

# 一种基于轨迹大数据离线挖掘与在线实时监测的出租车异常轨迹检测算法

韩博洋, 汪兆洋, 金蓓弘

(中国科学院软件所软件工程技术研究开发中心, 北京 100190)

**摘要:**以防止出租车欺诈绕路为例,提出一种基于出租车GPS时空轨迹数据离线挖掘与在线实时检测相结合的异常轨迹检测算法,获得快速反馈实时检测的结果.首先,将路网地图进行网格化切分并编号,用Pathlet方法优化常用的以GPS点组成的轨迹序列,并将轨迹通过匹配、补全等处理变换为Pathlet序列.然后,从大量出租车历史数据中,获得轨迹的Pathlet序列,并聚类得到起点与终点之间正常的 $K$ 类轨迹.当实时轨迹需要被检测时,便与 $K$ 类正常轨迹进行匹配,只需计算两段Pathlet序列的编辑距离,并同时考量时间和空间两个维度设定合理阈值,判断是否抛出异常.最后,基于北京地区2011年3月到5月出租车GPS轨迹的真实数据集进行了大量实验,对比了相关工作,印证了所提出算法的有效性和高效性.

**关键词:**GPS轨迹;异常轨迹检测;Pathlet方法;时空数据挖掘

**中图分类号:**TP311      **文献标识码:**A      doi:10.3969/j.issn.0253-2778.2016.03.010

**引用格式:** HAN Boyang, WANG Zhaoyang, JIN Beihong. An anomaly detection algorithm for taxis based on trajectory data mining and online real-time monitoring[J]. Journal of University of Science and Technology of China, 2016, 46(3):247-252.

韩博洋, 汪兆洋, 金蓓弘. 一种基于轨迹大数据离线挖掘与在线实时监测的出租车异常轨迹检测算法[J]. 中国科学技术大学学报, 2016, 46(3):247-252.

## An anomaly detection algorithm for taxis based on trajectory data mining and online real-time monitoring

HAN Boyang, WANG Zhaoyang, JIN Beihong

(Research and Development Center of Software Engineering and Technology,  
Institute of Software Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** Taking the prevention of taxi frauds as a motivating example, an anomalous spatio-temporal trajectory detection method that combines offline mining and online detection was proposed. A city roadmap was partitioned into a grid based on the longitude and latitude, using Pathlet sequences to express taxi trajectories instead of the traditional GPS sequences. Then,  $K$ -racial classes' normal sequences were clustered in the same origin-destination pair from history data sets. The incoming online GPS data was transformed into Pathlet sequences and matched with  $K$ -racial classes' normal sequences. The distance was computed and scored. Distance along with spatial and temporal factors together forms the criterion for

收稿日期:2015-09-12;修回日期:2015-12-29

作者简介:韩博洋,男,1993年生,硕士生.研究方向:数据挖掘. E-mail: hanboyang15@otcaix.iscas.ac.cn

通讯作者:金蓓弘,博士/研究员. E-mail: beihong@iscas.ac.cn

determining anomalous taxi trajectories. Finally, based on the real taxi GPS data sets in Beijing area during March, 2011 to May, 2011, experimental results indicate that the proposed method is able to detect online anomalous trajectories efficiently and quickly.

**Key words:** GPS trajectory; anomalous trajectory detection; pathlet method; spatio-temporal data mining

## 0 引言

当前,各地出租车绕路现象仍十分常见.特别在北京这样的一线城市,外地流动人口量大,遇到对出行线路不熟悉的非本地乘客,存在出租车司机为了赚钱舍近求远的现象.本文以此为出发点开展轨迹异常检测研究.

目前已有面向异常检测的若干工作,文献[1]作为这个领域的综述,全面地介绍了该方向的大背景,总结归纳了常用的检测手段.针对本文研究的出租车时空轨迹异常检测方向,通过挖掘分析出租车轨迹数据<sup>[2]</sup>,并用于分析城市人群的流通情况.文献[3]挖掘出乘客上下车的热点区域,规划城市公共交通.文献[4]根据出租车轨迹提供的道路信息,规划和推荐行车路线,文献[5-6]进行实时交通状况进行分析,报告区域拥堵等异常情况.检测出租车绕路异常方面的研究已有若干工作<sup>[7-11]</sup>,与他们相比,本文使用全新的方法解决这个问题.

