



## (12)发明专利

(10)授权公告号 CN 108955693 B

(45)授权公告日 2020.06.16

(21)申请号 201810872312.X

(22)申请日 2018.08.02

(65)同一申请的已公布的文献号

申请公布号 CN 108955693 A

(43)申请公布日 2018.12.07

(73)专利权人 吉林大学

地址 130000 吉林省长春市前进大街2699号

(72)发明人 王恩 黄秋阳 杨永健 黄丽平

(74)专利代理机构 北京高沃律师事务所 11569

代理人 王戈

(51)Int.Cl.

G01C 21/20(2006.01)

(56)对比文件

CN 107747947 A, 2018.03.02,

CN 107607121 A, 2018.01.19,

CN 105387864 A, 2016.03.09,

CN 105333874 A, 2016.02.17,

JP H02138813 A, 1990.05.28,

审查员 徐雅

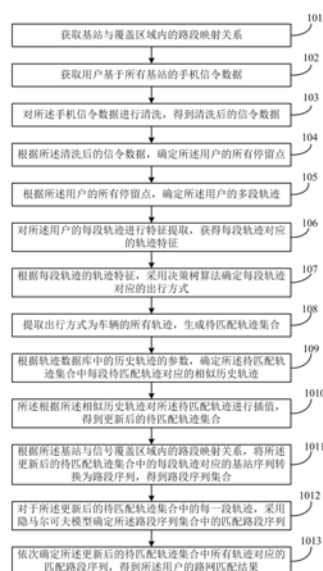
权利要求书4页 说明书14页 附图3页

## (54)发明名称

一种路网匹配的方法及系统

## (57)摘要

本发明公开一种路网匹配的方法及系统。该方法包括：获取基站与覆盖区域内的路段映射关系；获取用户的手机信令数据；对手机信令数据进行清洗；确定用户的所有停留点；确定用户的多段轨迹；对用户的每段轨迹进行特征提取；采用决策树算法确定每段轨迹对应的出行方式；提取出行方式为车辆的所有轨迹，生成待匹配轨迹集合；确定待匹配轨迹集合中每段待匹配轨迹对应的相似历史轨迹；根据相似历史轨迹对待匹配轨迹进行插值，得到更新后的待匹配轨迹集合；将每段轨迹对应的基站序列转换为路段序列；对于每一段轨迹，采用隐马尔可夫模型确定匹配路段序列，得到路网匹配结果。采用本发明的方法或系统，可以提高路网匹配的覆盖范围，提高路网匹配精度。



1. 一种路网匹配的方法,其特征在于,所述方法包括:
  - 获取基站与信号覆盖区域内的路段映射关系;
  - 获取用户基于所有基站的手机信令数据;
  - 对所述手机信令数据进行清洗,得到清洗后的信令数据;
  - 根据所述清洗后的信令数据,确定所述用户的所有停留点;所述停留点为所述用户在一个基站的信号覆盖区域内停留时间超过停留阈值的基站坐标;
  - 根据所述用户的所有停留点,确定所述用户的多段轨迹;每段轨迹对应一个基站序列;
  - 对所述用户的每段轨迹进行特征提取,获得每段轨迹对应的轨迹特征;所述轨迹特征包括:轨迹长度、轨迹半径、切换基站数量、平均速度和出行时段;
  - 根据每段轨迹的轨迹特征,采用决策树算法确定每段轨迹对应的出行方式;所述出行方式包括:步行、非机动车、车辆和地铁;
  - 提取出行方式为车辆的所有轨迹,生成待匹配轨迹集合;所述待匹配轨迹集合中包括多段待匹配轨迹;
  - 根据轨迹数据库中的历史轨迹的参数,确定所述待匹配轨迹集合中每段待匹配轨迹对应的相似历史轨迹;
  - 根据所述相似历史轨迹对所述待匹配轨迹进行插值,得到更新后的待匹配轨迹集合;
  - 根据所述基站与信号覆盖区域内的路段映射关系,将所述更新后的待匹配轨迹集合中的每段轨迹对应的基站序列转换为路段序列,得到路段序列集合,每一段轨迹对应的基站序列对应多个路段序列;
  - 对于所述更新后的待匹配轨迹集合中的每一段轨迹,采用隐马尔可夫模型确定所述路段序列集合中的匹配路段序列;
  - 依次确定所述更新后的待匹配轨迹集合中所有轨迹对应的匹配路段序列,得到所述用户的路网匹配结果。
2. 根据权利要求1所述的方法,其特征在于,所述获取基站与信号覆盖区域内的路段映射关系,之前还包括:
  - 获取最大阈值和最小阈值;
  - 判断第二基站与第一基站的距离是否小于最小阈值,得到第一判断结果;
  - 当所述第一判断结果表示第二基站与第一基站的距离不小于最小阈值时,判断第三基站与第一基站的距离是否小于最小阈值;
  - 当所述第一判断结果表示第二基站与第一基站的距离小于最小阈值时,将所述第二基站合并至第一基站集合;所述第一基站集合中包括第一基站;
  - 确定所述第一基站集合的中心位置的坐标,所述中心位置坐标的纬度为所述第一基站集合中所有基站的纬度的平均值,所述中心位置坐标的经度为所述第一基站集合中所有基站的经度的平均值;
  - 判断所述第一基站集合中每个基站距离所述中心位置的距离是否均小于最大阈值,得到第二判断结果;
  - 当所述第二判断结果表示所述第一基站集合中每个基站距离所述中心位置的距离不都小于最大阈值时,将所述第二基站从所述第一基站集合中移除,判断第三基站与第一基站的距离是否小于最小阈值;

当所述第二判断结果表示所述第一基站集合中每个基站距离所述中心位置的距离均小于最大阈值时,判断第三基站与第二基站的距离是否小于最小阈值;

直至遍历所有的基站,得到更新后的第一基站集合;

依次得到所有的基站集合;

确定每个基站集合的中心位置的坐标;

将每个基站集合的中心位置的坐标确定为更新后的第一基站的坐标,依次得到所有更新后的基站的坐标。

3. 根据权利要求2所述的方法,其特征在于,所述获取基站与覆盖区域内的路段映射关系,具体包括:

获取多个设定的覆盖区域半径;每个覆盖区域半径均大于所述最大阈值;

根据所述覆盖区域半径,获得基站与覆盖区域内的路段的多等级映射关系,每一个覆盖区域半径对应一个等级映射关系。

4. 根据权利要求1所述的方法,其特征在于,所述对所述手机信令数据进行清洗,得到清洗后的信令数据,具体包括:

过滤无效数据和冗余数据;所述无效数据包括字段缺失的数据和单据状态为失败的数据;所述冗余数据包括重复数据和连续处于同一基站的数据序列;

过滤乒乓数据;所述乒乓数据为:经过第一基站后,在设定阈值时间内经过其他基站再次返回第一基站的信令数据;

过滤漂移数据;所述漂移数据为经过第一基站后依次到达第二基站和第三基站的信令数据,且第一基站、第二基站之间的距离与时间的比值大于设定比值阈值,第三基站与第一基站之间的距离小于第一设定距离阈值;

分割中断信令数据。

5. 根据权利要求1所述的方法,其特征在于,所述根据所述用户的所有停留点,确定所述用户的多段轨迹,具体包括:

将所述用户的所有停留点按照时间顺序排列;

对于第*i*段轨迹,将第*i*个停留点作为第*i*段轨迹的起始点,将第*i*+1个停留点作为第*i*段轨迹的结束点,提取第*i*段轨迹,依次对所述用户所有的信令数据进行分割,得到所述用户的所有轨迹。

