

基于出租车 GPS 数据的道路平均速度估计方法

丁涛杰, 史殿习, 李永谋

(国防科学技术大学 计算机学院, 湖南 长沙 410073)

摘要: 浮动车技术是目前常用的获取道路交通信息技术手段之一,而出租车 GPS 数据是该技术重要数据来源,但其本身存在诸多不足。为更好地利用出租车 GPS 数据,文中提出了在定位误差、低频采样、数据量少等因素影响下的行程速度估计方法。首先结合城市复杂路段情况,提出基于车流方向、路段可达性、速度等约束条件下的实时地图匹配算法,引入反向检查机制对不确定点进行修正;然后针对出租车运营特性,建立了有效数据获取模型,并提出低频采样影响下的行程速度估计算法;最后利用北京出租车数据进行计算比较。结果表明了该方法的可用性,能满足道路状态判断的一般需求。

关键词: 出租车 GPS 数据;平均行程速度;有效数据获取;地图匹配

中图分类号: TP301

文献标识码: A

文章编号: 1673-629X(2015)07-0015-05

doi: 10.3969/j.issn.1673-629X.2015.07.004

Estimation Method for Average Traffic Speed Based on GPS Data of Taxi

DING Tao-jie, SHI Dian-xi, LI Yong-mou

(College of Computer, National University of Defense Technology, Changsha 410073, China)

Abstract: Floating car data technology is one of the techniques currently used to get traffic information, and GPS data of Taxi is an important source of it, but with many deficiencies. In order to availably utilize GPS data of Taxi, a travel speed estimation method under the influence of all kinds of external factors is presented in the paper. First, a map matching algorithm based on direction of traffic, road accessibility, speed and reverse check mechanism is presented to solve the problem of complex sections, introducing reverse check mechanism to modify the point of uncertainty. Second, the effective data acquisition model is established and the travel speed estimation algorithm under influence of low frequency sampling is presented. Finally, make a comparison using the Beijing Taxi data. The results verify the feasibility of this algorithm, which can well meet the requirements of road situation judgment.

Key words: GPS data of Taxi; average travel speed; effective data acquisition; map matching

0 引言

浮动车 (Floating Car Data) 技术是智能交通系统 (ITS) 领域近年来兴起的一种先进技术手段^[1-2]。其基本原理是对交通流中浮动车实时上传的 GPS 数据进行分析处理,使浮动车数据与城市道路数据在时间和空间上相关联,最终得到城市的交通状况信息^[1-3]。相比传统的交通信息采集技术,浮动车技术具有成本低、覆盖率高、实时性强等优点,得到了大量的推广应用。目前多数大城市都建立了 ITS 平台,并实现了基于出租车或公交车等浮动车的 GPS 数据采集,将其应用于交通状况估计等实时交通信息服务。

近几年,国内外学者对浮动车技术进行了丰富的

理论和实践研究,涌现出大量相关研究文献^[4-10]。然而,目前的研究主要针对理想的高频采样的 GPS 数据,对低频采样的 GPS 数据研究^[8]甚少。但在实际应用中,由于通讯负载、投资成本等一些客观因素的限制,目前我国城市交通中通过浮动车进行 GPS 数据的中低频采样更为普遍。

出租车 GPS 数据是浮动车技术的重要数据来源,除了低频采样造成的影响外,还存在空车寻客、上下载客、停运等特殊特性,很难将其直接应用到道路路况计算中去。为更好利用出租车 GPS 数据,文中改进了低频采样下的地图匹配算法,提出了一种基于出租车数据特性的有效数据获取方法和道路状态估计算法。

1 研究对象与方法

选取北京市 12 000 余辆出租车一个月的 GPS 数据(基本覆盖了北京的主要交通路段)作为研究数据,选取一条主干路段作为研究对象。原始浮动车数据有 9 个属性项,具体格式如表 1 所示。据统计,采样频率在 30 s 以下的约占 29%,30 s ~ 1 min 的约占 61%,1 ~ 2 min 的约占 6%,2min 以上的约占 4%。

表 1 浮动车数据原始数据格式

属性项	具体说明
车辆标识	6 个字符,用于标识唯一车辆
触发事件	0=变空车,1=变载客,2=设防,3=撤防,4=其他
运营状态	0=空车,1=载客,2=驻车,3=停运,4=其他
GPS 时间	格式 yyyyymmddhhnnss
GPS 经度	格式 ddd. dddddd
GPS 纬度	格式 dd. dddddd
GPS 速度	格式 ddd,取值 000 ~ 255 内整数
GPS 方位	格式 ddd,取值 000 ~ 360 内整数