为了准确并高效地进行实时异常轨迹检测,我们首先对路网地图进行网格化切分并编号,使用 Pathlet 方法代替 GPS 序列表示轨迹,将轨迹通过匹配、补全等处理变化为 Pathlet 序列.从大量历史数据中聚类得到每一对起点与终点之间正常的  $K$  类轨迹.当实时新轨迹需要被检测时,便与同一起点终点对的  $K$  类正常轨迹进行匹配,计算两段 Pathlet 序列的编辑距离.由于在具体的一次行程中,起点和目的点之前的线路种类较多,司机们不一定都按照固定路线行驶,并且不同的司机有不同的驾车习惯、技能以及对路段的熟悉程度,因此片面地关注不同的行驶路线与轨迹形态并不能正确反映轨迹的异常程度,本文加入时间方面的考量,并设定合理阈值,判断是否抛出异常.

本文的主要贡献是:①提出了一种有效的筛选出租车欺诈行为的算法,通过结合离线挖掘和在线监测两种模式,实时检测出租车的欺诈绕路行为;②使用 Pathlet 方法表示每条出租车轨迹,建立 Pathlet 字典集.简化 GPS 点之间的匹配,可大幅度提升在线实时检测的效率.文中基于北京地区 2011

年 3 月到 5 月出租车 GPS 轨迹的真实数据集进行了实验验证.

## 1 问题分析和概念定义

本文的本质问题是将离线挖掘和在线检测紧密结合,通过离线挖掘和聚类处理大量出租车历史 GPS 点轨迹,并将 GPS 点轨迹变换为 Pathlet 序列,提升在线实时异常检测的效率,同时还要考虑时空要素和特殊事件或重大活动.下面我们先给出概念定义,然后分析本文着重求解的难点问题.

**路网地图:**  $L_a$  为经度,  $L_o$  为纬度,将地图划分为网格状后采用横纵坐标  $x, y$  进行编号,即路网  $G=(L_a, L_o)=(x, y)$ ;

**轨迹起点终点有序对:** 起点设为  $O$ , 终点设为  $D$ , 可得有序对  $(O, D)$ , 其中  $O, D$  均使用网格编号.

**Pathlet 字典集  $P$ :**  $P$  由不同 Pathlet  $p$  组成,  $p$  由一个或多个网格的连续序列组成,即  $[(x_1, y_1), (x_2, y_2)]$ .

**Pathletlist  $L$ :** 给定 Pathlet 字典集  $P$ ,  $P$  的子集即为 Pathletlist.

**轨迹  $t$ :** 单一轨迹由(车辆编号, 总时间, 总路程,  $O, D, L$ )组成. 轨迹集为  $T$ .

**原始出租车轨迹数据结构:** (时间, 车辆编号, 经度, 纬度, 载客标识, 速度, 方向角).

**轨迹相似度度量:** 编辑距离, 对于给定 Pathletlist  $L_1=[p_1, p_2, p_3]$  与 Pathletlist  $L_2=[p_2, p_4, p_5]$ , 计算将  $L_2$  变换为  $L_1$  所需所有步骤的得分总和即为编辑距离.

定义描述长度(description length, DL):

最大 Pathlet 集合  $\bar{P}=\bigcup_{t \in T} P(t)$ ;

轨迹  $t$  包含的所有 Pathlet  $|P(t)|=\frac{|t|(|t|+1)}{2}$ .

描述长度:  $DL(t, P)=\min_{P_{sub} \subset P} |P_{sub}|$ ,

$\exists P_{sub} \subset P, t=L_t$

**异常:** 当被检测轨迹与相同  $(O, D)$  对聚类的  $K$  条轨迹的编辑距离得分均大于阈值时, 认为其是异

常的.

实现该算法的难点主要有 Pathlet 字典集的获取、 $K$  类聚类中心的选取标准与阈值定义以及编辑距离方程的可靠性. 不失普遍性, 我们做了以下三点基本假设用以简化问题:

(I) 北京地区路网道路不发生改变(如修路、封路、或出现新路线);

(II) 预处理后的离线轨迹 Pathlet 集合均正确;

(III) 抛出异常的阈值不因外界条件而改变(如高峰时段, 节假日, 重大活动).