6. 根据权利要求1所述的方法,其特征在于,所述根据轨迹数据库中的历史轨迹的参数,确定所述待匹配轨迹集合中每段待匹配轨迹对应的相似历史轨迹,具体包括:

获取待匹配轨迹中待匹配的轨迹段;所述待匹配的轨迹段为由A基站指向B基站的轨迹段,所述A基站和B基站为所述待匹配轨迹中连续的两个基站,且A基站和B基站之间的距离大于第二设定距离阈值;

获取所述轨迹数据库中先后经过所述A基站和B基站,且在所述A基站和B基站之间存在其他基站的历史轨迹;

计算所述历史轨迹中从A基站到B基站所用的时间与所述待匹配的轨迹段中从A基站到B基站所用的时间的比值;

判断所述比值是否不大于设定比值阈值,得到第三判断结果;

当所述第三判断结果表示所述比值不大于设定比值阈值时,将所述历史轨迹确定为所

述待匹配轨迹对应的相似历史轨迹；

当所述第三判断结果表示所述比值大于设定比值阈值时，获取下一个历史轨迹进行判断。

7. 根据权利要求6所述的方法，其特征在于，所述根据所述相似历史轨迹对所述待匹配轨迹进行插值，得到更新后的待匹配轨迹集合，具体包括：

将所述相似历史轨迹中从A基站到B基站的轨迹段，插入至对应的待匹配轨迹中待匹配的轨迹段，得到更新后的待匹配轨迹。

8. 根据权利要求1所述的方法，其特征在于，所述对于所述更新后的待匹配轨迹集合中的每一段轨迹，采用隐马尔可夫模型确定所述路段序列集合中的匹配路段序列，具体包括：

对于所述更新后的待匹配轨迹集合中的第j段轨迹，将所述第j段轨迹对应的基站序列确定为隐马尔可夫模型的观测状态；

将所述用户实际通过的路段确定为隐马尔可夫模型的隐藏状态；

采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的观测状态概率；

采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的状态转移概率；

采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的初始状态概率；

根据所述第j段轨迹对应的每一个路段序列的观测状态概率、状态转移概率和初始状态概率，确定所述第j段轨迹对应的每一个路段序列的总体概率；

将总体概率最大的路段序列确定为所述第j段轨迹的匹配路段序列。

9. 一种路网匹配的系统，其特征在于，所述系统包括：

映射关系获取模块，用于获取基站与信号覆盖区域内的路段映射关系；

手机信令数据获取模块，用于获取用户基于所有基站的手机信令数据；

清洗模块，用于对所述手机信令数据进行清洗，得到清洗后的信令数据；

停留点确定模块，用于根据所述清洗后的信令数据，确定所述用户的所有停留点；所述停留点为所述用户在一个基站的信号覆盖区域内停留时间超过停留阈值的基站坐标；

轨迹确定模块，用于根据所述用户的所有停留点，确定所述用户的多段轨迹；每段轨迹对应一个基站序列；

特征提取模块，用于对所述用户的每段轨迹进行特征提取，获得每段轨迹对应的轨迹特征；所述轨迹特征包括：轨迹长度、轨迹半径、切换基站数量、平均速度和出行时段；

出行方式确定模块，用于根据每段轨迹的轨迹特征，采用决策树算法确定每段轨迹对应的出行方式；所述出行方式包括：步行、非机动车、车辆和地铁；

轨迹提取模块，用于提取出行方式为车辆的所有轨迹，生成待匹配轨迹集合；所述待匹配轨迹集合中包括多段待匹配轨迹；

相似历史轨迹确定模块，用于根据轨迹数据库中的历史轨迹的参数，确定所述待匹配轨迹集合中每段待匹配轨迹对应的相似历史轨迹；

插值模块，用于根据所述相似历史轨迹对所述待匹配轨迹进行插值，得到更新后的待匹配轨迹集合；

路段序列转换模块,用于根据所述基站与信号覆盖区域内的路段映射关系,将所述更新后的待匹配轨迹集合中的每段轨迹对应的基站序列转换为路段序列,得到路段序列集合,每一段轨迹对应的基站序列对应多个路段序列;

匹配路段序列确定模块,用于对于所述更新后的待匹配轨迹集合中的每一段轨迹,采用隐马尔可夫模型确定所述路段序列集合中的匹配路段序列;

路网匹配结果确定模块,用于依次确定所述更新后的待匹配轨迹集合中所有轨迹对应的匹配路段序列,得到所述用户的路网匹配结果。

10. 根据权利要求9所述的系统,其特征在于,所述匹配路段序列确定模块具体包括:

观测状态确定单元,用于对于所述更新后的待匹配轨迹集合中的第j段轨迹,所述观测状态确定单元用于将所述第j段轨迹对应的基站序列确定为隐马尔可夫模型的观测状态;

隐藏状态确定单元,用于将所述用户实际通过的路段确定为隐马尔可夫模型的隐藏状态;

观测状态概率确定单元,用于采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的观测状态概率;

状态转移概率确定单元,用于采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的状态转移概率;

初始状态概率确定单元,用于采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的初始状态概率;

总体概率确定单元,用于根据所述第j段轨迹对应的每一个路段序列的观测状态概率、状态转移概率和初始状态概率,确定所述第j段轨迹对应的每一个路段序列的总体概率;

匹配路段序列确定单元,用于将总体概率最大的路段序列确定为所述第j段轨迹的匹配路段序列。



## 一种路网匹配的方法及系统

### 技术领域

[0001] 本发明涉及交通领域,特别是涉及一种路网匹配的方法及系统。

### 背景技术

[0002] 路网匹配是轨迹数据处理的关键步骤之一,一般我们获取到的轨迹数据由于设备、信号、精度等原因,与用户的实际位置存在一定偏差,通过路网匹配算法能够将获取的轨迹数据匹配到实际的路网中,确定用户的路网数据,有利于后续对城市交通、人口分布等问题的研究。目前大部分的研究都是基于车辆 GPS数据进行路网匹配,如出租车安装的GPS设备上传到服务器的数据、手机软件上传的GPS数据。虽然这些数据能够解决一定的问题,但是,仍然存在一定的不足,如数据倾斜、覆盖面小等,因此,路网匹配的效果不够理想。

### 发明内容

[0003] 本发明的目的是提供一种路网匹配的方法及系统,以提高路网匹配的覆盖范围,提高路网匹配精度。

[0004] 为实现上述目的,本发明提供了如下方案:

[0005] 一种路网匹配的方法,所述方法包括:

[0006] 获取基站与覆盖区域内的路段映射关系;

[0007] 获取用户基于所有基站的手机信令数据;

[0008] 对所述手机信令数据进行清洗,得到清洗后的信令数据;

[0009] 根据所述清洗后的信令数据,确定所述用户的所有停留点;所述停留点为所述用户在一个基站的信号覆盖区域内停留时间超过停留阈值的基站坐标;

[0010] 根据所述用户的所有停留点,确定所述用户的多段轨迹;每段轨迹对应一个基站序列;

[0011] 对所述用户的每段轨迹进行特征提取,获得每段轨迹对应的轨迹特征;所述轨迹特征包括:轨迹长度、轨迹半径、切换基站数量、平均速度和出行时段;

[0012] 根据每段轨迹的轨迹特征,采用决策树算法确定每段轨迹对应的出行方式;所述出行方式包括:步行、非机动车、车辆和地铁;

[0013] 提取出行方式为车辆的所有轨迹,生成待匹配轨迹集合;所述待匹配轨迹集合中包括多段待匹配轨迹;

[0014] 根据轨迹数据库中的历史轨迹的参数,确定所述待匹配轨迹集合中每段待匹配轨迹对应的相似历史轨迹;