文中按一定时间间隔(设定为 5 min)接收来自各运营公司的出租车 GPS 数据,先对原始数据进行实时地图匹配,再根据路段上的匹配位置和时间间隔信息对单辆浮动车行驶速度进行计算,最后根据一定规则估计该路段的平均行驶速度并进行误差分析。

2 实时地图匹配

目前针对 GPS 数据的地图匹配算法可大致分为三种:基于道路几何信息的匹配算法,基于概率统计的匹配算法,以及基于卡尔曼滤波、模糊理论和人工神经网络等现代理论的匹配算法。后两种算法在固定场合下具有很高的匹配精度,但复杂性相对较高,并且需要大量的数据进行统计分析或是对算法参数的前期学习和总结。因而对于实时性要求高、数据量较大、行驶道路环境复杂多变的出租车,这两种算法很难符合要求。文中提出了基于车流方向、路段可达性、速度等约束条件下的地图匹配算法,并在准确匹配的情况下引入反向检查机制对不确定匹配点进行修正。

文中定义准确匹配的启发式规则如下:

- (1) 行驶方向与路段方向夹角小于 5°;
- (2) 只有一个候选路段,或者 GPS 点到候选路段集的最短距离比其余距离相差 5 倍以上。

当两个条件都满足时,认为该 GPS 点准确匹配到了目标路段,进行反向检查,对该点与上一准确匹配点之间的不确定匹配点进行修正,并将其标为准确匹配。

在匹配过程中,文中假定 GPS 点误差基本在百米以内,算法描述如下:

算法:基于反向检查机制的实时地图匹配算法。

输入:GPS 原始点序列 $C = \{c_1, c_2, \dots, c_n\}$, 地图

数据;

输出:GPS 匹配点序列 $G = \{g_1, g_2, \dots, g_n\}$ 。

Begin

Step1:设置准确匹配标志 F 序列 $\{false, false, \dots, false\}$;

Step2:从 C 中依次获取 GPS 点,在矢量地图上搜索 100 m 范围内的路段信息,将其加入候选路段集 R ;

Step3:if 约束条件成立:从 R 中删去相应路段;

Step4:if $SIZE(R) = 0$:没有合适路段,设置该点 F 值为 false,重复 Step2 直至取完 C 中数据

else if $SIZE(R) = 1$:将剩余路段作为匹配路段

else if $SIZE(R) > 1$:比较剩余路段,优先选取车行方向与路段方向一致度较高、点与路段距离较近的作为匹配路段

Step5:根据启发式规则判断设置该点 F 值;

if $F = 1$:在 F 序列中搜索上一值为 true 的点,获取两点间的所有未准确匹配点,根据序列连通性对其进行修正,并将 F 值标为 true;

Step6:重复 Step2 直至取完 C 中数据,查看 F 序列是否全为 true,对 false 的点数据根据连通性进行修正,得到最终匹配点序列 $\{g_1, g_2, \dots, g_n\}$ 。

End

在上述算法中,文中定义以下几条约束条件:

约束 1:在候选路段集中,删去当前 GPS 点与上一 GPS 点之间的实际路段长度与时间之商大于限定值的候选路段。

约束 2:当车行方向与路段方向的差值大于 90° 时,删除该候选路段;并且差值越小,匹配可能性越高。(这里双向车道视为两个不同方向的路段)

约束 3:若候选路段对上一匹配路段不可达,删去该路段。

约束 4:若候选路段不存在 GPS 点的垂点,则认为匹配到该路段的可能性很小;当候选路段存在多个时,删去该路段。

最终匹配效果如图 1 所示。符号“X”为原始点,“+”为最终匹配点,车辆从左至右行驶。可以看到,匹配算法通过考虑道路的连通性等约束条件,准备匹配图 1 中右边的 GPS 点;由于左侧的 GPS 点准确匹配,通过反向检查机制修正了图 1 中部的 GPS 点。

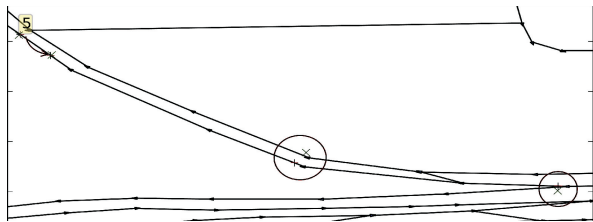


图 1 地图匹配效果图

3 道路平均行程速度估计

利用车载 GPS 数据估计单车行程速度的模型主要有速度积分模型^[11]和位置插值模型^[12-13]。前者利用 GPS 瞬时速度序列进行积分估计单车路段行程速度,后者利用路段边界两侧 GPS 定位点的位置、时间信息,采用内插法估计车辆通过路段边界的时间,进而计算出单车路段行程速度。但高频采样的 GPS 数据是以上两个模型的前提,否则计算得到的平均行程速度精度会显著下降。