## 2 轨迹异常检测框架

如图 1 所示, 本文的异常检测基于离线挖掘与在线实时监测两部分相结合. 离线端主要包括: 轨迹的预处理, GPS 点轨迹向 Pathlet 序列转换, 根据  $(O, D)$  对进行聚类. 在线端主要包括: 与  $K$  类正常轨迹进行匹配, 计算两段 Pathlet 序列的编辑距离, 判断是否抛出异常.

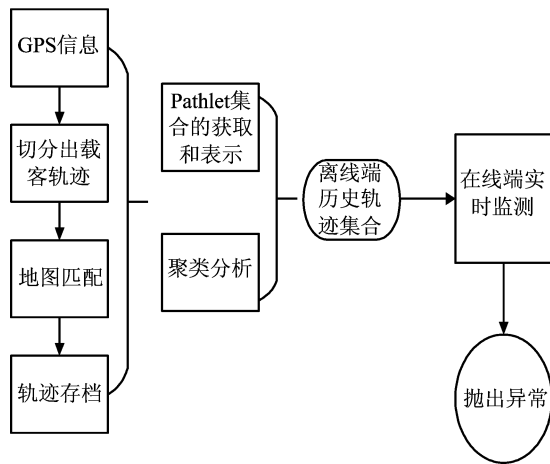


图 1 系统框架

Fig. 1 System framework

首先服务器端通过离线挖掘, 取得历史正常出租车轨迹的集合. 出租车在载客过程中, 乘客不断地通过 3G/4G 网络向服务器发送信息, 包括当前位置和时间信息, 此次载客轨迹的起始点位置和时间以及目的地信息, 进行在线实时检测. 当服务器收到信息后, 通过一系列变换操作, 与预先处理好的正常轨迹集做对比, 根据阈值判断是否抛出异常.

### 2.1 离线端轨迹预处理

轨迹预处理方面主要借鉴了已有工作<sup>[12]</sup>. 首先, 根据出租车轨迹中是否载有乘客的信息(有 0, 1

标识位)切分 GPS 轨迹, 串联相同车辆每段时间连续的 GPS 点, 得到出租车载客时的轨迹集(没乘客的轨迹点不能反映司机的正常行驶状态). 遇到格式有误或信息不全的轨迹, 视情况采取直接删除或提取有用信息并补充记录的处理方式.

第二步, 将北京市路网地图(图 2)划分成  $n * m$  个网格, 并使用坐标对其进行编号.

第三步, 将 GPS 点匹配到网格上, 再对轨迹做地图匹配, 完成轨迹预处理. 对每一条有效轨迹(一次上下客), 我们借用已有工作<sup>[12]</sup>, 先将其精确匹配到每一条街道上, 再将 GPS 轨迹点分别映射到网格中, 记录每条轨迹的网格序列.

由于不同时间段、不同日期类型(工作日、休息日)都会影响人们的出行方式, 会有不同的交通状况, 相同起点终点网格对的轨迹路线会出现变化, 因此需要考虑不同时段、不同日期类型的影响, 将时间维度加入起点终点网格对中. 最后使用临近网格补充方法形成连续的网格序列<sup>[13]</sup>.

