

· 计算机工程与应用 ·

## 基于路径旅行时间分析的交通异常检测方法

王璞, 熊雨沙, 王骋程, 郑治豪, 鲁恒宇

(中南大学交通运输工程学院 长沙 410075)

**【摘要】**为了综合考虑连续路段通行能力波动对旅行时间的影响,避免由单一路段通行能力的常规性波动所导致的交通异常误判,提出了一种基于路径旅行时间分析的交通异常检测算法。该算法将深圳市路网网格化为若干个地理子区,以地理子区为单位,使用ST-matching地图匹配算法将深圳市出租车GPS坐标记录点匹配到相应路段,采用基于密度的DBSCAN聚类算法计算路径旅行时间的时变异常阈值,来判定旅行时间的异常。该方法成本低廉,实施难度小,能精确灵敏地检测交通网络异常。

**关键词** DBSCAN聚类算法; GPS; 地图匹配; 交通异常检测; 路径旅行时间  
**中图分类号** U91 **文献标志码** A **doi:**10.3969/j.issn.1001-0548.2018.06.011

## Traffic Anomaly Detection Method Based on Travel Time of Path

WANG Pu, XIONG Yu-sha, WANG Cheng-cheng, ZHENG Zhi-hao, and LU Heng-yu

(School of Transportation Engineering, Central South University Changsha 410075)

**Abstract** To take the capacity variations of road segments into consideration, and to avoid the fake traffic anomaly detections caused by the normal fluctuation of road segment capacity, a new traffic anomaly detection method based on travel time of path is proposed. Shenzhen road network as sub spatial regions is divided into grids, and then ST-matching algorithm is applied to match Shenzhen taxi GPS coordinate to corresponding road segments, after that the DBSCAN algorithm is adopted to compute the time-varying threshold of path travel time, the travel time above threshold is recognized as anomaly. While maintaining high accuracy and sensitivity of detecting traffic network anomaly, this method is cost efficient and easy to implement.

**Key words** DBSCAN algorithm; GPS; map-matching; traffic anomaly detection; travel time of path

交通异常状态检测是当前交通领域中一个非常重要的研究方向。迄今为止,国内外已有的相关研究多数利用了线圈<sup>[1-2]</sup>、视频<sup>[3-5]</sup>、红外传感装置<sup>[6]</sup>等固定传感设备采集数据。然而,由固定传感设备采集的数据多限于检测器断面<sup>[7]</sup>,使交通异常检测受到了很大局限。随着全球定位系统(GPS)的日益普及,许多城市已可以实现对移动对象的实时跟踪,用装有GPS接收器的车辆作为移动传感器来采集交通状态信息,具有成本低、灵活性强、数据量大等优点<sup>[8]</sup>。近年来,基于车载GPS数据的研究主要集中在车辆路径选择模式及路段异常判别两个方面。

在车辆路径的选择模式方面,文献[9]将路径分割为若干子路径,通过比较子路径组成的相关多维向量,找出其中的异常子路径,从而发现城市的异常路径选择行为;文献[10]挖掘相同OD之间的不同

路径,提出了用流量构成描述向量的方法,挖掘人类出行异常模式子图,并利用社交媒体数据进行异常信息的补充;文献[11]提出了基于车辆移动轨迹在线检测异常轨迹的算法。这类方法研究重点在于发现人类的出行规律。

在路段异常判别方面,文献[12]通过比较观测路段与其他路段的相似性,构造包含历史趋势信息的状态向量,提取出离群点作为异常值;文献[13]通过比较路段真实值流量和期望值流量,判断路段是否为异常路段。基于路段的异常判断能直观清晰地反映路段通行情况,但缺乏对路段之间内在联系和相互作用的融合,难以确定异常产生的根源并进行疏导;文献[14]针对车辆运行状态如速度、加速度等的变化,推测当前行驶路段是否异常,但这类方法需要考虑不同等级路段固有属性对车辆通行状态的限

收稿日期: 2017-07-21; 修回日期: 2017-11-01

基金项目: “十三五”国家重点研发计划(2016YFB1200400); 国家自然科学基金面上项目(61473320)

作者简介: 王璞(1983-),男,博士,教授,主要从事交通运输规划与管理、统计物理学、复杂网络和数据挖掘方面的研究。

制,应用难度较大。

为了解决现有交通异常检测方法遇到的难点,本文提出一种基于路径旅行时间分析的交通网络异常检测算法。该算法以Map-matching技术和DBSCAN聚类算法为基础,且与传统方法相比具有以下特点:

1) 以路径为研究对象。

与基于路段的交通异常检测方法相比,基于路径的交通异常检测具有以下优点:首先,该方法可以获得更多样本量,减小偶发情况对整体检测结果的影响。其次,该方法能综合考虑连续路段交通状况的累加影响,对不同程度交通状况的检测结果更为精确。最后,异常路径的发现能直观显示交通异常状况影响区域,多个异常路径的累加更容易形成交通异常子网,便于发现交通异常的影响程度及传播规律。

2) 以旅行时间为判断标准。

本文利用浮动车GPS数据得到的路径旅行时间数据能够避免路段固有属性比如车道数、路面状况等对上述指标的影响,无需相应的路段参数即可对交通状况进行判断,方法简单,通用性强。

## 1 路网数据与出租车GPS数据

### 1.1 路网数据

使用由图1所示,26 242个节点和43 134条边组成的高精度深圳路网数据。为了提高算法效率,将城市分割为 $83 \times 43$ 个 $1000 \text{ m} \times 1000 \text{ m}$ 的地理子区,以地理子区为单位进行交通异常检测。

### 1.2 出租车GPS数据

采用由13 584辆配备GPS接收器的出租车采集的GPS数据,平均数据采集频率为15 s/次,使用2014年10月所有工作日共916 851 503个GPS坐标记录。由于出租车空载寻客时一般车速较慢,路段速度不具有代表性,因此删除出租车在空载状态下的GPS数据。由于少量GPS采集设备故障,还需要对GPS数据进行清洗,删除出租车日出行次数(由载客状态转变为空载状态为一次出行)大于500次的GPS设备所产生数据。为保证单位时间内充足的GPS记录数量,只使用6:00-23:00时间段的数据。经过以上数据预处理,平均每天有13 446 868条出租车坐标点记录被保留下来。

## 2 出租车行驶路径的获取

在一个地理子区中,满足以下条件的GPS轨迹

点组成轨迹 $T = p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$ 。

1) GPS轨迹点有可能匹配到与其临近的任意一条路段上,所以需要采用缓冲半径公式,以GPS轨迹点为中心确定搜索范围,选择范围内的路段作为候选匹配路段,有:

$$\rho = \alpha + \frac{w}{2} + \beta - \frac{m}{2} \quad (1)$$

式中,参数 $\alpha$ 是道路网定位误差;参数 $w$ 是单向道路的宽度;参数 $\beta$ 是车辆定位的最大误差;参数 $m$ 是车辆宽度<sup>[15]</sup>。经调研,深圳市最宽道路为深南大道,路面净宽80 m,除去16 m绿化带宽度,单向道路宽32 m,其他参数的取值参考文献[15]: $\alpha = 5$ , $\beta = 10$ , $m = 2$ ,得到 $\rho = 30$ 。本文将搜索范围略微扩大,取值35 m,并删除35 m范围内没有路段的GPS轨迹点。

2)  $p_i$  对应的时间标签值递增。

3)  $p_i$  和  $p_{i+1}$  之间的时间间隔小于45 s。

如图1所示,虚线框表示一个地理子区,图中 $p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$  为地理子区中一条满足条件的轨迹。

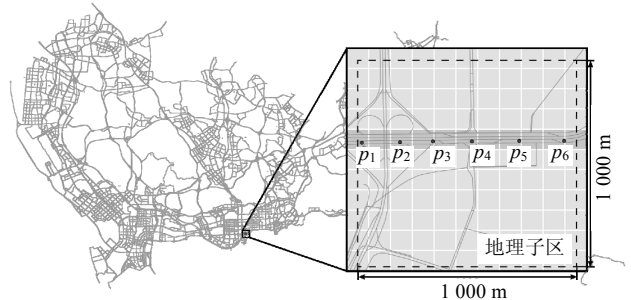


图1 深圳路网和地理子区GPS轨迹

为了计算出出租车的行驶路径和路段的通行速度,使用ST-matching地图匹配算法<sup>[16]</sup>将GPS轨迹匹配到路段上。该算法以路径为匹配对象,结合了路网拓扑结构和地理信息,是一种全局的地图匹配算法,对低频的GPS数据也能保持较高的匹配精度,并能方便地获取路径和路径的旅行时间。算法步骤如下:

1) 获得候选路径。

对于一条GPS轨迹 $T = p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$ ,当 $p_i$ 与路段 $e_i^j$ 的距离小于 $d = 35 \text{ m}$ 时,将 $e_i^j$ 加入 $p_i$ 的候选匹配边集合中, $p_i$ 在 $e_i^j$ 上的投影 $c_i^j$ 则为 $p_i$ 的第 $j$ 个候选匹配点,将 $n$ 个轨迹点相邻两点的候选匹配点两两组合,可得到该轨迹的若干条候选匹配路径。

2) 空间分析。

ST-matching算法中,路网的地理信息和拓扑结构都被用来评价候选匹配路径与GPS轨迹的匹配程度,地理信息评价用观测可能性来衡量,拓扑结

构评价用传递可能性来衡量。

观测可能性用  $p_i$  与  $c_i^j$  的距离来衡量  $p_i$  匹配到  $c_i^j$  的概率。定义  $p_i$  匹配到  $c_i^j$  的观测可能性  $f(c_i^j)$  为:

$$f(c_i^j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i^j - \mu)^2}{2\sigma^2}} \quad \mu = 0, \sigma = 20 \quad (2)$$

式中,  $x_i^j$  是  $p_i$  与  $c_i^j$  的距离,  $\text{dist}(c_i^j, p_i)$  越小,  $p_i$  点匹配到  $c_i^j$  的可能性越大。传递可能性  $V(c_{i-1}^t \rightarrow c_i^s)$  基于候选匹配路径上相邻两点  $c_{i-1}^t$  和  $c_i^s$  的最短路径长度计算得到, 计算公式为:

$$V(c_{i-1}^t \rightarrow c_i^s) = \frac{d_{i-1 \rightarrow i}}{w_{(i-1,t) \rightarrow (i,s)}} \quad (3)$$

式中,  $d_{i-1 \rightarrow i} = \text{dist}(p_{i-1}, p_i)$  为  $p_{i-1}$  与  $p_i$  之间的欧式距离;  $w_{(i-1,t) \rightarrow (i,s)}$  为  $c_{i-1}^t$  和  $c_i^s$  的最短路径长度。

结合观测可能性和传递可能性,  $c_{i-1}^t \rightarrow c_i^s$  的空间分析函数定义为:

$$F(c_{i-1}^t \rightarrow c_i^s) = V(c_{i-1}^t \rightarrow c_i^s) f(c_i^s) \quad 2 \leq i \leq n \quad (4)$$

### 3) 路径匹配。

轨迹  $T$  的候选匹配路径组成候选匹配图  $G_T'(V_T', E_T')$ ,  $V_T'$  是GPS轨迹点的候选点集合,  $E_T'$  是相邻两点之间最短路径的集合, 一条候选匹配路径  $P_c$  可表示为  $P_c: c_1^{s_1} \rightarrow c_2^{s_2} \rightarrow \dots \rightarrow c_n^{s_n}$ , 那么匹配到路径  $P_c$  的可能性为:

$$F(P_c) = \sum_{i=2}^n F(c_{i-1}^{s_{i-1}} \rightarrow c_i^{s_i}) \quad (5)$$

匹配可能性最高的路径为  $T$  的最终匹配路径为:

$$P = \arg \max F(P_c) \quad \forall P_c \in G_T'(V_T', E_T') \quad (6)$$

图2表示一条待匹配轨迹  $T = p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$ , 三角形标识的点都是待选的候选匹配点,  $p_5$  和  $p_6$  各有两个候选匹配点, 实线和虚线分别为两条候选匹配路径, 经计算, 经过  $c_5^1$  和  $c_6^1$  的路径匹配度更高。该路径的旅行时间为起点与终点, 即  $p_1$  与  $p_6$  的时间标签的差值,  $p_{i-1}$  到  $p_i$  的行驶速度为:

$$v_{(c_{i-1}^t \rightarrow c_i^s)} = \frac{w_{(i-1,t) \rightarrow (i,s)}}{t_{(i-1,t) \rightarrow (i,s)}} \quad (7)$$

式中,  $w_{(i-1,t) \rightarrow (i,s)}$  是  $p_{i-1}$  到  $p_i$  的最终匹配点之间的最短路径;  $t_{(i-1,t) \rightarrow (i,s)}$  是它们之间的时间差。如果同时同一段一条路段上有多个速度记录, 可计算得到路段的平均通行速度。

由于所使用的路网精度较高, 可能存在因起点或终点不在同一条路段的细微差别, 而将两条轨迹判定为经过不同路径的情况。为了保证路径判定有一定的容差率, 将地理子区进一步划分为  $100 \text{ m} \times$

