广州市城市规划勘测设计研究院
地址 510030 广东省广州市越秀区建设大
马路10号

G06F 16/2458 (2019.01)

G06F 16/2457 (2019.01)

(72) 发明人 邓兴栋 刘洋 李冠耀 廖顺意
张佶 王鹏 王楠

(74) 专利代理机构 广州三环专利商标代理有限
公司 44202

专利代理师 胡静

(51) Int.Cl.

H04W 4/029 (2018.01)

H04W 4/02 (2018.01)

H04W 4/20 (2018.01)

H04W 8/18 (2009.01)

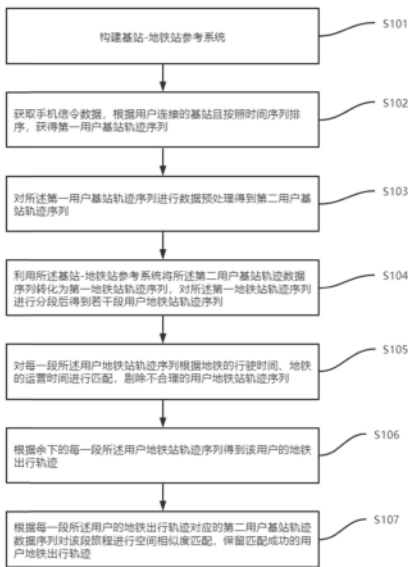
权利要求书3页 说明书9页 附图2页

(54) 发明名称

基于手机信令的地铁出行识别方法、装置、
设备及介质

(57) 摘要

本发明公开了一种基于手机信令的地铁出行识别方法,包括:获取手机信令数据,根据用户连接的基站获得用户基站轨迹序列;对用户基站轨迹序列进行数据预处理;利用预先构建的基站-地铁站参考系统将基站轨迹序列转化为地铁站轨迹序列,分段后得到若干段用户地铁站轨迹序列;根据地铁的行驶时间、运营时间剔除不合理的用户地铁站轨迹序列;根据实际地铁线路得到完整的用户地铁出行轨迹,通过地铁线路空间相似度剔除不合理的用户地铁出行轨迹。本发明还公开了一种基于手机信令的地铁出行识别装置、设备及介质。本发明可以满足大规模地铁出行识别的需求,在不依赖训练数据与第三方导航数据的情况下,实现及时、准确的大规模地铁出行轨迹识别。



1. 一种基于手机信令的地铁出行识别方法,其特征在于,包括:

构建基站-地铁站参考系统;所述基站-地铁站参考系统是根据计算地铁站与所有基站的距离,选取所述地铁站的用户能连接到的基站作为参考基站;

获取手机信令数据,记用户在时刻 t_i 连接到的基站为 C_i ,并规定 $t_i < t_{i+1}$,根据用户连接的基站且按照时间序列排序,获得第一用户基站轨迹序列;其中,所述第一用户基站轨迹序列表示为: $\{(C_1, t_1), (C_2, t_2), \dots, (C_i, t_i), \dots (C_n, t_n)\}$;其中 $C_i = (lat_i, lon_i)$ 表示一个用户 t_i 时刻连接的基站, (lat_i, lon_i) 表示基站的经纬度坐标;

对所述第一用户基站轨迹序列进行数据预处理得到第二用户基站轨迹序列;

利用所述基站-地铁站参考系统将所述第二用户基站轨迹数据序列转化为第一地铁站轨迹序列,对所述第一地铁站轨迹序列进行分段后得到若干段用户地铁站轨迹序列;

对每一段所述用户地铁站轨迹序列根据地铁的行驶时间及地铁的运营时间进行匹配,剔除不合理的用户地铁站轨迹序列;

根据余下的每一段所述用户地铁站轨迹序列得到该用户的地铁出行轨迹;

根据每一段所述用户的地铁出行轨迹对应的第二用户基站轨迹数据序列对该段旅程进行空间相似度匹配,保留匹配成功的旅程对应的用户的地铁出行轨迹。

2. 如权利要求1所述的基于手机信令的地铁出行识别方法,其特征在于,所述构建基站-地铁站参考系统,具体包括:

基于k近邻的算法构建基站-地铁站参考系统,计算该地铁站与所有基站的距离,并选取距离最近的k个基站作为参考基站;

基于覆盖范围的算法构建基站-地铁站参考系统,计算该地铁站与所有基站的距离,并选取距离为d米范围内的基站作为参考基站;

对两种算法的结果取并集,得到以下形式的基站-地铁站参考系统:

$$f(C_i) = S_i,$$

其中, $f(\cdot)$ 表示索引函数, C_i 表示基站, S_i 表示 C_i 对应的地铁站,即 C_i 是 S_i 的参考基站;如果 C_i 不为任何地铁站的参考基站,则定义 $f(C_i)$ 为空集;如果 C_i 同时属于多个地铁站的参考基站,则 $f(C_i)$ 为距离 C_i 最近的地铁站。

3. 如权利要求1所述的基于手机信令的地铁出行识别方法,其特征在于,所述对所述第一用户基站轨迹序列进行数据预处理得到第二用户基站轨迹序列,具体包括:

剔除因乒乓效应产生的数据,当所述第一用户基站轨迹序列中的数据存在 $C_{i-1} = C_{i+1}$, $C_{i-1} \neq C_i$,且 $t_{i+1} - t_{i-1} < \theta$ 时,将符合上述条件的 (C_i, t_i) 数据进行剔除;其中 θ 为时间阈值;

剔除错误数据,计算用户的移动速度,当所述第一用户基站轨迹序列中的数据存在 $\frac{dis(C_{i-1}, C_i)}{(t_i - t_{i-1})} > \rho$ 且 $\frac{dis(C_i, C_{i+1})}{(t_{i+1} - t_i)} > \rho$ 时,将符合上述条件的 (C_i, t_i) 数据进行剔除;其中, $dis(\cdot)$ 为两个基站之间的距离, ρ 为速度阈值;