[0015] 所述根据所述相似历史轨迹对所述待匹配轨迹进行插值,得到更新后的待匹配轨迹集合;

[0016] 根据所述基站与信号覆盖区域内的路段映射关系,将所述更新后的待匹配轨迹集合中的每段轨迹对应的基站序列转换为路段序列,得到路段序列集合,每一段轨迹对应的基站序列对应多个路段序列;

[0017] 对于所述更新后的待匹配轨迹集合中的每一段轨迹,采用隐马尔可夫模型确定所述路段序列集合中的匹配路段序列;

[0018] 依次确定所述更新后的待匹配轨迹集合中所有轨迹对应的匹配路段序列,得到所述用户的路网匹配结果。

[0019] 可选的,所述获取基站与信号覆盖区域内的路段映射关系,之前还包括:

[0020] 获取最大阈值和最小阈值;

[0021] 判断第二基站与第一基站的距离是否小于最小阈值,得到第一判断结果;

[0022] 当所述第一判断结果表示第二基站与第一基站的距离不小于最小阈值时,判断第三基站与第一基站的距离是否小于最小阈值;

[0023] 当所述第一判断结果表示第二基站与第一基站的距离小于最小阈值时,将所述第二基站合并至第一基站集合;所述第一基站集合中包括第一基站;

[0024] 确定所述第一基站集合的中心位置的坐标,所述中心位置坐标的纬度为所述第一基站集合中所有基站的纬度的平均值,所述中心位置坐标的经度为所述第一基站集合中所有基站的经度的平均值;

[0025] 判断所述第一基站集合中每个基站距离所述中心位置的距离是否均小于最大阈值,得到第二判断结果;

[0026] 当所述第二判断结果表示所述第一基站集合中每个基站距离所述中心位置的距离不都小于最大阈值时,将所述第二基站从所述第一基站集合中移除,判断第三基站与第一基站的距离是否小于最小阈值;

[0027] 当所述第二判断结果表示所述第一基站集合中每个基站距离所述中心位置的距离均小于最大阈值时,判断第三基站与第二基站的距离是否小于最小阈值;

[0028] 直至遍历所有的基站,得到更新后的第一基站集合;

[0029] 依次得到所有的基站集合;

[0030] 确定每个基站集合的中心位置的坐标;

[0031] 将每个基站集合的中心位置的坐标确定为更新后的第一基站的坐标,依次得到所有更新后的基站的坐标。

[0032] 可选的,所述获取基站与覆盖区域内的路段映射关系,具体包括:

[0033] 获取多个设定的覆盖区域半径;每个覆盖区域半径均大于所述最大阈值;

[0034] 根据所述覆盖区域半径,获得基站与覆盖区域内的路段的多等级映射关系,每一个覆盖区域半径对应一个等级映射关系。

[0035] 可选的,所述对所述手机信令数据进行清洗,得到清洗后的信令数据,具体包括:

[0036] 过滤无效数据和冗余数据;所述无效数据包括字段缺失的数据和单据状态为失败的数据;所述冗余数据包括重复数据和连续处于同一基站的数据序列;

[0037] 过滤乒乓数据;所述乒乓数据为:经过第一基站后,在设定阈值时间内经过其他基站再次返回第一基站的信令数据;

[0038] 过滤漂移数据;所述漂移数据为经过第一基站后依次到达第二基站和第三基站的信令数据,且第一基站、第二基站之间的距离与时间的比值大于设定比值阈值,第三基站与第一基站之间的距离小于第一设定距离阈值。

[0039] 分割中断信令数据。

- [0040] 可选的,所述根据所述用户的所有停留点,确定所述用户的多段轨迹,具体包括:
- [0041] 将所述用户的所有停留点按照时间顺序排列;
- [0042] 对于第i段轨迹,将第i个停留点作为第i段轨迹的起始点,将第i+1个停留点作为第i段轨迹的结束点,提取第i段轨迹,依次对所述用户所有的信令数据进行分割,得到所述用户的所有轨迹。
- [0043] 可选的,所述根据轨迹数据库中的历史轨迹的参数,确定所述待匹配轨迹集合中每段待匹配轨迹对应的相似历史轨迹,具体包括:
- [0044] 获取待匹配轨迹中待匹配的轨迹段;所述待匹配的轨迹段为由A基站指向B基站的轨迹段,所述A基站和B基站为所述待匹配轨迹中连续的两个基站,且A基站和B基站之间的距离大于第二设定距离阈值;
- [0045] 获取所述轨迹数据库中先后经过所述A基站和B基站,且在所述A基站和B基站之间存在其他基站的历史轨迹;
- [0046] 计算所述历史轨迹中从A基站到B基站所用的时间与所述待匹配的轨迹段中从A基站到B基站所用的时间的比值;
- [0047] 判断所述比值是否不大于设定比值阈值,得到第三判断结果;
- [0048] 当所述第三判断结果表示所述比值不大于设定比值阈值时,将所述历史轨迹确定为所述待匹配轨迹对应的相似历史轨迹;
- [0049] 当所述第三判断结果表示所述比值大于设定比值阈值时,获取下一个历史轨迹进行判断。
- [0050] 可选的,所述根据所述相似历史轨迹对所述待匹配轨迹进行插值,具体包括:
- [0051] 将所述相似历史轨迹中从A基站到B基站的轨迹段,插入至对应的待匹配轨迹中待匹配的轨迹段,得到更新后的待匹配轨迹。
- [0052] 可选的,所述对于所述更新后的待匹配轨迹集合中的每一段轨迹,采用隐马尔可夫模型确定所述路段序列集合中的匹配路段序列,具体包括:
- [0053] 对于所述更新后的待匹配轨迹集合中的第j段轨迹,将所述第j段轨迹对应的基站序列确定为隐马尔可夫模型的观测状态;
- [0054] 将所述用户实际通过的路段确定为隐马尔可夫模型的隐藏状态;
- [0055] 采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的观测状态概率;
- [0056] 采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的状态转移概率;
- [0057] 采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的初始状态概率;
- [0058] 根据所述第j段轨迹对应的每一个路段序列的观测状态概率、状态转移概率和初始状态概率,确定所述第j段轨迹对应的每一个路段序列的总体概率;
- [0059] 将总体概率最大的路段序列确定为所述第j段轨迹的匹配路段序列。
- [0060] 一种路网匹配的系统,所述系统包括:
- [0061] 映射关系获取模块,用于获取基站与覆盖区域内的路段映射关系;
- [0062] 手机信令数据获取模块,用于获取用户基于所有基站的手机信令数据;