$100 \text{ m}$  的格子, 如果两条轨迹途经的细分网格相同, 则认为它们经过同一条路径。

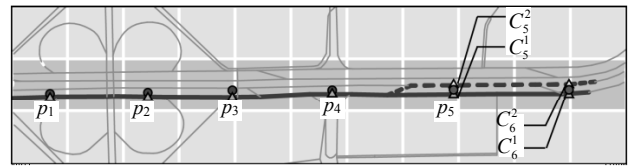


图2 地图匹配算法示例

## 3 路径的交通异常判断算法

本文采用基于密度的DBSCAN聚类算法<sup>[17]</sup>对通过同一条路径所有轨迹的旅行时间进行聚类。该算法需要确定两个参数: EPS邻域和最小样本数MinPts, 参数的选择对聚类的结果影响较大, 较小的EPS和较大的MinPts会导致算法对路径旅行时间异常的判断不够敏感, 反之, 会过于敏感。使用欧式距离计算两点之间的距离为:

$$\text{dist}(t_1, t_2) = |\Delta t_1 - \Delta t_2| \quad (8)$$

30 min是目前交通状态短时预测常用时间段之一<sup>[18]</sup>, 且能保证一条路径有足够数据量来进行聚类分析, 所以, 使用30 min作为时间窗, 定义每个时间窗的数据为一个聚类数据集, 并将聚类数据集中的每个实例除以数据集中实例的最大值, 实现所有实例的归一化。根据文献<sup>[17]</sup>对参数EPS值和MinPts值选取方法的介绍, 先确定MinPts取值为4, 然后找到4-距离的突增位置来确定参数EPS的值, 方法如下: 首先计算每两个实例之间的欧式距离, 找到与每个实例最邻近的第4个实例, 二者之间的距离称为4-距离。然后对一个聚类数据集中实例的4-距离进行概率统计, 发现绝大多数数据集的4-距离概率呈指数分布。最后对概率分布进行指数拟合, 选取斜率为-1处对应的4-距离作为EPS值。图3a是对某个聚类数据集的4-距离进行归一化后的概率分布图, 图3b是将图3a中局部区域放大后, 做拟合曲线斜率为-1处的切线, 对应的横坐标为EPS取值。

经过聚类后, 聚类数据集中的实例被分成  $m$  个团簇, 定义拥有最多实例的团簇为中心团簇, 中心团簇中最大值为初始异常阈值, 初步认为数据集中所有小于初始异常阈值的对象是初始正常值, 初始正常值的平均值  $\langle t \rangle$  为路径在所测时间段的正常旅行时间。由于存在少量聚类数据集, 使实例的4-距离不满足指数分布, 且拟合的效果不稳定, 为减少DBSCAN参数对阈值计算结果的影响, 根据常用异常检测方法三倍标准差准则<sup>[19]</sup>, 计算初始正常值的

标准差  $\sigma$ 。定义  $\langle t \rangle + 3 \times \sigma$  为路径在该时间段内的异常阈值。如图4所示，图4a为某条路径10月份工作日10:20-12:30所有旅行时间，虚线为每30 min路径旅行时间的异常阈值。每10 min计算一次待测路径的平均旅行时间，图4b为某条待测路径在10:20-12:30不同时间窗进入地理子区时，路径的平均旅行时间，与对应时间窗的异常阈值对比，虚线空心点所表示的旅行时间值超出了异常阈值，由图可知，异常持续了50 min，即11:00-11:50。

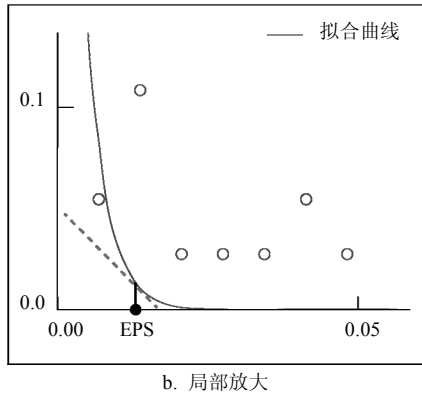
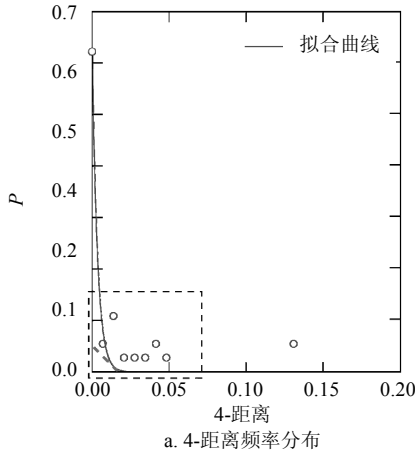