对经过剔除乒乓效应和错误数据后的用户基站轨迹序列进行合并,对于 (C_i, t_i) 与 (C_{i+1}, t_{i+1}) ,如果 $C_i = C_{i+1}$,则将数据合并为 (C_i, t_i, t_{i+1}) ,其中 t_i 表示起始时间, t_{i+1} 表示终止时间;如果 $C_i \neq C_{i+1}$,则将数据转换为 (C_i, t_i, t_i) 与 $(C_{i+1}, t_{i+1}, t_{i+1})$;经过上述的数据预处理后,第二用户基站轨迹序列中的数据表示为 (C, t^s, t^e) ,其中C为基站, t^s 表示起始时间, t^e 表示终止时间。

4. 如权利要求1所述的基于手机信令的地铁出行识别方法,其特征在于,所述利用所述基站-地铁站参考系统将所述第二用户基站轨迹数据序列转化为第一地铁站轨迹序列,对所述第一地铁站轨迹序列进行分段后得到若干段用户地铁站轨迹序列,具体包括:

利用所述基站-地铁站参考系统将所述第二用户基站轨迹序列:

$$\{(C_1, t_1^s, t_1^e), (C_2, t_2^s, t_2^e), \dots, (C_n, t_n^s, t_n^e)\},$$

转换为第一地铁站序列:

$$\{(f(C_1), t_1^s, t_1^e), (f(C_2), t_2^s, t_2^e), \dots, (f(C_i), t_i^s, t_i^e), \dots, (f(C_n), t_n^s, t_n^e)\},$$

其中,若 $f(C_i)$ 为空集,则将 $(f(C_i), t_i^s, t_i^e)$ 从第一地铁站序列中剔除,并且最终把第一地铁站轨迹序列表示为:

$$\{(S_1, t_1^s, t_1^e), (S_2, t_2^s, t_2^e), \dots, (S_j, t_j^s, t_j^e), \dots, (S_m, t_m^s, t_m^e)\},$$

其中 S_j 表示地铁站, t_j^s 表示起始时间, t_j^e 表示终止时间;

对所述第一地铁站轨迹序列进行分段,如果 $t_j^s - t_{j-1}^s > \tau$, τ 为时间阈值,则认为 S_{j-1} 属于上一段旅程的终点, S_j 属于新一段旅程的起点;另外,如果 $t_j^e - t_j^s > \tau$,则认为 S_j 既是上一段旅程的终点,同时也是新一段旅程的起点,最终得到若干段用户地铁站轨迹序列。

5. 如权利要求1所述的基于手机信令的地铁出行识别方法,其特征在于,所述对每一段所述用户地铁站轨迹序列根据地铁的行驶时间及地铁的运营时间进行匹配,剔除不合理的用户地铁站轨迹序列,具体包括:

一段用户地铁站轨迹序列定义为:

$$\{(S_1, t_1^s, t_1^e), (S_2, t_2^s, t_2^e), \dots, (S_k, t_k^s, t_k^e)\},$$

利用地铁的行驶时间对每一段用户地铁站轨迹序列进行验证,其实际行驶时间为 $T = t_k^s - t_1^e$, S_1 到 S_k 的旅程期望时间为 \hat{T} ,如果满足 $|\hat{T} - T| < \beta$, β 为时间阈值,则认为这段用户地铁站轨迹序列对应的旅程合理;否则,剔除该段用户地铁站轨迹序列;

利用地铁的运营时间对每一段旅程合理的用户地铁站轨迹序列进行验证,保留满足 t_1^e 与 t_k^s 均在地铁的运营时间内的用户地铁站轨迹序列,否则剔除该段用户地铁站轨迹序列。

6. 如权利要求1所述的基于手机信令的地铁出行识别方法,其特征在于,所述根据余下的每一段所述用户地铁站轨迹序列得到该用户的地铁出行轨迹,具体包括:

根据余下的每一段所述用户地铁站轨迹序列,将用户的地铁出行轨迹表示为:

$$\{T_k^s, T_k^e, S_1^k, S_2^k, \dots, S_m^k\},$$

其中, T_k^s 为该段用户的地铁出行轨迹的出发时间, T_k^e 为该段用户的地铁出行轨迹的结束时间, $\{S_1^k, S_2^k, \dots, S_m^k\}$ 为一段用户经过的地铁轨迹;

通过比较用户的地铁出行轨迹中的 $\{S_1^k, S_2^k, \dots, S_m^k\}$ 和实际地铁线路中从 S_1^k 到 S_m^k 的站点是否一一对应,若一一对应,则得到该段用户的地铁出行轨迹;反之,根据实际地铁线路将该段地铁出行轨迹补充完整后,得到该段用户的地铁出行轨迹。

7. 如权利要求1所述的基于手机信令的地铁出行识别方法,其特征在于,所述根据每一段所述用户的地铁出行轨迹对应的第二用户基站轨迹数据序列对该段旅程进行空间相似

度匹配,保留匹配成功的旅程对应的用户的地铁出行轨迹,具体包括:

利用地铁线路空间相似度对每一段用户的地铁出行轨迹进行验证,比较该段用户地铁站轨迹序列中对应的基站轨迹与 S_1 到 S_k 地铁线路 L_{S_1,S_K} 的相似度;其中,定义基站到地铁线路的距离为基站到该线路的地铁站最近的距离:

$$dis(C_i, L_{S_1,S_K}) = \min\{dis(C_i, S_j), S_j \in L_{S_1,S_K}\},$$

其中, $dis(\cdot)$ 为距离度量方法, $S_j \in L_{S_1,S_K}$ 表示 S_j 是地铁线路 L_{S_1,S_K} 上的一个站点;基站轨迹与地铁相似度定义为所有基站与该线路距离的平均值;如果相似度超过阈值 γ ,则将剔除该段用户地铁出行轨迹,反之,则保留匹配成功的旅程对应的用户的地铁出行轨迹。

8.一种基于手机信令的地铁出行识别装置,其特征在于,包括:

构建参考系统模块,用于构建基站-地铁站参考系统;所述基站-地铁站参考系统是根据计算地铁站与所有基站的距离,选取所述地铁站的用户能连接到的基站作为参考基站;

获取第一用户基站轨迹序列模块,用于获取手机信令数据,记用户在时刻 t_i 连接到的基站为 C_i ,并规定 $t_i < t_{i+1}$,根据用户连接的基站且按照时间序列排序,获得第一用户基站轨迹序列;