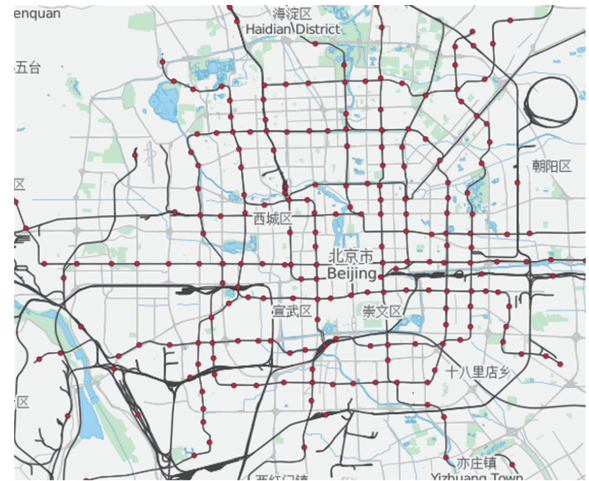


图 2 北京路网地图

Fig. 2 Beijing road network map

### 2.2 Pathlet 轨迹集合的获取与表示

做完轨迹预处理步骤后, 我们得到了每条轨迹网格序列的表示, 基于此再使用 Pathlet 方法来表示每一条轨迹. 我们使用有约束的整数线性规划方法来取得 Pathlet 字典集<sup>[14]</sup>, 在求解过程中, 目标有两点: ①得到数据量最小、最优化(一个 Pathlet  $p$  可以在尽可能多的轨迹中使用)的 Pathlet 字典集; ②表示轨迹  $T$  所用的 Pathletlist 平均长度最短.

首先, 定义动态规划方程

$$\min_{P \subset \bar{P}} \text{DL}(P) + \lambda \sum_{t \in T} \text{DL}(t, P)$$

$$s. t. \forall t \in T, \exists P_{\text{sub}} \in P, t = L_t.$$

变量  $\lambda$  为控制变量, 增大则表示轨迹的 Pathlet 个数减少, 但是也增大了 pathlet 字典的容量.

在实现本文目标时, 达到最优化的 Pathlet 字典集和 Pathletlist 长度对提升在线实时检测的效率都有很大的帮助, 所以此处控制变量设为 1.

将每一个 Pathlet 序列  $p$  用二进制指数  $x_p \in \{0, 1\}$  表示, 如果  $p \in \bar{P}$ , 则  $x_p = 1$ .

$$\text{可推出 } P = \{p \mid p \in \bar{P}, x_p = 1\}.$$

设二进制指数  $x_{t,p} \in \{0, 1\}$ , 如果在轨迹  $t$  中使用了  $p$ , 则  $x_{t,p} = 1$ .

由此做出等价变换, 变为求解整数方程.

$$\min \sum_{p \in \bar{P}} x_p + \lambda \sum_{t \in T} \sum_{p \in P(t)} x_{t,p}$$

$$s. t. x_{t,p} \leq x_p, \forall p \in P(t), t \in T.$$

通过缩放变换求解

$$\begin{aligned} \sum_{p \in \bar{P}} x_p + \lambda \sum_{t \in T} \sum_{p \in P(t)} x_{t,p} &= \\ \sum_{p \in \bar{P}} \sum_{t \in T(p)} \frac{x_p}{|T(p)|} + \lambda \sum_{t \in T} \sum_{p \in P(t)} x_{t,p} &\geq \\ \sum_{p \in \bar{P}} \sum_{t \in T(p)} \frac{x_{t,p}}{|T(p)|} + \lambda \sum_{t \in T} \sum_{p \in P(t)} x_{t,p} &= \end{aligned}$$

$$\sum_{t \in T} \sum_{p \in P(t)} \left( \lambda + \frac{1}{|T(p)|} \right) x_{t,p};$$

$$T(p) = \{t \mid t \in T, p \in P(t)\}.$$

则变为求下式最小值

$$\min_{x_{t,p} \in \{0,1\}} \sum_{p \in P(t)} \left( \lambda + \frac{1}{|T(p)|} \right) x_{t,p}.$$

至此, 可以通过动态规划求解, 将优化目标值  $t$  中的子集段看作  $v_i \sim v_j$ , 则  $t = v_1 \sim v_n$ , 记作  $f^*(v_i \sim v_j)$ . 则有

$$\begin{aligned} f^*(v_i \sim v_j) &= \min_{k \in \{i+1, \dots, j-1\}} (f^*(v_i \sim v_k) + \\ f^*(v_k \sim v_j)) &= \lambda + 1/|T(v_i v_{i+1})|, \end{aligned}$$

可通过回溯求解最优  $t$ .

### 2.3 Pathlet 轨迹集合聚类与在线异常监测

在得到 Pathlet 轨迹字典集后, 我们使用聚类分析, 得到每一个  $O-D$  点对之间的  $K$  类正常轨迹, 完成离线端的工作.