图3 DBSCAN算法参数选择

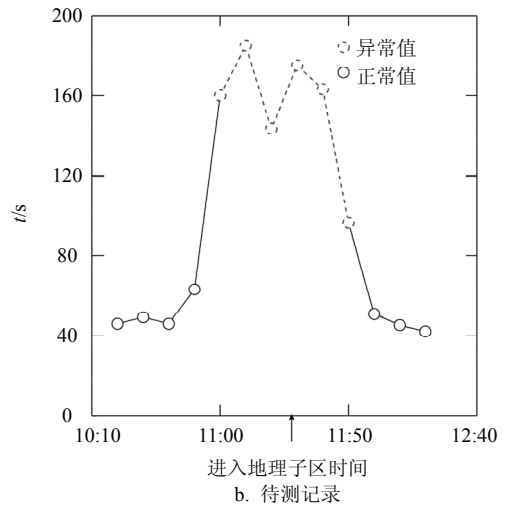
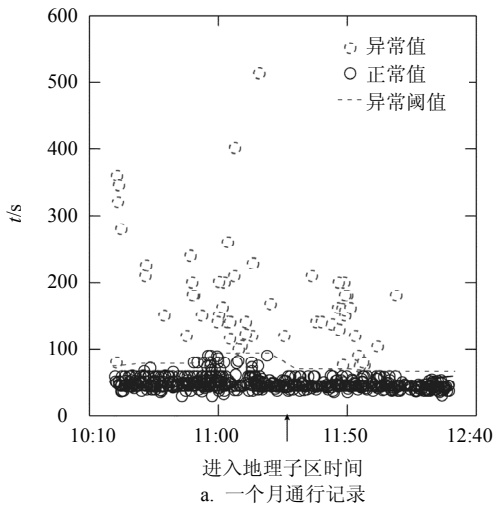


图4 路径旅行时间异常判断

### 4 地理子区交通异常检测案例

交通事故或交通拥堵等交通异常事件的显著特点是造成路段通行速度降低，因此，大量的研究基于路段的速度、流量等属性进行路段状态识别。然而，这种异常识别方法也存在一些不足。一些路段的速度表现出不规则的波动给交通异常判断带来困难，如图5a和图5b所示。因此，选择路径的旅行时间作为研究对象，能综合考虑多个邻接路段的通行情况，减小异常检测误差，如图4b所示。

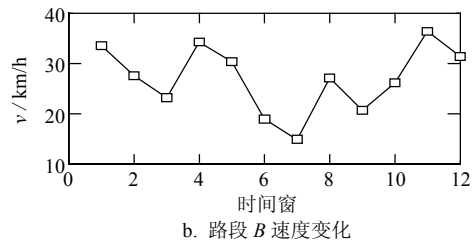
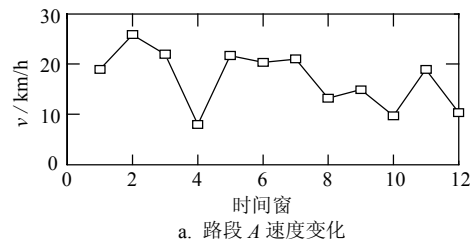


图5 路段速度计算

使用上一节阐述的方法计算各路径的异常阈值，图6为某地理子区中检测出的4条异常路径旅行时间随时间的变化，4条路径的旅行时间均在第6个时间窗开始突增并超出异常阈值，于第8个时间窗附近达到相对高峰之后逐步恢复正常。由图6a可知，该条路径正常通行时的旅行时间约50 s，而旅行时间异常值最高达280 s。