获取第二用户基站轨迹序列模块,用于对所述第一用户基站轨迹序列进行数据预处理得到第二用户基站轨迹序列;

分段模块,用于利用所述基站-地铁站参考系统将所述第二用户基站轨迹数据序列转化为第一地铁站轨迹序列,对所述第一地铁站轨迹序列进行分段后得到若干段用户地铁站轨迹序列;

匹配模块,用于对每一段所述用户地铁站轨迹序列根据地铁的行驶时间、地铁的运营时间进行匹配,剔除不合理的用户地铁站轨迹序列;

获取地铁出行轨迹模块,用于根据余下的每一段所述用户地铁站轨迹序列得到该用户的地铁出行轨迹;根据每一段所述用户的地铁出行轨迹对应的第二用户基站轨迹数据序列对该段旅程进行空间相似度匹配,保留匹配成功的旅程对应的用户的地铁出行轨迹。

9.一种终端设备,其特征在于,包括处理器、存储器以及存储在所述存储器中且被配置为由所述处理器执行的计算机程序,所述处理器执行所述计算机程序时实现如权利要求1至7中任意一项所述的一种基于手机信令的地铁出行识别方法。

10.一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包括存储的计算机程序,其中,在所述计算机程序运行时控制所述计算机可读存储介质所在设备执行如权利要求1至7中任意一项所述的一种基于手机信令的地铁出行识别方法。

基于手机信令的地铁出行识别方法、装置、设备及介质

技术领域

[0001] 本发明涉及智慧交通技术领域,尤其涉及一种基于手机信令的地铁出行识别方法、装置、设备及介质。

背景技术

[0002] 地铁作为大型城市公共交通系统中最重要的交通工具,在城市运转中发挥着重要的作用。掌握城市居民使用地铁的情况,有助于分析城市居民的日常流动,在城市规划、道路资源分配、公共交通系统优化、地铁线路班次优化、班次优化、商业选址、用户画像等方面有着重要的应用。

[0003] 传统的居民地铁出行调查大多基于地铁支付数据或采用人工调查的方式,例如调查问卷、电话访问等。随着移动互联网的发展以及移动设备的普及,基于手机传感器数据、GPS数据、手机信令等数据的技术方法逐渐引起学术界与工业界的关注。由于手机信令数据具有群体覆盖范围广、数据实时性强、采集成本低等优势,更适合用于全面大规模地进行城市居民交通方式推演。目前基于手机信令数据的方法主要分为两类:基于机器学习的方法与基于导航数据匹配的方法。

[0004] 针对手机信令数据的识别方法中,基于机器学习的方法往往依赖于大量的训练数据,需要耗费大量人力与时间进行数据的采集与标注;而基于导航数据的技术,需要借助第三方API,例如高德、百度,在城市大规模的地铁识别中有所受限。此外,绝大多数现有的技术需要先将用户的轨迹分成不同的旅程,并假设得到的每段用户轨迹只包含一种交通方式。然而,对于存在交通工具有换乘的旅程,这些方法容易产生误判,错误地把多种交通工具的使用识别为一种。

发明内容

[0005] 本发明提供一种基于手机信令的地铁出行识别方法,面向大规模城市居民地铁出行识别的需求,能够在不依赖训练数据与第三方API的情况下,实现及时、准确的大规模城市居民地铁出行轨迹识别。

[0006] 为实现上述目的,本发明实施例提供了一种基于手机信令的地铁出行识别方法,包括:

[0007] 构建基站-地铁站参考系统;所述基站-地铁站参考系统是根据计算地铁站与所有基站的距离,选取所述地铁站的用户能连接到的基站作为参考基站;

[0008] 获取手机信令数据,记用户在时刻 t_i 连接到的基站为 C_i ,并规定 $t_i < t_{i+1}$,根据用户连接的基站且按照时间序列排序,获得第一用户基站轨迹序列;其中,所述第一用户基站轨迹序列表示为: $\{(C_1, t_1), (C_2, t_2), \dots, (C_i, t_i), \dots (C_n, t_n)\}$;其中 $C_i = (lat_i, lon_i)$ 表示一个用户 t_i 时刻连接的基站, (lat_i, lon_i) 表示基站的经纬度坐标;

[0009] 对所述第一用户基站轨迹序列进行数据预处理得到第二用户基站轨迹序列;

[0010] 利用所述基站-地铁站参考系统将所述第二用户基站轨迹数据序列转化为第一地

铁路站迹序列,对所述第一地铁站迹序列进行分段后得到若干段用户地铁站迹序列;

[0011] 对每一段所述用户地铁站迹序列根据地铁的行驶时间、地铁的运营时间进行匹配,剔除不合理的用户地铁站迹序列;

[0012] 根据余下的每一段所述用户地铁站迹序列得到该用户的地铁出行轨迹;

[0013] 根据每一段所述用户的地铁出行轨迹对应的第二用户基站迹数据序列对该段旅程进行空间相似度匹配,保留匹配成功的旅程对应的用户的地铁出行轨迹。

[0014] 其中,所述构建基站-地铁站参考系统,具体包括:

[0015] 基于k近邻的算法构建基站-地铁站参考系统,计算该地铁站与所有基站的距离,并选取距离最近的k个基站作为参考基站;

[0016] 基于覆盖范围的算法构建基站-地铁站参考系统,计算该地铁站与所有基站的距离,并选取距离为d米范围内的基站作为参考基站;

[0017] 对两种算法的结果取并集,得到以下形式的基站-地铁站参考系统:

[0018] $f(C_i) = S_i$,

[0019] 其中, $f(\cdot)$ 表示索引函数, C_i 表示基站, S_i 表示 C_i 对应的地铁站,即 C_i 是 S_i 的参考基站;如果 C_i 不为任何地铁站的参考基站,则定义 $f(C_i)$ 为空集;如果 C_i 同时属于多个地铁站的参考基站,则 $f(C_i)$ 为距离 C_i 最近的地铁站。

[0020] 进一步的,所述对所述第一用户基站迹序列进行数据预处理得到第二用户基站迹序列,具体包括:

[0021] 剔除因乒乓效应产生的数据,当所述第一用户基站迹序列中的数据存在 $C_{i-1} = C_{i+1}, C_{i-1} \neq C_i$,且 $t_{i+1} - t_{i-1} < \theta$ 时,将符合上述条件的 (C_i, t_i) 数据进行剔除;其中 θ 为时间阈值;

[0022] 剔除错误数据,计算用户的移动速度,当所述第一用户基站迹序列中的数据存在 $\frac{dis(C_{i-1}, C_i)}{(t_i - t_{i-1})} > \rho$ 且 $\frac{dis(C_i, C_{i+1})}{(t_{i+1} - t_i)} > \rho$ 时,将符合上述条件的 (C_i, t_i) 数据进行剔除;其中,

$dis(\cdot)$ 为两个基站之间的距离, ρ 为速度阈值;

[0023] 对经过剔除乒乓效应和错误数据后的用户基站迹序列进行合并,对于 (C_i, t_i) 与 (C_{i+1}, t_{i+1}) ,如果 $C_i = C_{i+1}$,则将数据合并为 (C_i, t_i, t_{i+1}) ,其中 t_i 表示起始时间, t_{i+1} 表示终止时间;如果 $C_i \neq C_{i+1}$,则将数据转换为 (C_i, t_i, t_i) 与 $(C_{i+1}, t_{i+1}, t_{i+1})$;经过上述的数据预处理后,第二用户基站迹序列中的数据表示为 (C, t^s, t^e) ,其中C为基站, t^s 表示起始时间, t^e 表示终止时间。

[0024] 其中,所述利用所述基站-地铁站参考系统将所述第二用户基站迹数据序列转化为第一地铁站迹序列,对所述第一地铁站迹序列进行分段后得到若干段用户地铁站迹序列,具体包括:

[0025] 利用所述基站-地铁站参考系统将所述第二用户基站迹序列:

[0026] $\{(C_1, t_1^s, t_1^e), (C_2, t_2^s, t_2^e), \dots, (C_n, t_n^s, t_n^e)\}$,

[0027] 转换为第一地铁站序列:

[0028] $\{(f(C_1), t_1^s, t_1^e), (f(C_2), t_2^s, t_2^e), \dots, (f(C_i), t_i^s, t_i^e), \dots, (f(C_n), t_n^s, t_n^e)\}$,

[0029] 其中,若 $f(C_i)$ 为空集,则将 $(f(C_i), t_i^s, t_i^e)$ 从第一地铁站序列中剔除,并且最终把第一地铁站迹序列表示为:

[0030] $\{(S_1, t_1^s, t_1^e), (S_2, t_2^s, t_2^e), \dots, (S_j, t_j^s, t_j^e), \dots, (S_m, t_m^s, t_m^e)\},$

[0031] 其中 S_j 表示地铁站, t_j^s 表示起始时间, t_j^e 表示终止时间;

[0032] 对所述第一地铁站轨迹序列进行分段, 如果 $t_j^s - t_{j-1}^s > \tau$, τ 为时间阈值, 则认为 S_{j-1} 属于上一段旅程的终点, S_j 属于新一段旅程的起点; 另外, 如果 $t_j^e - t_j^s > \tau$, 则认为 S_j 既是上一段旅程的终点, 同时也是新一段旅程的起点, 最终得到若干段用户地铁站轨迹序列。

[0033] 进一步的, 所述对每一段所述用户地铁站轨迹序列根据地铁的行驶时间及地铁的运营时间进行匹配, 剔除不合理的用户地铁站轨迹序列, 具体包括:

[0034] 一段用户地铁站轨迹序列定义为:

[0035] $\{(S_1, t_1^s, t_1^e), (S_2, t_2^s, t_2^e), \dots, (S_k, t_k^s, t_k^e)\},$

[0036] 利用地铁的行驶时间对每一段用户地铁站轨迹序列进行验证, 其实际行驶时间为 $T = t_k^s - t_1^e$, S_1 到 S_k 的旅程期望时间为 \hat{T} , 如果满足 $|\hat{T} - T| < \beta$, β 为时间阈值, 则认为这段用户地铁站轨迹序列对应的旅程合理; 否则, 剔除该段用户地铁站轨迹序列;

[0037] 利用地铁的运营时间对每一段旅程合理的用户地铁站轨迹序列进行验证, 保留满足 t_1^e 与 t_k^s 均在地铁的运营时间内的用户地铁站轨迹序列, 否则剔除该段用户地铁站轨迹序列。

[0038] 再进一步的, 所述根据余下的每一段所述用户地铁站轨迹序列得到该用户的地铁出行轨迹, 具体包括:

[0039] 根据余下的每一段所述用户地铁站轨迹序列, 将用户的地铁出行轨迹表示为:

[0040] $\{T_k^S, T_k^E, S_1^k, S_2^k, \dots, S_m^k\},$

[0041] 其中, T_k^S 为该段用户的地铁出行轨迹的出发时间, T_k^E 为该段用户的地铁出行轨迹的结束时间, $\{S_1^k, S_2^k, \dots, S_m^k\}$ 为一段用户经过的地铁轨迹;

[0042] 通过比较用户的地铁出行轨迹中的 $\{S_1^k, S_2^k, \dots, S_m^k\}$ 和实际地铁线路中从 S_1^k 到 S_m^k 的站点是否一一对应, 若一一对应, 则得到该段用户的地铁出行轨迹; 反之, 根据实际地铁线路将该段地铁出行轨迹补充完整后, 得到该段用户的地铁出行轨迹。

[0043] 进一步的, 所述根据每一段所述用户的地铁出行轨迹对应的第二用户基站轨迹数据序列对该段旅程进行空间相似度匹配, 保留匹配成功的旅程对应的用户的地铁出行轨迹, 具体包括:

[0044] 利用地铁线路空间相似度对每一段用户的地铁出行轨迹进行验证, 比较该段用户地铁站轨迹序列中对应的基站轨迹与 S_1 到 S_k 地铁线路 L_{S_1, S_K} 的相似度; 其中, 定义基站到地铁线路的距离为基站到该线路的地铁站最近的距离:

[0045] $dis(C_i, L_{S_1, S_K}) = \min\{dis(C_i, S_j), S_j \in L_{S_1, S_K}\},$

[0046] 其中, $\text{dis}(\cdot)$ 为距离度量方法, $S_j \in L_{S_1, S_K}$ 表示 S_j 是地铁线路 L_{S_1, S_K} 上的一个站点; 基站轨迹与地铁相似度定义为所有基站与该线路距离的平均值; 如果相似度超过阈值 γ , 则将剔除该段用户地铁出行轨迹, 反之, 则保留匹配成功的旅程对应的用户的地铁出行轨迹。

[0047] 相应地, 本发明实施例还提供了一种基于手机信令的地铁出行识别装置, 包括:

[0048] 构建参考系统模块, 用于构建基站-地铁站参考系统; 所述基站-地铁站参考系统是根据计算地铁站与所有基站的距离, 选取所述地铁站的用户能连接到的基站作为参考基站;

[0049] 获取第一用户基站轨迹序列模块, 用于获取手机信令数据, 记用户在时刻 t_i 连接到的基站为 C_i , 并规定 $t_i < t_{i+1}$, 根据用户连接的基站且按照时间序列排序, 获得第一用户基站轨迹序列;

[0050] 获取第二用户基站轨迹序列模块, 用于对所述第一用户基站轨迹序列进行数据预处理得到第二用户基站轨迹序列;

[0051] 分段模块, 用于利用所述基站-地铁站参考系统将所述第二用户基站轨迹数据序列转化为第一地铁站轨迹序列, 对所述第一地铁站轨迹序列进行分段后得到若干段用户地铁站轨迹序列;

[0052] 匹配模块, 用于对每一段所述用户地铁站轨迹序列根据地铁的行驶时间及地铁的运营时间进行匹配, 剔除不合理的用户地铁站轨迹序列;

[0053] 获取地铁出行轨迹模块, 用于获取地铁出行轨迹模块, 用于根据余下的每一段所述用户地铁站轨迹序列得到该用户的地铁出行轨迹; 根据每一段所述用户的地铁出行轨迹对应的第二用户基站轨迹数据序列对该段旅程进行空间相似度匹配, 保留匹配成功的旅程对应的用户的地铁出行轨迹。

[0054] 相应地, 本发明实施例还提供一种终端设备, 包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序, 所述处理器执行所述程序时实现上述基于手机信令的地铁出行识别方法的步骤。

[0055] 相应地, 本发明实施例还提供了一种计算机可读存储介质, 其上存储有计算机程序, 该程序被处理器执行时实现上述基于手机信令的地铁出行识别方法的步骤。

[0056] 相比于现有技术, 本发明具有如下有益效果:

[0057] 本发明基于手机信令数据, 提出一种融合时间和空间两方面的精准识别方法, 在不依赖任何的训练数据与第三方导航数据的情况下, 实现及时、准确的大规模城市居民地铁出行方式识别。其中, 手机信令数据具有群体覆盖范围广、数据实时性强、采集成本低等优势, 更适合用于全面大规模地进行城市居民交通方式推演, 本发明有利于实现低成本、大规模、高精度的地铁出行识别以及方案的快速和稳定部署; 另外, 本发明不需要事先对用户的轨迹进行分段, 不依赖于其他假设条件, 能更好适用于用户换乘交通的情况, 能较完整的识别用户的地铁交通链。

附图说明

[0058] 图1是本发明实施例提供的基于手机信令的地铁出行识别方法的流程示意图;

[0059] 图2是本发明实施例提供的基于手机信令的地铁出行识别装置的结构示意图。

具体实施方式

[0060] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0061] 参见图1,是本发明提供的基于手机信令的地铁出行识别方法的一个实施例的流程图示意图,该方法包括步骤S101至S107,具体如下:

[0062] S101,构建基站-地铁站参考系统;所述基站-地铁站参考系统是根据计算地铁站与所有基站的距离,选取所述地铁站的用户能连接到的基站作为参考基站;

[0063] 具体的,所述构建基站-地铁站参考系统,具体包括:

[0064] 基于k近邻的算法构建基站-地铁站参考系统,计算该地铁站与所有基站的距离,并选取距离最近的k个基站作为参考基站;

[0065] 基于覆盖范围的算法构建基站-地铁站参考系统,计算该地铁站与所有基站的距离,并选取距离为d米范围内的基站作为参考基站;

[0066] 对两种算法的结果取并集,得到以下形式的基站-地铁站参考系统:

[0067] $f(C_i) = S_i$,

[0068] 其中, $f(\cdot)$ 表示索引函数, C_i 表示基站, S_i 表示 C_i 对应的地铁站,即 C_i 是 S_i 的参考基站;如果 C_i 不为任何地铁站的参考基站,则定义 $f(C_i)$ 为空集;如果 C_i 同时属于多个地铁站的参考基站,则 $f(C_i)$ 为距离 C_i 最近的地铁站。

[0069] 在具体的实施例中,由于不同基站的覆盖范围存在重合的区域,同一个地铁站可能被多个基站的信号所覆盖,导致同一地铁站的用户可能同时连接到不同的基站。因此,对于每个地铁站,我们首先选取该地铁站用户最可能连接到的基站,作为该地铁站的参考基站。当需要判断用户与地铁站的距离时,我们先验证该用户连接的基站是否为某一地铁站的参考基站;若否,则直接跳过。

[0070] S102,获取手机信令数据,记用户在时刻 t_i 连接到的基站为 C_i ,并规定 $t_i < t_{i+1}$,根据用户连接的基站且按照时间序列排序,获得第一用户基站轨迹序列;

[0071] 其中,所述第一用户基站轨迹序列表示为:

[0072] $\{(C_1, t_1), (C_2, t_2), \dots, (C_i, t_i), \dots, (C_n, t_n)\}$;

[0073] 其中 $C_i = (\text{lat}_i, \text{lon}_i)$ 表示一个用户 t_i 时刻连接的基站, $(\text{lat}_i, \text{lon}_i)$ 表示基站的经纬度坐标。

[0074] 手机信令数据是由手机用户在发生通话、发短信或移动位置等事件时,被运营商的通信基站捕获并记录同一用户信令轨迹所产生。

[0075] 在本发明实施例中,通过用户的手机信令数据中连接基站得到用户的基站轨迹序列,例如,当用户从广州火车站坐到淘金,其用户的基站轨迹序列表示为: $\{((113.256512, 23.148339), 9:00:00), ((113.282933, 23.145765), 9:03:00), ((113.293138, 23.142699), 9:06:00)\}$;