- [0063] 清洗模块,用于对所述手机信令数据进行清洗,得到清洗后的信令数据;
- [0064] 停留点确定模块,用于根据所述清洗后的信令数据,确定所述用户的所有停留点;所述停留点为所述用户在一个基站的信号覆盖区域内停留时间超过停留阈值的基站坐标;
- [0065] 轨迹确定模块,用于根据所述用户的所有停留点,确定所述用户的多段轨迹;每段轨迹对应一个基站序列;
- [0066] 特征提取模块,用于对所述用户的每段轨迹进行特征提取,获得每段轨迹对应的轨迹特征;所述轨迹特征包括:轨迹长度、轨迹半径、切换基站数量、平均速度和出行时段;
- [0067] 出行方式确定模块,用于根据每段轨迹的轨迹特征,采用决策树算法确定每段轨迹对应的出行方式;所述出行方式包括:步行、非机动车、车辆和地铁;
- [0068] 轨迹提取模块,用于提取出行方式为车辆的所有轨迹,生成待匹配轨迹集合;所述待匹配轨迹集合中包括多段待匹配轨迹;
- [0069] 相似历史轨迹确定模块,用于根据轨迹数据库中的历史轨迹的参数,确定所述待匹配轨迹集合中每段待匹配轨迹对应的相似历史轨迹;
- [0070] 插值模块,用于所述根据所述相似历史轨迹对所述待匹配轨迹进行插值,得到更新后的待匹配轨迹集合;
- [0071] 路段序列转换模块,用于根据所述基站与信号覆盖区域内的路段映射关系,将所述更新后的待匹配轨迹集合中的每段轨迹对应的基站序列转换为路段序列,得到路段序列集合,每一段轨迹对应的基站序列对应多个路段序列;
- [0072] 匹配路段序列确定模块,用于对于所述更新后的待匹配轨迹集合中的每一段轨迹,采用隐马尔可夫模型确定所述路段序列集合中的匹配路段序列;
- [0073] 路网匹配结果确定模块,用于依次确定所述更新后的待匹配轨迹集合中所有轨迹对应的匹配路段序列,得到所述用户的路网匹配结果。
- [0074] 可选的,所述匹配路段序列确定模块具体包括:
- [0075] 观测状态确定单元,对于所述更新后的待匹配轨迹集合中的第j段轨迹,所述观测状态确定单元用于将所述第j段轨迹对应的基站序列确定为隐马尔可夫模型的观测状态;
- [0076] 隐藏状态确定单元,用于将所述用户实际通过的路段确定为隐马尔可夫模型的隐藏状态;
- [0077] 观测状态概率确定单元,用于采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的观测状态概率;
- [0078] 状态转移概率确定单元,用于采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的状态转移概率;
- [0079] 初始状态概率确定单元,用于采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的初始状态概率;
- [0080] 总体概率确定单元,用于根据所述第j段轨迹对应的每一个路段序列的观测状态概率、状态转移概率和初始状态概率,确定所述第j段轨迹对应的每一个路段序列的总体概率;
- [0081] 匹配路段序列确定单元,用于将总体概率最大的路段序列确定为所述第j段轨迹的匹配路段序列。
- [0082] 根据本发明提供的具体实施例,本发明公开了以下技术效果:

[0083] 本发明采用手机信令数据作为路网匹配的信息来源,扩大了覆盖范围。本发明提出的数据清洗方法,能够有效的解决手机信令的乒乓数据和漂移数据问题,极大地提高了手机信令定位的准确性;通过本发明的相似轨迹匹配算法能够找到大量相似路径,从而对轨迹点稀疏的轨迹进行插值,有利于提高路网匹配的准确性;本发明使用隐马尔可夫(HMM)模型与最短路径结合的方法进行路网匹配,通过建立基站与道路的多等级映射,并且将道路等级和道路直行程度两种因素添加到HMM模型中,实验结果表明本发明对粗粒度、低频率的手机信令数据实现路网匹配的效果较好,对城市道路规划以及交通管理方面具有积极意义。

## 附图说明

[0084] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0085] 图1为本发明路网匹配的方法的流程示意图;

[0086] 图2为本发明路网匹配的系统的结构示意图;

[0087] 图3为本发明具体实施方式中原始基站轨迹的示意图;

[0088] 图4为本发明具体实施方式中经过数据清洗与相似轨迹插值处理之后的轨迹;

[0089] 图5为本发明具体实施方式中最终的路网匹配结果。

## 具体实施方式

[0090] 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做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。

[0091] 为使本发明的上述目的、特征和优点能够更加明显易懂,下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0092] 图1为本发明路网匹配的方法的流程示意图。如图1所示,所述方法包括:

[0093] 步骤101:获取基站与覆盖区域内的路段映射关系。

[0094] 在确定映射关系之前,首先需要将距离较近的基站进行合并。具体如下:

[0095] 获取设定的最大阈值和最小阈值;

[0096] 判断第二基站与第一基站的距离是否小于最小阈值,得到第一判断结果;

[0097] 当所述第一判断结果表示第二基站与第一基站的距离不小于最小阈值时,判断第三基站与第一基站的距离是否小于最小阈值;第三基站为不同于第一基站和第二基站的其它任意一个基站;

[0098] 当所述第一判断结果表示第二基站与第一基站的距离小于最小阈值时,将所述第二基站合并至第一基站集合;所述第一基站集合中包括第一基站;

[0099] 确定所述第一基站集合的中心位置的坐标,所述中心位置坐标的纬度为所述第一基站集合中所有基站的纬度的平均值,所述中心位置坐标的经度为所述第一基站集合中所

有基站的经度的平均值；

[0100] 判断所述第一基站集合中每个基站距离所述中心位置的距离是否均小于最大阈值,得到第二判断结果；

[0101] 当所述第二判断结果表示所述第一基站集合中每个基站距离所述中心位置的距离不都小于最大阈值时,将所述第二基站从所述第一基站集合中移除,判断第三基站与第一基站的距离是否小于最小阈值；

[0102] 当所述第二判断结果表示所述第一基站集合中每个基站距离所述中心位置的距离均小于最大阈值时,判断第三基站与第二基站的距离是否小于最小阈值；

[0103] 直至遍历所有的基站,得到更新后的第一基站集合；

[0104] 依次得到所有的基站集合；

[0105] 确定每个基站集合的中心位置的坐标；

[0106] 将每个基站集合的中心位置的坐标确定为更新后的第一基站的坐标,依次得到所有更新后的基站的坐标。

[0107] 将较近的基站进行合并后,得到所有的基站集合对应的新的基站,以新的基站作为后续信令数据处理的基础。

[0108] 确定基站与覆盖区域内的路段映射关系时,首先获取多个设定的覆盖区域半径；每个覆盖区域半径均大于所述最大阈值；根据所述覆盖区域半径,获得基站与覆盖区域内的路段的多等级映射关系,每一个覆盖区域半径对应一个等级映射关系。

[0109] 步骤102:获取用户基于所有基站的手机信令数据。

[0110] 每个城市中分布着很多手机信号基站,每个基站会记录与其连接的手机用户信息,每当手机用户从一个基站覆盖范围移动到另一个基站覆盖范围时都会产生一条记录。目前随着智能手机的普及,基本上每个人都有一部智能手机,因此手机信令数据具有接近全样本的高覆盖率。每个基站都有精确的经纬度信息,手机信令数据通过大区编号(LAC)和小区编号(CI)可以唯一关联一个基站,因此我们能够从信令数据中提取每个用户的移动轨迹,每条移动轨迹用一系列的基站表示。与普通的轨迹数据相比,手机信令数据具有用户数量广、覆盖率高、并且以被动方式采集数据无需用户主动上传,因此通过手机信令数据能够以接近全样本的角度,更好的研究城市交通、人口分布等热点问题。

[0111] 步骤103:对所述手机信令数据进行清洗,得到清洗后的信令数据。对手机信令数据进行清洗包括以下方面:

[0112] 1. 过滤无效数据和冗余数据:字段缺失的数据以及单据状态为失败的数据定义为无效数据;重复数据以及连续处于同一基站的序列视为冗余数据。将无效数据和冗余数据删除。

[0113] 2. 过滤乒乓数据:设定时间阈值 $T_p$ ,遍历用户的一条轨迹,如果一条轨迹经过基站A之后,在阈值时间 $T_p$ 之内经过若干个其他基站再次返回A点,则视为是一段乒乓切换数据,删除中间的基站序列。

[0114] 3. 过滤漂移数据:有时由于反射折射等原因,会导致用户连接的基站突然从A切换到一个很远的基站B,处理方法是,设定一个阈值V,当A、B两点的距离与时间的比值超过阈值V时,并且下一个位置又返回距离A较近的位置,则视为漂移数据,将B点删除。