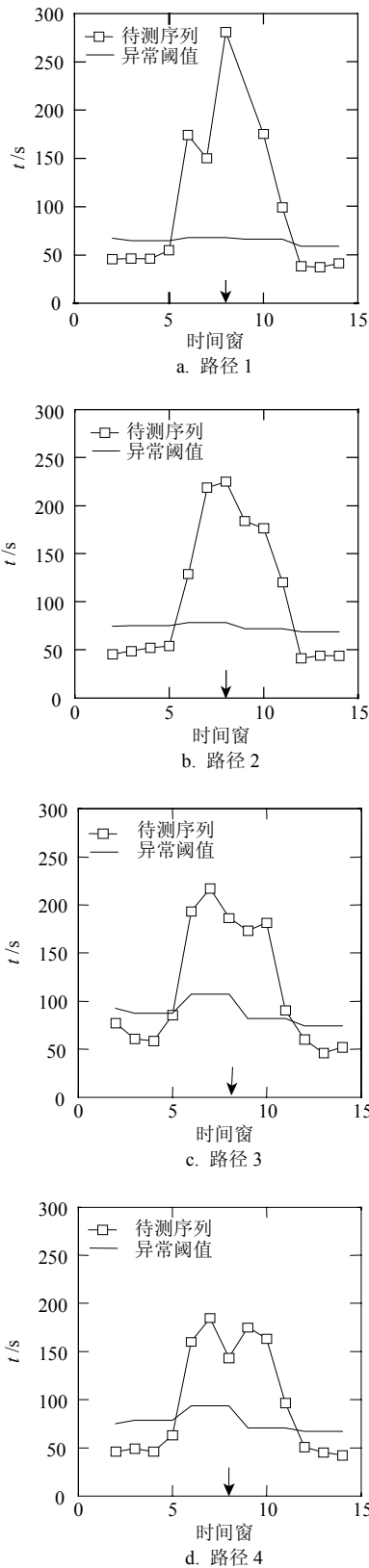


图6 路径旅行时间随时间变化

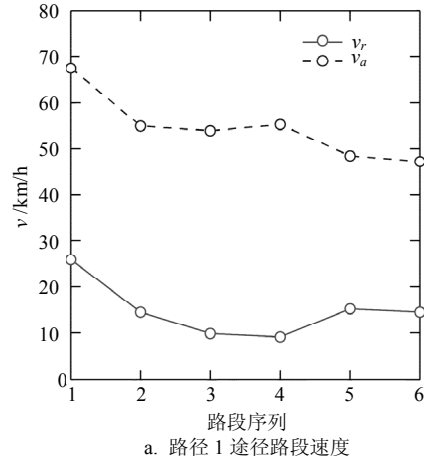
为衡量异常路径覆盖路段受影响程度, 引入 $R$ 值进行计算: 计算路径在各个时间窗的受影响程度=平均旅行时间/ $\langle t \rangle$ , 取受影响程度最大的时间窗进行

各路段  $R$  值的计算:

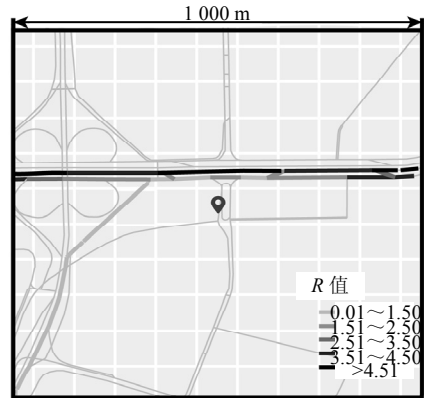
$$R = v_r / v_a \tag{9}$$

式中,  $v_r$  为一个月中该时间窗的平均通行速度;  $v_a$  为受影响程度最大时间窗的平均通行速度。图7a所示为路径1在第8个时间窗依次经过路段  $v_r$  和  $v_a$  的对比。

图7b中加粗路段为异常路径覆盖路段, 由  $R$  值分布可看出各路段受影响的程度。



a. 路径 1 途径路段速度



b. 路段  $R$  值分布

图7 路段速度分析

为验证该交通异常检测结果, 从新浪微博中搜索信息, 发现如表 1 所示事件。

表1 微博实例

微博发出时间	初始坐标	内容
October 17 11:28	114.077 22 528	11 时 25 分, 滨河大道福田南路西往东方向, 发生两小车事故, 占据中间车道, 导致车多缓行, 交警正赶往现场, 请途经车友注意避让

该事件初始定位在图7b所示的地理图标位置, 事件描述与所测交通异常基本相符, 微博所反映的事件发生在11:28, 为图6中箭头所示位置。而算法在11:00就检测出交通异常, 比微博发出提前28分钟, 微博发出时间基本位于路径旅行时间最大的时间窗内, 从微博内容可看出此时交警还在赶往途中。

图6b展示了交通异常的产生和消散过程,图7b展示了交通异常的集中区域。

将地理子区内所有交通异常路径经过的节点和边组成的网络定义为“基于路径的交通异常子网”,如图8a中的粗线所示。建立基于路段的交通异常子网用于比较。定义路段在检测时间窗的车辆行驶平均速度 $\mu_c$ ,路段在工作日相同时段的车辆行驶平均速度为 $\mu_m$ ,标准差为 $\sigma_e$ ,当 $\mu_c < \mu_m - 3 \times \sigma_e$ 时认为路段的车辆行驶速度发生异常。提取车辆行驶速度发生异常的边组成的网络,得到“基于路段的交通异常子网”,如图8b、图8d所示。