[0076] S103,对所述第一用户基站轨迹序列进行数据预处理得到第二用户基站轨迹序列;

[0077] 具体的,所述对所述第一用户基站轨迹序列进行数据预处理得到第二用户基站轨

迹序列,具体包括:

[0078] 剔除因乒乓效应产生的数据,当所述第一用户基站轨迹序列中的数据存在 $C_{i-1} = C_{i+1}, C_{i-1} \neq C_i$,且 $t_{i+1} - t_{i-1} < \theta$ 时,将符合上述条件的 (C_i, t_i) 数据进行剔除;其中 θ 为时间阈值;

[0079] 剔除错误数据,计算用户的移动速度,当所述第一用户基站轨迹序列中的数据存在 $\frac{dis(C_{i-1}, C_i)}{(t_i - t_{i-1})} > \rho$ 且 $\frac{dis(C_i, C_{i+1})}{(t_{i+1} - t_i)} > \rho$ 时,将符合上述条件的 (C_i, t_i) 数据进行剔除;其中,

$dis(\cdot)$ 为两个基站之间的距离, ρ 为速度阈值;

[0080] 对经过剔除乒乓效应和错误数据后的用户基站轨迹序列进行合并,对于 (C_i, t_i) 与 (C_{i+1}, t_{i+1}) ,如果 $C_i = C_{i+1}$,则将数据合并为 (C_i, t_i, t_{i+1}) ,其中 t_i 表示起始时间, t_{i+1} 表示终止时间;如果 $C_i \neq C_{i+1}$,则将数据转换为 (C_i, t_i, t_i) 与 $(C_{i+1}, t_{i+1}, t_{i+1})$;经过上述的数据预处理后,第二用户基站轨迹序列中的数据表示为 (C, t^s, t^e) ,其中 C 为基站, t^s 表示起始时间, t^e 表示终止时间。

[0081] 在具体的实施例中,通过剔除因乒乓效应产生的以及错误的对第一用户基站轨迹的数据进行清洗。

[0082] S104,利用所述基站-地铁站参考系统将所述第二用户基站轨迹数据序列转化为第一地铁站轨迹序列,对所述第一地铁站轨迹序列进行分段后得到若干段用户地铁站轨迹序列;

[0083] 其中,所述利用所述基站-地铁站参考系统将所述第二用户基站轨迹数据序列转化为第一地铁站轨迹序列,对所述第一地铁站轨迹序列进行分段后得到若干段用户地铁站轨迹序列,具体包括:

[0084] 利用所述基站-地铁站参考系统将所述第二用户基站轨迹序列:

[0085] $\{(C_1, t_1^s, t_1^e), (C_2, t_2^s, t_2^e), \dots, (C_n, t_n^s, t_n^e)\}$,

[0086] 转换为第一地铁站序列:

[0087] $\{(f(C_1), t_1^s, t_1^e), (f(C_2), t_2^s, t_2^e), \dots, (f(C_i), t_i^s, t_i^e), \dots, (f(C_n), t_n^s, t_n^e)\}$,

[0088] 其中,若 $f(C_i)$ 为空集,则将 $(f(C_i), t_i^s, t_i^e)$ 从第一地铁站序列中剔除,并且最终把第一地铁站轨迹序列表示为:

[0089] $\{(S_1, t_1^s, t_1^e), (S_2, t_2^s, t_2^e), \dots, (S_j, t_j^s, t_j^e), \dots, (S_m, t_m^s, t_m^e)\}$,

[0090] 其中 S_j 表示地铁站, t_j^s 表示起始时间, t_j^e 表示终止时间;

[0091] 对所述第一地铁站轨迹序列进行分段,如果 $t_j^s - t_{j-1}^e > \tau$, τ 为时间阈值,则认为 S_{j-1} 属于上一段旅程的终点, S_j 属于新一段旅程的起点;另外,如果 $t_j^e - t_j^s > \tau$,则认为 S_j 既是上一段旅程的终点,同时也是新一段旅程的起点,最终得到若干段用户地铁站轨迹序列。

[0092] 在本发明实施例中,根据预先构建的基站-地铁站参考系统将用户连接基站的轨迹转化成地铁站轨迹序列,需要说明的是,该段地铁站轨迹序列的时间跨度可能会很大,即用户可能存在多段地铁旅程。

[0093] 在具体实施例中,早上8点30用户从家里出发坐地铁到公司,晚上6点用户下班坐

地铁从公司回到家里,那么在用户的地铁站轨迹序列中应存在两段旅程,因此需要将该段地铁站轨迹序列通过时间分段的方式将其划分成若干段用户地铁站轨迹序列,每一段用户地铁站轨迹序列代表一段用户完整的地铁旅程。

[0094] S105,对每一段所述用户地铁站轨迹序列根据地铁的行驶时间及地铁的运营时间进行匹配,剔除不合理的用户地铁站轨迹序列;

[0095] 具体的,所述对每一段所述用户地铁站轨迹序列根据地铁的行驶时间及地铁的运营时间进行匹配,剔除不合理的用户地铁站轨迹序列,具体包括:

[0096] 一段用户地铁站轨迹序列定义为:

[0097] $\{(S_1, t_1^s, t_1^e), (S_2, t_2^s, t_2^e), \dots, (S_k, t_k^s, t_k^e)\}$,

[0098] 利用地铁的行驶时间对每一段用户地铁站轨迹序列进行验证,其实际行驶时间为 $T = t_k^s - t_1^e$, S_1 到 S_k 的旅程期望时间为 \hat{T} ,如果满足 $|\hat{T} - T| < \beta$, β 为时间阈值,则认为这段用户地铁站轨迹序列对应的旅程合理;否则,剔除该段用户地铁站轨迹序列;

[0099] 利用地铁的运营时间对每一段旅程合理的用户地铁站轨迹序列进行验证,保留满足 t_1^e 与 t_k^s 均在地铁的运营时间内的用户地铁站轨迹序列,否则剔除该段用户地铁站轨迹序列;