[0115] 4. 分割中断信令数据:由于用户手机信号不良或关机等原因,会产生部分轨迹中

间较长一段时间数据缺失严重的现象,导致相邻两个基站距离很远,无法对其进行路网匹配,因此设定距离阈值D,如果一段轨迹中存在相邻两个基站间超过距离阈值D,则认为轨迹中断,在中断处分割成两段轨迹。

[0116] 步骤104:根据所述清洗后的信令数据,确定所述用户的所有停留点。所述停留点为所述用户在一个基站的信号覆盖区域内停留时间超过停留阈值的基站坐标。

[0117] 步骤105:根据所述用户的所有停留点,确定所述用户的多段轨迹。每段轨迹对应一个基站序列。具体如下:

[0118] 将所述用户的所有停留点按照时间顺序排列;

[0119] 对于第i段轨迹,将第i个停留点作为第i段轨迹的起始点,将第i+1个停留点作为第i段轨迹的结束点,提取第i段轨迹,依次对所述用户所有的信令数据进行分割,得到所述用户的所有轨迹。

[0120] 步骤106:对所述用户的每段轨迹进行特征提取,获得每段轨迹对应的轨迹特征。所述轨迹特征包括:轨迹长度、轨迹半径、切换基站数量、平均速度和出行时段。

[0121] 步骤107:根据每段轨迹的轨迹特征,采用决策树算法确定每段轨迹对应的出行方式。所述出行方式包括:步行、非机动车、车辆和地铁。

[0122] 步骤108:提取出行方式为车辆的所有轨迹,生成待匹配轨迹集合。所述待匹配轨迹集合中包括多段待匹配轨迹。

[0123] 步骤109:根据轨迹数据库中的历史轨迹的参数,确定所述待匹配轨迹集合中每段待匹配轨迹对应的相似历史轨迹。具体包括:

[0124] 获取待匹配轨迹中待匹配的轨迹段;所述待匹配的轨迹段为由A基站指向B基站的轨迹段,所述A基站和B基站为所述待匹配轨迹中连续的两个基站,且A基站和B基站之间的距离大于第二设定距离阈值;

[0125] 获取所述轨迹数据库中先后经过所述A基站和B基站,且在所述A基站和B基站之间存在其他基站的历史轨迹;

[0126] 计算所述历史轨迹中从A基站到B基站所用的时间与所述待匹配的轨迹段中从A基站到B基站所用的时间的比值;

[0127] 判断所述比值是否不大于设定比值阈值,得到第三判断结果;

[0128] 当所述第三判断结果表示所述比值不大于设定比值阈值时,将所述历史轨迹确定为所述待匹配轨迹对应的相似历史轨迹;

[0129] 当所述第三判断结果表示所述比值大于设定比值阈值时,获取下一个历史轨迹进行判断。

[0130] 步骤1010:所述根据所述相似历史轨迹对所述待匹配轨迹进行插值,得到更新后的待匹配轨迹集合。具体的,将所述相似历史轨迹中从A基站到B基站的轨迹段,插入至对应的待匹配轨迹中待匹配的轨迹段,得到更新后的待匹配轨迹。

[0131] 步骤1011:根据所述基站与信号覆盖区域内的路段映射关系,将所述更新后的待匹配轨迹集合中的每段轨迹对应的基站序列转换为路段序列,得到路段序列集合。每一段轨迹对应的基站序列对应多个路段序列。按照步骤101中建立的基站-路网映射关系进行转换,转换时根据相邻基站距离选择不同等级的映射关系。最终将一条基站序列转换成路段集合的序列。

[0132] 步骤1012:对于所述更新后的待匹配轨迹集合中的每一段轨迹,采用隐马尔可夫模型确定所述路段序列集合中的匹配路段序列。隐马尔可夫模型主要包含5个要素:2个状态集合和3个概率矩阵,对于所述更新后的待匹配轨迹集合中的第j段轨迹,确定匹配路段序列的过程具体如下:

[0133] 1.观测状态O:将所述第j段轨迹对应的基站序列确定为隐马尔可夫模型的观测状态。

[0134] 2.隐藏状态S:将所述用户实际通过的路段确定为隐马尔可夫模型的隐藏状态。

[0135] 3.观测状态概率B:采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的观测状态概率。基站与路段最近点的距离的高斯分布作为观测状态概率,本发明中以 $\mu=0$ , $\sigma$ 为1(或道路集合到基站最大距离)的高斯分布模型计算观测概率。观测概率计算公式如下:

$$[0136] \quad b_j(k) = P(o_t = o_k | s_t = s_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\|o_t - s_j\|^2}{2\sigma^2}}$$

$$[0137] \quad \sigma = \begin{cases} 1, \max(\|o_t - s_j\|) = 0 \\ \max(\|o_t - s_j\|), \max(\|o_t - s_j\|) \neq 0 \end{cases}$$

[0138] 其中, $b_j(k)$ 代表第k个基站 $o_k$ 与其对应路段集合中第j个路段 $s_j$ 相匹配的概率值; $o_t$ 、 $s_t$ 分别代表t时刻用户所连接的基站和待匹配的路段; $\|o_t - s_j\|$ 代表基站与路段地最短距离; $\sigma$ 当基站与路段集合中所有路段最大距离等于0时取1,否则取最大距离。

[0139] 4.状态转移概率A:采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的状态转移概率。现有的研究与发明大都针对采样的GPS数据,而由于手机信令数据的特殊性,相邻两个基站往往间隔较远,其所对应的路段无法直接相连。因此本发明在构建概率转移矩阵时,首先要获取两个路段的最短路径。主要考虑方向和加权最短路径长度因素。

[0140] 加权最短路径采用Dijkstra算法,权值计算公式如下:

$$[0141] \quad D(m \rightarrow n) = \begin{cases} (1 + \omega_1 * \text{direct}(m \rightarrow n) + \omega_2 * \text{level}(m \rightarrow n)) * l_n, m、n \text{ 连通} \\ \infty, m、n \text{ 不连通} \end{cases}$$

$$[0142] \quad \text{direct}(m \rightarrow n) = \begin{cases} 1, m \text{ 到 } n \text{ 转弯} \\ 0, m \text{ 到 } n \text{ 直行} \end{cases}$$

$$[0143] \quad \text{level}(m \rightarrow n) = \begin{cases} 1, m \text{ 到 } n \text{ 道路等级切换} \\ 0, m \text{ 到 } n \text{ 道路等级不变} \end{cases}$$

[0144] 其中, $D(m \rightarrow n)$ 表示路段m到n的权值; $\text{direct}(m \rightarrow n)$ 在发生转弯时为1,否则为0; $\text{level}(m \rightarrow n)$ 在道路等级发生变换时为1,不变为0; $\omega_1, \omega_2$ 为前二者的权重; $l_n$ 为路段n的长度。通过Dijkstra得到路段 $s_i \rightarrow s_j$ 的最短路径L。

[0145] 得到最短路径之后计算 $s_i \rightarrow s_j$ 的转移概率,本专利主要通过角度、最短路径长度、转弯次数和道路等级切换次数几个因素确定转移概率,具体计算公式如下所示:

$$[0146] \quad a_{ij} = P(s_{t+1} = s_j | s_t = s_i) = F_\theta(s_i \rightarrow s_j) * F_d(s_i \rightarrow s_j) * F_c(s_i \rightarrow s_j)$$