在发生交通事故的时间窗:基于路径的交通子网连成一个大团簇,如图8a所示;而基于路段的交通异常子网分布较为杂乱,有几个零散的团簇,如图8b所示。在交通运行正常的时间窗:基于路径的交通子网根本不存在,如图8c所示;而基于路段的交通异常子网还是有一些零散的小团簇存在,这是由于个别路段偶发速度波动引起的,如图8d所示。可以看出,基于路径的交通异常子网在交通正常与异常状态下的区别更为明显,有助于提高交通异常判断的准确性,上述结果也体现了基于路径的交通异常检测方法的优点:有利于减小偶发情况对整体检测结果的影响。

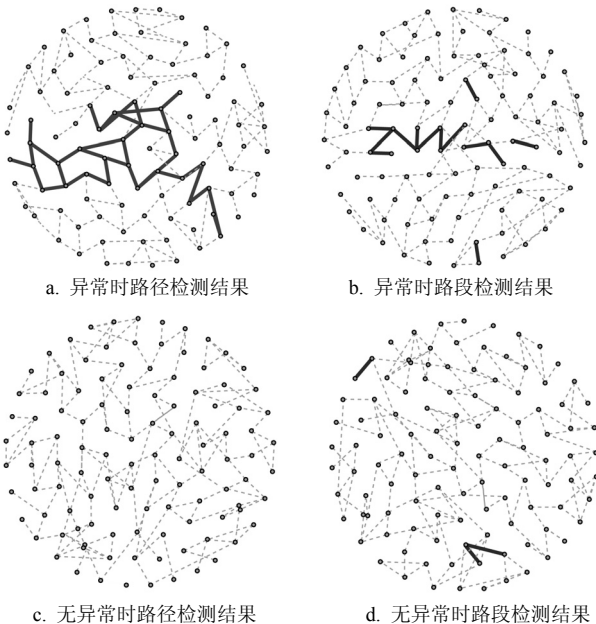


图8 异常网络规模

## 5 结束语

近年来,随着数据分析挖掘技术的发展,路段旅行时间和拥堵的判别方法已经日益成熟,出行者在出行时可以借助各种软件,查看到各个路段的实

时路况信息,比如拥堵情况、运行速度、限速情况等。但是,局部的、可以自行消散的路段异常出现频繁,使交管部门较难准确识别出真正存在严重问题的区域。

本文针对有一定规模的道路交通异常突发情况,提出了基于路径旅行时间的交通异常检测算法。该算法不同于常用的基于路段通行能力和状态进行异常检测的方法,可以减少由单一的路段信息和车辆偶然因素导致的异常误判,检测出具有规模的、自行消散困难的、异于历史通行状态的交通异常网络,为交通管理部门提供可信用度更高的交通异常监测结果。算法在现实路网和真实数据环境下的应用证明其检测的交通异常时空信息及时有效,所检测出的带有时间和空间信息的交通异常,能为交通管理者疏导交通提供重要管控和决策信息。

本文研究得到霍英东青年教师基金基础研究课题(141075)及中南大学创新驱动计划(2016CSX014)的资助,在此表示感谢。

## 参 考 文 献

- [1] ZHANG Zhen-hua, HE Qing, TONG Hang-hang, et al. Spatial-temporal traffic flow pattern identification and anomaly detection with dictionary-based compression theory in a large-scale urban network[J]. *Transportation Research Part C Emerging Technologies*, 2016, 71: 284-302.
- [2] 姜桂艳, 闫赫, 李琦, 等. 基于悉尼协调自适应交通控制系统线圈数据的交通事件自动检测算法[J]. *吉林大学学报(工学版)*, 2012, 42(3): 575-580.  
JIANG Gui-yan, YAN He, LI Qi, et al. Automated incident detection algorithm based on inductive loop data on Sydney coordinated adaptive traffic system[J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2012, 42(3): 575-580.
- [3] DANG T T, NGAN H Y T, LIU Wei. Distance-based k-nearest neighbors outlier detection method in large-scale traffic data[C]//2015 IEEE International Conference on Digital Signal Processing. Singapore: IEEE, 2015: 507-510.
- [4] 方晓莹, 王小君. 基于视频的智能交通异常检测系统[J]. *微型机与应用*, 2010, 29(6): 35-38.  
FANG Xiao-ying, WANG Xiao-jun. A traffic abnormality detection algorithm based on vision in the ITS[J]. *Microcomputer Its Applications*, 2010, 29(6): 35-38.
- [5] 万福才, 尹承祥, 韩晓微, 等. 城市道路交通异常事件自动检测方法[J]. *沈阳大学学报(自然科学版)*, 2015, 27(1): 30-33.  
WAN Fu-cai, YIN Cheng-xiang, HAN Xiao-wei, et al. Automatic detection method of abnormal events in city road traffic[J]. *Journal of Shenyang University (Natural Science)*, 2015, 27(1): 30-33.
- [6] 姚殿梅, 周彬. 红外线在道路测试中的应用[J]. *交通科技*