[0100] 在具体的实施例中,通过地铁的行驶时间及运营时间将所有不合理的数据进行筛选和验证,最终得到的用户地铁站序列就是用户搭乘地铁实际的起始时间和实际的线路。

[0101] S106,根据余下的每一段所述用户地铁站轨迹序列得到该用户的地铁出行轨迹。

[0102] 具体的,所述根据余下的每一段所述用户地铁站轨迹序列得到该用户的地铁出行轨迹,具体包括:

[0103] 根据余下的每一段所述用户地铁站轨迹序列,将用户的地铁出行轨迹表示为:

[0104] $\{T_k^S, T_k^E, S_1^k, S_2^k, \dots, S_m^k\}$,

[0105] 其中, T_k^S 为该段用户的地铁出行轨迹的出发时间, T_k^E 为该段用户的地铁出行轨迹的结束时间, $\{S_1^k, S_2^k, \dots, S_m^k\}$ 为一段用户经过的地铁轨迹;

[0106] 通过比较用户的地铁出行轨迹中的 $\{S_1^k, S_2^k, \dots, S_m^k\}$ 和实际地铁线路中从 S_1^k 到 S_m^k 的站点是否一一对应,若一一对应,则得到该段用户的地铁出行轨迹;反之,根据实际地铁线路将该段地铁出行轨迹补充完整后,得到该段用户的地铁出行轨迹。

[0107] 在具体实施例中,由于基站数据的采样间隔可能较大,一段地铁轨迹中的 S_i 与 S_{i+1} 可能不是地铁线路中相邻的地铁站;因此,对于经过上述步骤得到的地铁出行轨迹,例如用户从广州火车站出发到淘金,用户的地铁轨迹序列中地铁轨迹为: {广州火车站, 淘金}, 应按照从广州火车站到淘金的实际地铁线路站点进行比较,根据地铁线路得知站点应为 {广州火车站, 小北, 淘金}, 于是应将其补充完整,即可得到正确的用户的地铁出行轨迹。

[0108] S107,根据每一段所述用户的地铁出行轨迹对应的第二用户基站轨迹数据序列对该段旅程进行空间相似度匹配,保留匹配成功的旅程对应的用户的地铁出行轨迹。

[0109] 具体的,利用地铁线路空间相似度对每一段用户的地铁出行轨迹进行验证,比较

该段用户地铁站轨迹序列中对应的基站轨迹与 S_1 到 S_k 地铁线路 L_{S_1, S_K} 的相似度;其中,定义基站到地铁线路的距离为基站到该线路的地铁站最近的距离:

$$[0110] \quad dis(C_i, L_{S_1, S_K}) = \min\{dis(C_i, S_j), S_j \in L_{S_1, S_K}\},$$

[0111] 其中, $dis(\cdot)$ 为距离度量方法, $S_j \in L_{S_1, S_K}$ 表示 S_j 是地铁线路 L_{S_1, S_K} 上的一个站点;基站轨迹与地铁相似度定义为所有基站与该线路距离的平均值;如果相似度超过阈值 γ ,则将剔除该段用户地铁出行轨迹,反之,则保留匹配成功的旅程对应的用户的地铁出行轨迹。

[0112] 具体的,给定两个位置A、B的经纬度坐标, $dis(\cdot)$ 的计算如下:

$$[0113] \quad S = 2 * \arcsin \sqrt{\sin^2 \frac{a}{2} + \cos(lat1) * \cos(lat2) * \sin^2 \frac{b}{2}} * 6378.137,$$

[0114] 其中,(lat1,lon1)表示A的经纬度坐标;(lat2,lon2)表示B的经纬度坐标; $a = lat1 - lat2$ 表示两点纬度之差, $b = lon1 - lon2$ 为两点经度之差;6378.137为地球半径,单位为千米。

[0115] 综上,本发明实施例提供的基于手机信令的地铁出行识别方法,融合时间和空间两方面进行精准识别,能够在不依赖任何的训练数据与第三方导航数据的情况下,实现及时、准确的大规模城市居民地铁出行方式识别。其中,手机信令数据具有群体覆盖范围广、数据实时性强、采集成本低等优势,更适合用于全面大规模地进行城市居民交通方式推演,本发明有利于实现低成本、大规模、高精度的地铁出行识别以及方案的快速和稳定部署;另外,本发明利用基站-地铁站参考系统,可以帮助减少大量的非必要计算,大大提高识别系统的效率;并且不需要事先对用户的轨迹进行分段,不依赖于其他假设条件,能更好适用于用户换乘交通的情况,能较完整的识别用户的地铁交通链。

[0116] 参见图2,是本发明实施例提供的基于手机信令的地铁出行识别装置的结构示意图。

[0117] 本发明实施例提供了一种基于手机信令的地铁出行识别装置,包括:

[0118] 构建参考系统模块1,用于构建基站-地铁站参考系统;所述基站-地铁站参考系统是根据计算地铁站与所有基站的距离,选取所述地铁站的用户能连接到的基站作为参考基站;

[0119] 获取第一用户基站轨迹序列模块2,用于获取手机信令数据,记用户在时刻 t_i 连接到的基站为 C_i ,并规定 $t_i < t_{i+1}$,根据用户连接的基站且按照时间序列排序,获得第一用户基站轨迹序列;

[0120] 获取第二用户基站轨迹序列模块3,用于对所述第一用户基站轨迹序列进行数据预处理得到第二用户基站轨迹序列;

[0121] 分段模块4,用于利用所述基站-地铁站参考系统将所述第二用户基站轨迹数据序列转化为第一地铁站轨迹序列,对所述第一地铁站轨迹序列进行分段后得到若干段用户地铁站轨迹序列;

[0122] 匹配模块5,用于对每一段所述用户地铁站轨迹序列根据地铁的行驶时间、地铁的运营时间进行匹配,剔除不合理的用户地铁站轨迹序列;

[0123] 获取地铁出行轨迹模块6,用于根据余下的每一段所述用户地铁站轨迹序列得到该用户的地铁出行轨迹;根据每一段所述用户的地铁出行轨迹对应的第二用户基站轨迹数据序列对该段旅程进行空间相似度匹配,保留匹配成功的旅程对应的用户地铁出行轨迹。