[0147] 其中,  $t$ 时刻基站对应路段 $s_i$ ,  $t+1$ 时刻基站对应路段为 $s_j$ ,  $a_{ij}$ 表示 $s_i$ 到 $s_j$ 的转移概率;  $F_\theta(s_i \rightarrow s_j)$ 表示余弦相似度;  $F_d(s_i \rightarrow s_j)$ 表示两个相邻基站距离与加权最短路径的比值;  $F_c(s_i \rightarrow s_j)$ 表示转弯次数与道路等级切换次数的正态分布。

$$[0148] \quad F_\theta(s_i \rightarrow s_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{||\cos\theta - \cos\varphi||^2}{2}}$$

$$[0149] \quad F_d(s_i \rightarrow s_j) = \frac{||o_t - o_{t+1}||}{length(L)}$$

$$[0150] \quad F_c(s_i \rightarrow s_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{changedirect(L) * changelevel(L)}{2 * num(L)^2}}$$

[0151] 5、初始状态概率 $\pi$ :采用所述隐马尔可夫模型确定所述第 $j$ 段轨迹对应的每一个路段序列的初始状态概率,第一个基站所对应路段的概率,使用基站的观测状态概率来表示。

[0152] 最后,根据所述第 $j$ 段轨迹对应的每一个路段序列的观测状态概率、状态转移概率和初始状态概率,确定所述第 $j$ 段轨迹对应的每一个路段序列的总体概率,例如,将每一个路段序列的观测状态概率、状态转移概率和初始状态概率相乘,得到总体概率。将总体概率最大的路段序列确定为所述第 $j$ 段轨迹的匹配路段序列,即每个基站转换为一个映射路段与加权最短路径路段序列,完成所有基站的匹配,最终形成一条连续的路段序列,得到最终匹配结果。

[0153] 步骤1013:依次确定所述更新后的待匹配轨迹集合中所有轨迹对应的匹配路段序列,得到所述用户的路网匹配结果。

[0154] 图2为本发明路网匹配的系统的结构示意图。如图2所示,所述系统包括:

[0155] 映射关系获取模块201,用于获取基站与覆盖区域内的路段映射关系;

[0156] 手机信令数据获取模块202,用于获取用户基于所有基站的手机信令数据;

[0157] 清洗模块203,用于对所述手机信令数据进行清洗,得到清洗后的信令数据;

[0158] 停留点确定模块204,用于根据所述清洗后的信令数据,确定所述用户的所有停留点;所述停留点为所述用户在一个基站的信号覆盖区域内停留时间超过停留阈值的基站坐标;

[0159] 轨迹确定模块205,用于根据所述用户的所有停留点,确定所述用户的多段轨迹;每段轨迹对应一个基站序列;

[0160] 特征提取模块206,用于对所述用户的每段轨迹进行特征提取,获得每段轨迹对应的轨迹特征;所述轨迹特征包括:轨迹长度、轨迹半径、切换基站数量、平均速度和出行时段;

[0161] 出行方式确定模块207,用于根据每段轨迹的轨迹特征,采用决策树算法确定每段轨迹对应的出行方式;所述出行方式包括:步行、非机动车、车辆和地铁;

[0162] 轨迹提取模块208,用于提取出行方式为车辆的所有轨迹,生成待匹配轨迹集合;所述待匹配轨迹集合中包括多段待匹配轨迹;

[0163] 相似历史轨迹确定模块209,用于根据轨迹数据库中的历史轨迹的参数,确定所述



待匹配轨迹集合中每段待匹配轨迹对应的相似历史轨迹；

[0164] 插值模块2010,用于所述根据所述相似历史轨迹对所述待匹配轨迹进行插值,得到更新后的待匹配轨迹集合；

[0165] 路段序列转换模块2011,用于根据所述基站与信号覆盖区域内的路段映射关系,将所述更新后的待匹配轨迹集合中的每段轨迹对应的基站序列转换为路段序列,得到路段序列集合,每一段轨迹对应的基站序列对应多个路段序列；

[0166] 匹配路段序列确定模块2012,用于对于所述更新后的待匹配轨迹集合中的每一段轨迹,采用隐马尔可夫模型确定所述路段序列集合中的匹配路段序列；

[0167] 路网匹配结果确定模块2013,用于依次确定所述更新后的待匹配轨迹集合中所有轨迹对应的匹配路段序列,得到所述用户的路网匹配结果。

[0168] 其中,匹配路段序列确定模块2012具体包括：

[0169] 观测状态确定单元,对于所述更新后的待匹配轨迹集合中的第j段轨迹,所述观测状态确定单元用于将所述第j段轨迹对应的基站序列确定为隐马尔可夫模型的观测状态；

[0170] 隐藏状态确定单元,用于将所述用户实际通过的路段确定为隐马尔可夫模型的隐藏状态；

[0171] 观测状态概率确定单元,用于采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的观测状态概率；

[0172] 状态转移概率确定单元,用于采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的状态转移概率；

[0173] 初始状态概率确定单元,用于采用所述隐马尔可夫模型确定所述第j段轨迹对应的每一个路段序列的初始状态概率；

[0174] 总体概率确定单元,用于根据所述第j段轨迹对应的每一个路段序列的观测状态概率、状态转移概率和初始状态概率,确定所述第j段轨迹对应的每一个路段序列的总体概率；

[0175] 匹配路段序列确定单元,用于将总体概率最大的路段序列确定为所述第j段轨迹的匹配路段序列。

[0176] 下面给出本发明一个具体实施方式的实施过程：

[0177] 使用联通系统集成有限公司吉林省分公司提供的脱敏后的长春市基站数据和手机信令数据,基站数据包括基站的唯一标识BTS\_ID、LAC(大区编号)、CI(小区编号)、经度和纬度,如表1所示。手机信令数据包括用户ID、连接基站的时间、LAC(大区编号)、CI(小区编号)、事件类型,如表2所示。两个表通过LAC和CI关联,每条信令通过LAC和CI唯一确定连接的基站,每当手机用户发生位置更新、开关机或接打电话时就会产生一条记录,将一个用户的记录按时间顺序排列则构成了一个人的出行轨迹。

[0178] 表1基站数据样例

[0179]

| JID   | LAC   | CI    | 经度          | 纬度          |
|-------|-------|-------|-------------|-------------|
| 12679 | 45831 | 10231 | 125.3330315 | 43.876557   |
| 12677 | 45831 | 24932 | 125.33183   | 43.8784815  |
| 10292 | 37128 | 63563 | 125.3306003 | 43.87540033 |

[0180] 表2手机信令数据样例

|        |                                     |                |       |       |      |
|--------|-------------------------------------|----------------|-------|-------|------|
| [0181] | ID                                  | 时间戳            | LAC   | CI    | 事件类型 |
|        | 725E781255570F39642906102C81E70B213 | 20170705025954 | 45831 | 10231 | 位置更新 |
|        | C09A4DECDCAF15DC969AB79F01B         | 20170705030028 | 45831 | 24932 | 开关机  |
|        | 624DFC89E314632CE52AFA0C6A738E20    | 20170705025956 | 37128 | 63563 | 呼叫   |

[0182] 第一步:基站数据处理以及手机信令数据清洗

[0183] (一) 基站数据处理

[0184] 1. 设定最大阈值500米和最小阈值100米,将基站距离小于100米且合并后每个基站距离基站集合的中心位置不超过500米的基站进行合并,并将基站集合的中心位置作为合并后的基站位置坐标。