- 与经济, 2013, 15(3): 45-48.
- YAO Dian-mei, ZHOU Bin. Application of infrared ray to road tests[J]. Technology & Economy in Areas of Communications, 2013, 15(3): 45-48.
- [7] 王尧, 邵长桥, 刘洋. 城市快速路交通拥堵判定方法研究[J]. 交通信息与安全, 2014, 32(2): 23-27.
- WANG Yao, SHAO Chang-qiao, LIU Yang. Traffic congestion identification method for urban expressway[J]. Journal of Transport Information and Safety, 2014, 32(2): 23-27.
- [8] 余柳, 于雷, 戚懿, 等. 基于浮动车数据的城市快速路交通事件检测算法研究[J]. 交通运输系统工程与信息, 2008, 8(4): 36-41.
- YU Liu, YU Lei, QI Yi, et al. Traffic incident detection algorithm for urban expressways based on probe vehicle data[J]. Journal of Transportation Systems Engineering and Information Technology, 2008, 8(4): 36-41.
- [9] LEE J G, HAN Jia-wei, LI Xiao-lei. Trajectory outlier detection: a partition-and-detect framework[C]//2008 IEEE 24th International Conference on Data Engineering. Cancun: IEEE, 2008: 140-149.
- [10] PAN Bei, ZHENG Yu, WILKIE D, et al. Crowd sensing of traffic anomalies based on human mobility and social media[C]//ACM Sigspatial International Conference on Advances in Geographic Information Systems. Orlando: ACM, 2013: 344-353.
- [11] LAXHAMMAR R, FALKMAN G. Online learning and sequential anomaly detection in trajectories[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(6): 1158-1173.
- [12] LI Xiao-lei, LI Zhen-hui, HAN Jia-wei, et al. Temporal outlier detection in vehicle traffic data[C]//2009 IEEE 25th International Conference on Data Engineering. Shanghai: IEEE, 2009: 1319-1322.
- [13] HUANG Chao, WU Xian. Discovering road segment-based outliers in urban traffic network[C]//IEEE GLOBECOM Workshops. Atlanta: IEEE, 2013: 1350-1354.
- [14] 冯星宇, 周晨静, 荣建. 基于速度特性的城市快速路常发性交通拥堵研究[J]. 交通信息与安全, 2014, 32(1): 29-33.
- FENG Xing-yu, ZHOU Chen-jing, RONG Jian. Recurrent congestion analysis based on speed variation characteristics on urban expressway[J]. Journal of Transport Information and Safety, 2014, 32(1): 29-33.
- [15] 钱颖. 基于GIS的出租车辆GPS地图匹配研究[D]. 昆明: 云南大学, 2015.
- QIAN Ying. Research on map-matching of GPS based on GIS[D]. Kunming: Yunnan University, 2015.
- [16] LOU Yin, ZHANG Cheng-yang, ZHENG Yu, et al. Map-matching for low-sampling-rate GPS trajectories[C]//ACM Sigspatial International Symposium on Advances in Geographic Information System. Seattle: IEEE, 2009: 352-361.
- [17] ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//International Conference Knowledge Discovery and Data Mining. Portland: KDD, 1996: 226-231.
- [18] 郭敏, 肖翔, 蓝金辉, 等. 道路交通流短时预测方法综述[J]. 自动化技术与应用, 2009, 28(6): 8-9.
- GUO Min, XIAO Xiang, LAN Jin-hui, et al. A summary of the short-time traffic flow forecasting methods[J]. Techniques of Automation & Applications, 2009, 28(6): 8-9.
- [19] VYSOCHANSKIJ D F, PETUNIN Y I. Justification of the  $3\sigma$  rule for unimodal distributions[J]. Theory of Probability & Mathematical Statistics, 1980, 21(21): 25-36.

编辑 蒋晓