本文中聚类的规则是轨迹间相似度的度量, 这是实现轨迹匹配、轨迹聚类、轨迹模式发现、轨迹异

常检测等问题的基础. 针对本文轨迹 Pathlet 序列的特点以及在线检测的实时性, 我们在使用  $K$ -Means 进行聚类, 选择编辑距离作为度量标准. 在定义编辑距离得分方程和合理阈值时, 我们通过大量实验验证每种假设, 选取效果最理想的一组.

定义: 交换得分 1; 插入得分 2; 删除得分 2; 时间延长(每 10 分钟)得分 1. 当被检测轨迹与相同  $OD$  对聚类的  $k$  条轨迹的编辑距离得分均大于阈值时, 我们认为其是异常.

$K$ -Means 聚类分析具体流程如下.

Input: 路网  $G = (La, Lo) = (x, y)$ , Pathlet 字典集  $P$ , Pathletlist  $L$ , 单一轨迹  $t$ , 轨迹集  $T$

Output:  $(O, D, K$  类中心)

1 轨迹集  $T$  中随机挑选  $k$  条轨迹(挑选相同  $O, D$  点)

2 随机选择  $k$  条轨迹构成初始聚类中心集合  $C$

3 将  $T$  中每条轨迹与  $C$  中聚类中心计算编辑距离, 并标记类别

4 重新计算聚类中心集合  $C$ , 使用每一个类别的编辑距离均值

5 迭代至收敛

最终, 根据不同的  $(O, D)$  点对进行分类, 储存在离线端. 为了最大程度地提升在线端的检测效率, 我们为每一个聚类选取一个聚类中心, 即聚类中与所有轨迹距离最近的轨迹.

在出租车载客过程中, 乘客会不断地将当前的时间和位置信息发送给服务器. 当服务器收到数据, 新的 GPS 点到达时, 需要进行两步操作: ① GPS 序列到网格序列的转化; ② 每次获取 3~5 个网格(或路段)后匹配到一个 Pathlet, 形成这条轨迹的新的 Pathletlist. 随着服务器动态更新当前车辆的轨迹信息, 将根据  $(O, D)$  点对选取相应的聚类, 进行比对分析, 判断是否抛出异常.

聚类归属的判定主要有如下三种方式:

(I) 使用轨迹已有的 Pathlet 和聚类的中心进行编辑距离的计算.

(II) 使用已有 Pathlet + 从当前位置到  $D$  点的最优 Pathlet 合成最终轨迹和聚类的中心进行编辑距离的计算(使用这种方法的时候, 需要在上一步 clustering 的时候额外存储  $O-D$  之间每个用到的 Pathlet 到  $D$  点的最优路径, 最优指的是现有轨迹中走的最频繁的轨迹).

(III) 使用轨迹已有的 Pathlet, 并在聚类的中心中取出相应长度或者用与最后一个 Pathlet 点相同的 Pathlet 序列, 然后计算编辑距离.

我们选择第三种方法. 当服务器端根据对比聚类分析的结果, 向乘客端进行及时反馈, 还可以给相关部门发送异常通知.

### 3 结果分析与评价

本文的实验数据集来自 2011 年 3 月到 5 月北京地区真实出租车 GPS 轨迹数据, 详见表 1、2. 选择 70% 作为离线挖掘数据集, 30% 作为实时监测数据集. 并与 IBOAT (isolation-based online anomalous

trajectory detection) 算法、地标图算法进行对比.

表 1 实验数据集

Tab. 1 Experimental datasets

数据集	2011. 3-5 北京地区出租车 GPS 信息
总出租车数	13 592
有效天数	76
平局采样率	70
平均采样点间距	337

表 2 由(车辆编号, 总时间, 总路程,  $O, D, L$ )组成的单一轨迹

Tab. 2 Single trajectory by (vehicle number, total time, total distance,  $O, D, L$ )

时间	车辆编号	经度	纬度	速度	方向角	载客标示(1 代表载客)
2011/03/05 00:01:59	8613436732400	39.950 227	116.358 539	36	102	1
2011/03/05 17:58:56	8613466431361	39.976 983	116.464 012	771	219	1
2011/03/05 02:57:57	8613466444679	39.935 898	116.454 697	617	87	0

#### 3.1 实验结果和分析

与已有算法 IBOAT<sup>[8-10]</sup> 及地标图方法<sup>[5,13]</sup> 进行对比分析, 采取文献[12]中的实验评估数据, 主要参考指标为在线检测的反应时间  $T$  与异常报告的准确率  $P$  以及召回率  $R$ . 我们使用相同的轨迹数据集, 进行对比分析.