[0185] 2. 分别以100米、300米为半径,建立基站与圆形区域内覆盖的路段多等级映射关系。

[0186] (二) 信令数据清洗

[0187] 1. 过滤无效数据和冗余数据:将字段缺失、单据状态为失败、重复数据以及连续处于同一基站数据删除。

[0188] 2. 过滤乒乓数据:设定时间阈值为10分钟,删除在阈值时间之内经过若干个其他基站再次返回同一个基站的序列。

[0189] 3. 过滤漂移数据:设定阈值120,当A、B两点的距离与时间的比值超过阈值时,并且下一个位置又返回距离A较近的位置,则视为漂移数据,将B点删除。

[0190] 4. 分割中断轨迹:设定距离阈值1000米,如果一段轨迹中存在相邻两个基站间超过距离阈值,则认为轨迹中断,在中断处分割成两段轨迹。

[0191] 第二步:用户O、D点提取以及轨迹特征提取

[0192] 1. O、D点提取:当用户在一个基站覆盖范围内停留时间超过1小时,则认为是一个停留点,找到第一个停留点记为第一个O点,然后找到下一个停留点以上一个停留点为O点,以当前停留点为D点,直到所有停留点都已经提取完毕,最终获得一系列O、D对。

[0193] 2. 轨迹分割:将每个用户的出行轨迹,按照提取的O、D对其进行分割,将每个用户一天的出行轨迹分割成多段轨迹。

[0194] 3. 轨迹特征提取:对每段轨迹提取特征,包括:轨迹长度、轨迹半径、切换基站数量、平均速度、出行时段等。

[0195] 第三步:出行方式判别并提取车辆出行轨迹

[0196] 1. 轨迹出行方式判别:根据提取的轨迹特征,对每段轨迹通过决策树算法对其出行方式进行判别,主要分为:步行、非机动车、车辆、地铁。

[0197] 2. 车辆轨迹提取:本发明只提取车辆的出行轨迹进行路网匹配。

[0198] 第四步:建立轨迹数据库,实现相似轨迹匹配。

[0199] 1. 建立轨迹数据库:

[0200] 建立数据库,轨迹表和经过基站表示例数据如表3和表4所示:

[0201] 表3轨迹表结构

[0202]

| GID | UID                                 | JIDs  |
|-----|-------------------------------------|---|
| 1   | 725E781255570F39642906102C81E70B213 | 10285, 12679, 12677, 10292, 12679, 10009, 10183,<br>10178, 10182, 10170 |
| 2   | C09A4DECDCFDCAF15DC969AB79F01B      | 10239, 12679, 10275, 10009, 10191, 10183, 10178,<br>10170, 10189        |
| 3   | 624DFC89E314632CE52AFA0C6A738E20    | 34, 76, 865, 895, 2343, 1232, 235, 976, 1246, 213                       |

[0203] 表4经过基站表结构

[0204]

| PID | JID   | GID | TIME           |
|-----|-------|-----|----------------|
| 1   | 10009 | 2   | 20170705082344 |
| 2   | 10191 | 2   | 20170705083118 |
| 3   | 10183 | 2   | 20170705084305 |

[0205] 2. 相似轨迹匹配, 根据历史轨迹以及同时段轨迹进行插值。

[0206] 以轨迹1为例进相似轨迹匹配, 其中基站10009和10183之间距离超过 100米, 则

[0207] 从数据库中进行相似轨迹选取:

[0208] (1) 首先选择先后经过基站10009、10183, 并且在两点之间存在其他基站的轨迹;

[0209] (2) 并且, 从A到B所用时间不超过待匹配轨迹A到B所用时间的  $(1 \pm 0.5)$  倍。

[0210] (3) 最终匹配轨迹为10009→10191→10183, 将10191插入到轨迹1中。

[0211] 3. 基站序列转换成路段集合序列: 对完成插值后的基站序列建立基站-路网映射关系进行转换。如图3、图4所示, 图3为本发明具体实施方式中原始基站轨迹的示意图, 图4为本发明具体实施方式中经过数据清洗与相似轨迹插值处理之后的轨迹。最终将一条基站ID序列转换成路段ID集合的序列。

[0212] 12679→{2621, 2622, 2623, 2624, 2625, 2627, 4874, 6142, 6143, 6153, 6154, 7084, 10635, 10636}

[0213] 10275→{24867, 4868, 4869, 6010, 6011, 6012, 6039, 6040, 13740, 15896, 17768}

[0214] 10009→{3394, 3395, 3396, 5629, 6019, 9928, 11964, 18748, 18970, 19946}

[0215] 10191→{4159, 8940, 12062, 16648, 16662, 18985}

[0216] 10183→{6056, 6066, 6067, 6717, 8564, 8565, 8566, 9175, 11919, 12095, 16648, 19516}

[0217] 10178→{7247, 7582, 8929, 8932, 8933, 8934, 8935, 9042, 9043, 9044, 18659, 18660, 18663}

[0218] 10170→{1654, 1676, 5923, 6508, 7354, 8539, 8673, 9019, 9020, 9026, 15237, 15241, 16139}

[0219] 第五步: 结合加权最短路径建立隐马尔科夫 (HMM) 模型, 对车辆出行轨迹进行路网匹配:

[0220] 1. 观测状态0: 待匹配基站序列 {12679, 10275, 10009, 10191, 10183, 10178,

10170}。

[0221] 2.隐藏状态S:该手机用户实际通过的路段为隐藏状态。

[0222] 3.观测状态概率B:

[0223] 计算每个基站与其映射集合中所有路段的最大距离 $\max(||o_t - s_j||)$ ,观测概率计算公式如下:

$$[0224] \quad b_j(k) = P(o_t = o_k | s_t = s_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{||o_t - s_j||^2}{2\sigma^2}}$$

$$[0225] \quad \sigma = \begin{cases} 1, \max(||o_t - s_j||) = 0 \\ \max(||o_t - s_j||), \max(||o_t - s_j||) \neq 0 \end{cases}$$

[0226] 4.状态转移概率

[0227] 利用加权最短路径Dijkstra算法计算转移概率,权值计算公式如下:

$$[0228] \quad D(m \rightarrow n) = \begin{cases} (1 + \omega_1 * \text{direct}(m \rightarrow n) + \omega_2 * \text{level}(m \rightarrow n)) * l_n, m、n \text{ 连通} \\ \infty, m、n \text{ 不连通} \end{cases}$$

$$[0229] \quad \text{direct}(m \rightarrow n) = \begin{cases} 1, m \text{ 到 } n \text{ 转弯} \\ 0, m \text{ 到 } n \text{ 直行} \end{cases}$$

$$[0230] \quad \text{level}(m \rightarrow n) = \begin{cases} 1, m \text{ 到 } n \text{ 道路等级切换} \\ 0, m \text{ 到 } n \text{ 道路等级不变} \end{cases}$$

[0231] 得到最短路径之后计算 $s_i \rightarrow s_j$ 的转移概率,本专利主要通过角度、最短路径长度、转弯次数和道路等级切换次数几个因素确定转移概率,具体计算公式如下所示:

$$[0232] \quad a_{ij} = P(s_{t+1} = s_j | s_t = s_i) = F_\theta(s_i \rightarrow s_j) * F_d(s_i \rightarrow s_j) * F_c(s_i \rightarrow s_j)$$

[0233] 其中,t时刻基站对应路段 $s_i$ ,t+1时刻基站对应路段为 $s_j$ , $a_{ij}$ 表示 $s_i$ 到 $s_j$ 的转移概率; $F_\theta(s_i \rightarrow s_j)$ 表示余弦相似度; $F_d(s_i \rightarrow s_j)$ 表示两个相邻基站距离与加权最短路径的比值; $F_c(s_i \rightarrow s_j)$ 表示转弯次数与道路等级切换次数的正态分布。

$$[0234] \quad F_\theta(s_i \rightarrow s_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{||\cos\theta - \cos\varphi||^2}{2}}$$

$$[0235] \quad F_d(s_i \rightarrow s_j) = \frac{||o_t - o_{t+1}||}{\text{length}(L)}$$

$$[0236] \quad F_c(s_i \rightarrow s_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\text{changedirect}(L) * \text{changelevel}(L)}{2 * \text{num}(L)^2}}$$