[0124] 本发明实施例还提供一种终端设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现上述基于手机信令的地铁出行识别方法的步骤。

[0125] 此外,本发明实施例还提供了一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现上述基于手机信令的地铁出行识别方法的步骤。

[0126] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到本发明可借助软件加必需的硬件平台的方式来实现,当然也可以全部通过硬件来实施。基于这样的理解,本发明的技术方案对背景技术做出贡献的全部或者部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例或者实施例的某些部分所述的方法。

[0127] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以所述权利要求的保护范围为准。

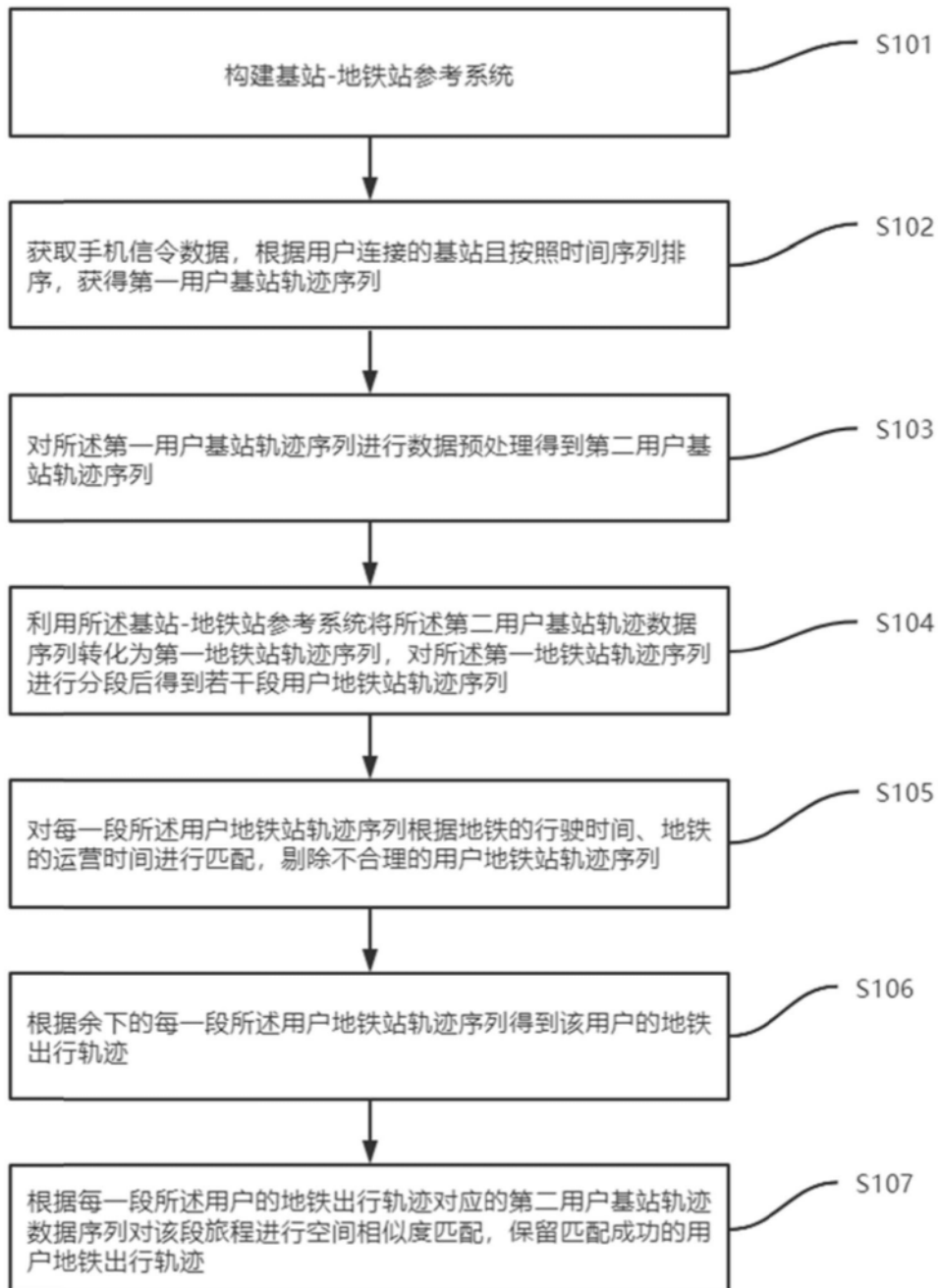


图1

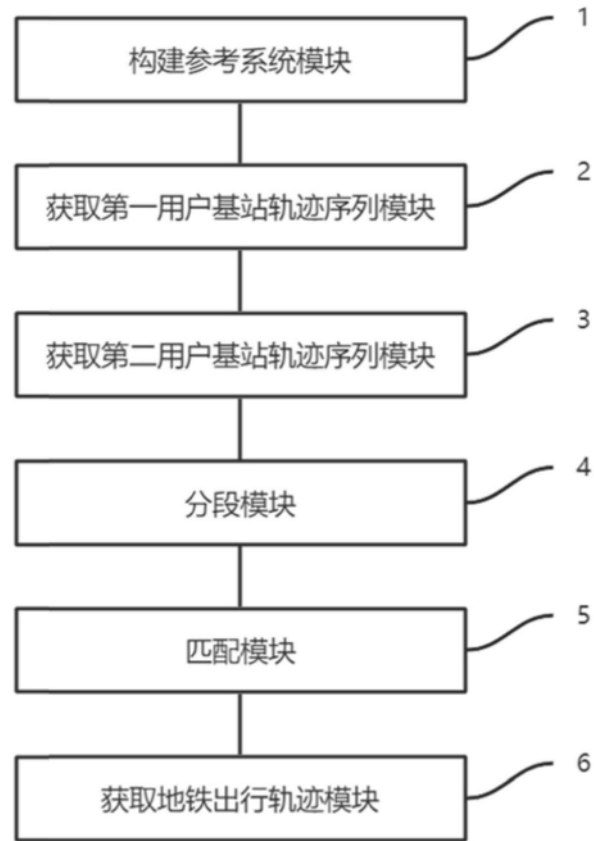


图2