通过大量基于真实出租车 GPS 轨迹数据的实验评估本文提出的异常轨迹检测算法, 采用 Java 实现, 运行实验的 PC 机配置如下: CPU 型号 Intel 酷睿 i7 4850HQ, CPU 主频 2.3 GHz, 内存容量 16 GB, 硬盘容量 512 GB, SSD 固态硬盘, 操作系统 Windows 8.1. 具体比对结果如表 3 所示.

表 3 实验结果

Tab. 3 Experimental results

准确率	工作日	休息日	平均
Pathlet 方法	0.74	0.70	0.732
基于地标图的检测算法	0.75	0.60	0.72
IBOAT 算法	0.39	0.48	0.41
召回率	工作日	休息日	平均
Pathlet 方法	0.90	0.78	0.86
基于地标图的检测算法	0.89	0.78	0.85
IBOAT 算法	0.91	0.82	0.88
匹配速度	平均速度		
Pathlet 方法	1.7s		
基于地标图的检测算法	3s		
IBOAT 算法	2.1s		

结果证明, 本文算法在该数据集上的正确性和效率等方面相对原有算法有一定程度的提升.

#### 3.2 相关工作比较

本文重点研究的是用于防止出租车欺诈的实时异常轨迹检测, 轨迹如何表示十分关键. 针对这个切入点, 文献[7,13]提出并使用了地标图方法, 使用热门地标表示轨迹序列, 进行异常检测. 文献[8]首先提出了 IBAT, 用网格编号表示轨迹进行异常检测, 随后在文献[9-10,14-15]加入在线检测部分, 提出了优化的 IBOAT. 在这些工作基础上, 借鉴他们的优势, 并提出了全新的算法, 提升了效率和准确率. 在轨迹数据预处理方面, 常用的两种地图划分模型为矢量<sup>[13]</sup>与栅格<sup>[12,14]</sup>. 本文采取文献[12,14]的栅格模型, 因为该方法更便于将 GPS 点轨迹转换为 Pathlet 序列. 轨迹与地图匹配方面, 本文主要借鉴了已有工作<sup>[12,16]</sup>, 该方法更适合我们的目标.

Pathlet 方法表示轨迹数据融合了文献[16]的思想, 其介绍使用有约束的整数线性规划方法来取得 Pathlet 集合. 本文还借鉴了文献[11-12]中用于分割轨迹的一些方法来达成目标.

轨迹聚类的方法众多, 文献[18]使用了经典的  $K$ -means<sup>[19]</sup> 讨论了 LP 距离和谱聚类, 文献[5]使用了 LRT (likelihood-test), 文献[12]使用了 Top- $k$  与 kNN. 针对本文轨迹 Pathlet 序列的特点, 我们选择了  $K$ -Means 算法做聚类分析.

对于不同的目标和数据形式, 需要有针对性的选择轨迹相似度的度量. 文献[6-8,15,17]使用最近点对 (closest-pair) 距离, sum-of-pairs 距离, 动态时间校正 (dynamic time warping, DTW)、最长公共序列、实序列编辑距离 (edit distance on real

sequence, EDR) 和实补偿编辑距离 (edit distance with real penalty, ERP) 等不同方式, 体现了每种度量的优缺点. 本文根据自身特点选择了编辑距离作为度量标准, 并加入时间维度的考量.

## 4 结论

本文以防止出租车欺诈绕路为例, 提出一种基于出租车 GPS 轨迹数据离线挖掘与在线实时监测相结合的异常轨迹检测算法. 算法使用 Pathlet 方法优化常用的以 GPS 点组成的轨迹序列, 从大量出租车历史数据中, 获得轨迹的 Pathlet 序列, 并聚类得到起点与终点之间正常的  $K$  类轨迹. 当实时新轨迹需要被检测时, 便与  $K$  类正常轨迹进行匹配, 同时考量时间和空间两个维度设定合理阈值, 判断是否抛出异常. 基于北京市 2011 年 3 月到 5 月出租车 GPS 轨迹的真实数据集进行了大量实验, 对比了相关工作, 印证了所提出算法的有效性和高效性.