[0237] 5.初始状态概率 $\pi$ :第一个基站所对应路段的概率,使用基站的观测状态概率来表示。

[0238] 最终匹配结果:选择概率最大的路径作为最终匹配结果,如图5所示,图5为本发明具体实施方式中最终的路网匹配结果。

[0239] 本说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他

实施例的不同之处,各个实施例之间相同相似部分互相参见即可。对于实施例公开的系统而言,由于其与实施例公开的方法相对应,所以描述的比较简单,相关之处参见方法部分说明即可。

[0240] 本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处。综上所述,本说明书内容不应理解为对本发明的限制。

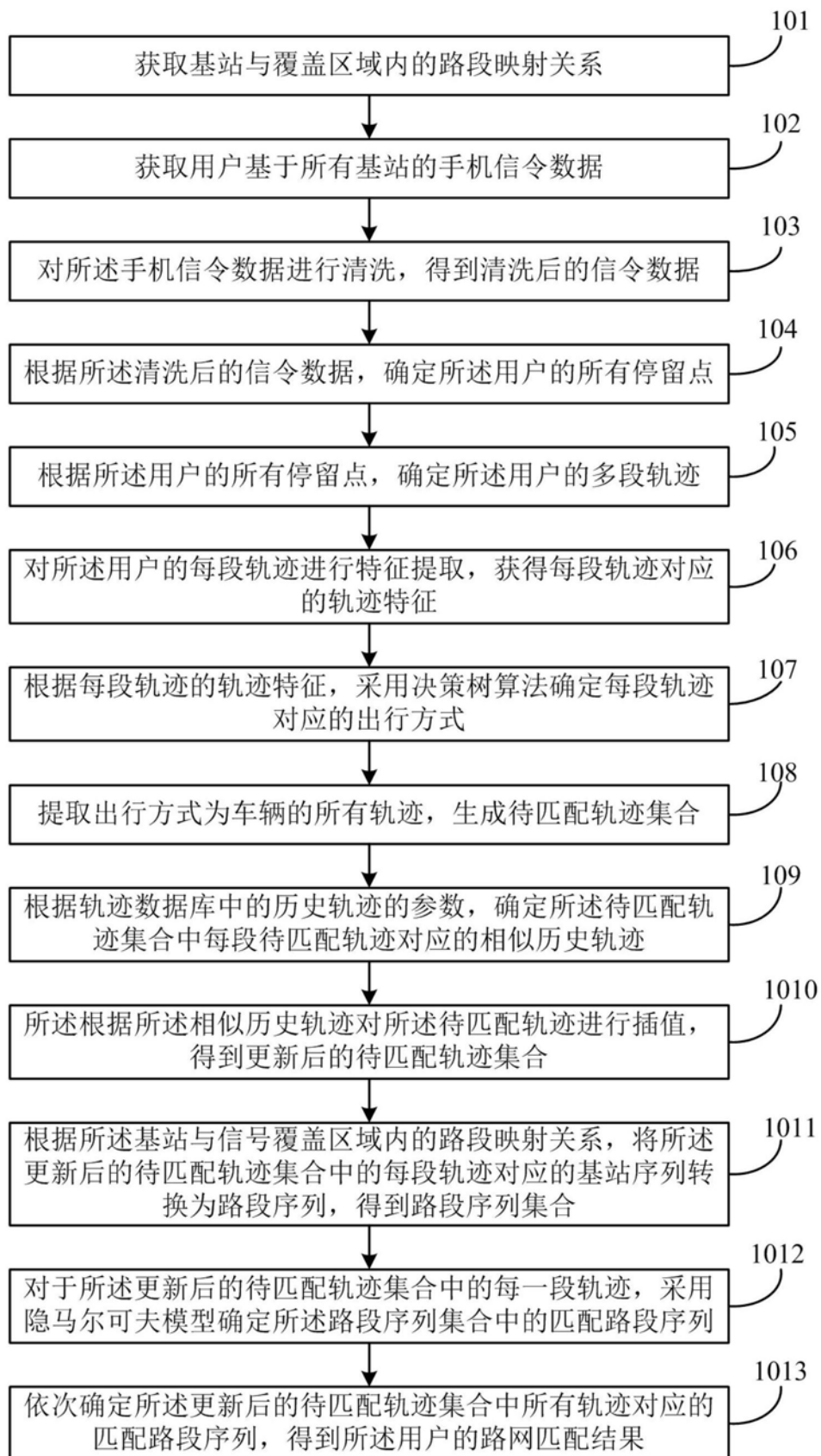


图1



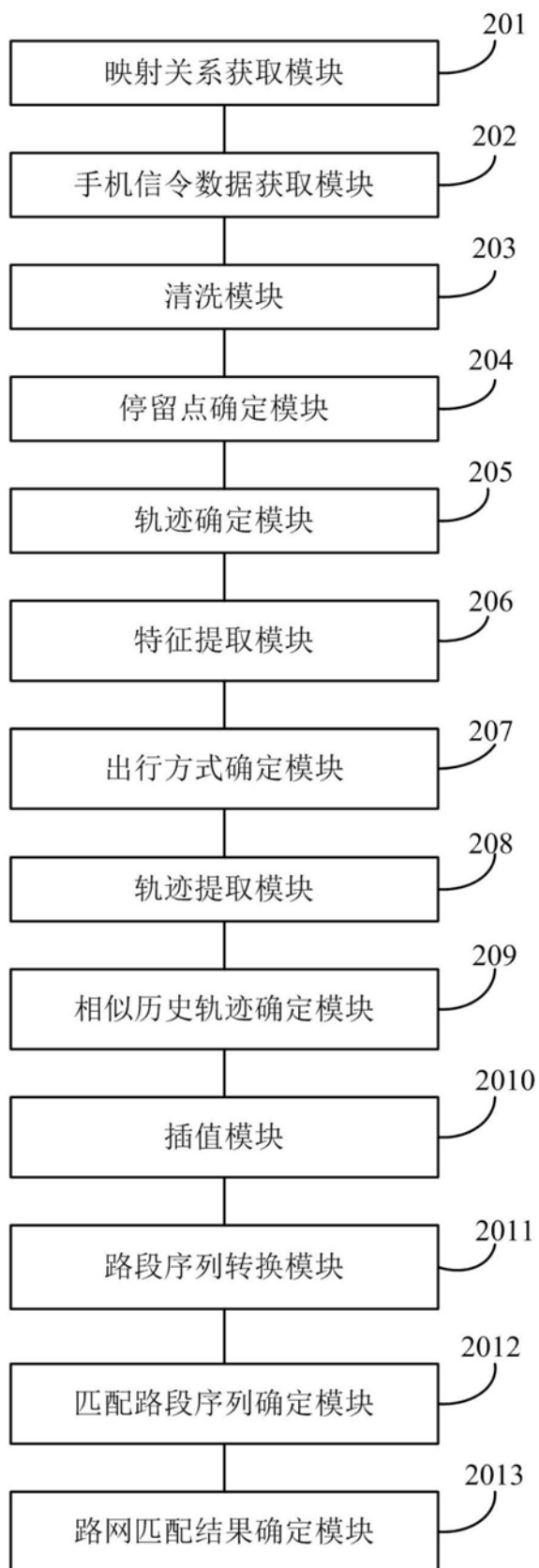


图2



图3



图4



图5