### 参考文献 (References)

- [1] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection a survey [J]. *ACM Computing Surveys*, 2009, 41(3): 75-79.
- [2] VELOSO M, PHITHAKKITNUKON S, BENTO C. Sensing urban mobility with taxi flow [C]// *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*. Chicago, USA: ACM Press, 2011: 41-44.
- [3] ZHU B, HUANG Q, GUIBAS L, et al. Urban population migration pattern mining based on taxi trajectories [C]// *Proceedings of the 3rd International Workshop on Mobile Sensing*. Springer, 2015, 9142: 172-181.
- [4] ZHENG Y, LIU Y C, YUAN J, et al. Urban computing with taxicabs [C]// *Proceedings of the 13th International Conference on Ubiquitous Computing*. Beijing, China: ACM Press, 2011: 89-98.
- [5] YUAN J, ZHENG Y, ZHANG C Y, et al. T-drive: Driving directions based on taxi trajectories [C]// *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. San Jose, USA: ACM Press, 2010: 99-108.
- [6] LEE J G, HAN J W, WHANG K Y. Trajectory clustering a partition-and-group framework [C]// *Proceedings of the SIGMOD International Conference on Management of Data*. Beijing, China: ACM Press, 2007: 593-604.
- [7] GE Y, XIONG H, LIU C R, et al. A taxi driving fraud detection system [C]// *Proceedings of the 11th International Conference on Data Mining*. Vancouver, Canada: IEEE Press, 2011: 181-190.
- [8] FU Z Y, HU W M, TAN T N. Similarity based vehicle trajectory clustering and anomaly detection [C]// *IEEE International Conference on Image Processing*. Beijing, China: IEEE Press, 2005, 2: 602-605.
- [9] ZHANG D, LI N, ZHOU Z H, et al. iBAT: Detecting anomalous taxi trajectories from GPS traces [C]// *Proceedings of the 13th International Conference on Ubiquitous Computing*. Beijing, China: ACM Press, 2011: 99-108.
- [10] CHEN C, ZHANG D Q, CASTRO P S, et al. Real-time detection of anomalous taxi trajectories from GPS traces [A]// *Mobile and Ubiquitous Systems Computing, Networking, and Services*. Berlin Heidelberg: Springer, 2012: 63-74.
- [11] GE Y, XIONG H, ZHOU Z H, et al. Top-eye Top-k evolving trajectory outlier detection [C]// *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*. Toronto, Canada: ACM Press, 2010: 1733-1736.
- [12] 胡佳峰. 基于轨迹数据的事件检测技术研究 [D]. 硕士学位论文, 中国科学院大学, 2014.
- [13] 袁晶. 大规模轨迹数据的检索、挖掘及应用 [D]. 博士学位论文, 中国科学技术大学, 2012.
- [14] CHEN C, ZHANG D Q, CASTRO P S, et al. iBOAT: Isolation-based online anomalous trajectory detection [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2013, 14(2): 806-818.
- [15] SUN L, ZHANG D, CHEN C, et al. Real time anomalous trajectory detection and analysis [J]. *Mobile Networks and Applications*, 2013, 18(3) 341-356.
- [16] CHEN C, SU H, HUANG Q, et al. Pathlet learning for compressing and planning trajectories [C]// *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Orlando, USA: ACM, 2013: 392-395.
- [17] LEE A J T, CHEN Y A, IP W C. Mining frequent trajectory patterns in spatial-temporal databases [J]. *Information Sciences*, 2009, 179(13): 2218-2231.
- [18] CHEN L, NG R. On the marriage of Lp-norms and edit distance [C]// *Proceedings of the 30th International Conference on Very Large Databases*. Toronto, Canada: VLDB Endowment, 2004: 792-803.
- [19] HAN B, LIU L, OMIECINSKI E. Neat Road network aware trajectory clustering [C]// *32nd International Conference on Distributed Computing Systems*. Macau, China: IEEE, 2012: 142-151.