此外,由于出租车的运营特性^[14],存在空车寻客、停运、载客上下车等情况,路段上的 GPS 数据不能完全反映当时的道路状况信息。如果单纯使用载客状态下的 GPS 数据,会因为数据量太少而降低了道路平均行程速度估计精度。

文中针对出租车运营特性及其 GPS 数据特点,首先对能用于道路速度的有效数据进行获取,再根据位置插值模型进行改进,使之能适用于低频采样的数据。

3.1 有效数据的获取

GPS 有效数据的不足,一是受出租车数量和采用频率低所限制,二是受出租车空车行驶的影响。对实验数据进行统计发现,白天出租车空车数约占总数的 42.5%~48.7%。对于这些 GPS 数据,除去停车或是载客上下车等原因,大多空车数据都较可靠,特别是快速路和高速公路上的空车数据,可以直接用于速度计算。另一方面,由于出租车存在载客上下车的情况,使得运营状态变化时段内载客状态下的 GPS 数据也变得不可靠(存在加减速的时间差)。因而有效数据主要分两部分:一是有效空车数据的获取,二是无效载客数据的删除。

文中通过对现有出租车载客上下车和空车数据情况进行分析,结合出租车实际情况,假设空出租车不会去拥堵路段寻找乘客,提出了利用 GPS 数据的瞬时速度、前后 GPS 数据的时间差、交叉路口以及交通信号灯位置等信息来判断出租车数据有效性的处理算法,从而筛选出有效 GPS 数据。

有效数据的判断流程如图 2 所示。

图中出租车瞬时车速限值($V_{\text{限}}$)的取值为当前路段等级拥堵车速的上限值;时间差 $t_{\text{后}}$ 指后续 GPS 数据与当前 GPS 数据的时间差。有效数据判断过程将车辆 GPS 序列 $\{g_1, g_2, \dots, g_n\}$ 处理为有效数据连续序列集合 $\{g_{11}, g_{12}, \dots, g_{1k_1}\}, \dots, \{g_{m1}, g_{m2}, \dots, g_{mk_m}\}$ 。

经过有效数据判断获取处理,可用数据量与纯粹的载客数据量相比有了较大提升,整体数据量提升约 18%。分析结果表明:该有效数据的获取算法能够满足一般交通信息发布需求,如城市道路车速分级显示。

图 3 为有效数据统计结果。

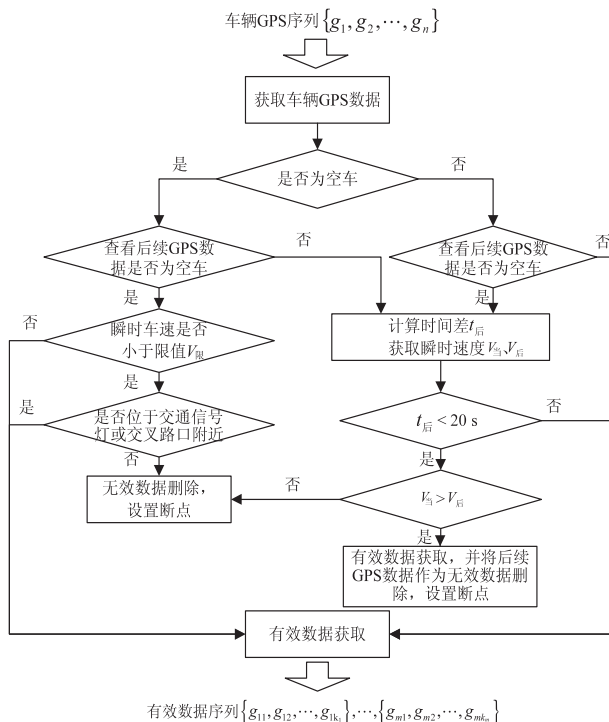


图 2 有效数据判断流程图

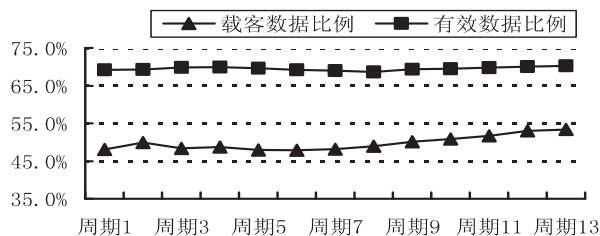


图 3 有效数据比较结果

3.2 改进的距离-时间速度计算方法

在对某一路段的单车 GPS 序列进行分析时,考虑加入相邻路段的连续 GPS 数据点来对车辆通过路段端点的时间进行判断。

在上一节中,通过对车辆原始 GPS 序列的处理,得到了有效数据序列集 $\{g_{11}, g_{12}, \dots, g_{1k_1}\}, \dots, \{g_{m1}, g_{m2}, \dots, g_{mk_m}\}$,有效避免了出租车运营特性造成的影响。文中基于有效数据序列集所覆盖的距离 s_1, s_2, \dots, s_m ,以及所花费的时间 t_1, t_2, \dots, t_m 提出了改进的距离-时间速度计算方法,公式如下:

$$v = \frac{\sum_{k=1}^m s_k}{\sum_{k=1}^m t_k}$$

(1) $1 < i < m$ 时,若有效序列 $\{g_{i1}, g_{i2}, \dots, g_{ik_i}\}$ 中 GPS 数据点个数不为 1, 则

$$\begin{cases} s_i = \text{dis}(g_{i1}, g_{ik_i}) \\ t_i = \text{time}(g_{i1}, g_{ik_i}) \end{cases}$$

其中, $\text{dis}()$, $\text{time}()$ 分别表示路段上两 GPS 数据点间的距离和时间差。若 k_i 为 1, 即只有一个 GPS

数据点时,根据瞬时速度 $v_{\text{瞬}}$ 设定最小行驶距离 s_i , 则 $t_i = s_i/v_{\text{瞬}}$ 。

(2) $i = 1, m$ 时,则可能存在前后两个 GPS 数据点分别落在不同路段上,如图 4 所示。路段 1 表示相邻路段,路段 2 表示需进行速度估计的路段。 L, s 分别表示序列中相邻路段 GPS 数据点 g_{i1} 到路段端点的距离、GPS 数据点 g_{ik_i} 到路段端点的距离。

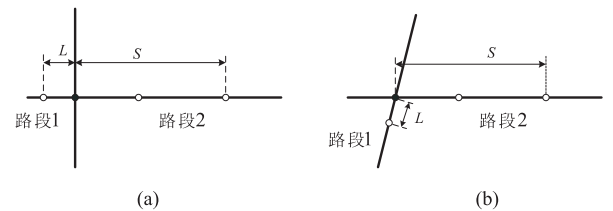


图 4 路段示意图

由于车辆出入交叉口会适当降速通过,文中设定交叉口范围为 30 m,并假设车辆在此范围内匀速行驶。若 $L > 30$ m,则当 g_{i1} 不在交叉口范围内,由于低频采样,两连续点间速度变化不可知,不考虑该点信息,则有

$$\begin{cases} s_i = \text{dis}(g_{i2}, g_{ik_i}) \\ t_i = \text{time}(g_{i2}, g_{ik_i}) \end{cases}$$

若 $L < 30$ m,则 g_{i1} 位于交叉口范围,根据文中假设,当车辆行驶方向未发生明显变化,如图 4(a)所示,则有

$$\begin{cases} s_i = s \\ t_i = \text{time}(g_{i1}, g_{ik_i}) - \frac{L}{v_{i1}} \end{cases}$$

其中, v_{i1} 为 g_{i1} 的瞬时速度。当车辆转弯,如图 4(b)所示,则有

$$\begin{cases} s_i = s - L \\ t_i = \text{time}(g_{i1}, g_{ik_i}) - \frac{2L}{v_{i1}} \end{cases}$$

3.3 路段平均行程速度计算

对于一个路段来说,它可以分为几个区域,各个区域之间的交通状况可能差别较大,行驶速度的变动也较大。那么对于路段行程速度的计算,需综合考虑在此路段上行驶的所有车辆样本速度。在前文获取的单车行程速度中,得到的速度结果均基于车辆行驶路程,因而充分考虑其在路段速度估计中的影响。

假设一条路段上,有 n 个车辆速度样本 (V_1, L_1) , (V_2, L_2) , \dots , (V_n, L_n) , 其中 V_n 为单车行程速度, L_n 为该车覆盖行驶路程。则路段行程速度为

$$V = \frac{\sum_{i=1}^n L_i}{\sum_{i=1}^n (L_i/V_i)}$$

相比传统的均值法,它缓和了高速、低速样本对最

终结果带来的影响,同时考虑文中行程距离的因素,使结果更加精确。

4 实际算例

为了验证文中所提出的计算方法,选取某主干道路段为研究对象进行实验。该路段全长约 381 m,两端出口为交叉路口,与大部分路段类似,具有代表性。选取该路段低峰时段及高峰时段的出租车 GPS 数据为实验数据,以 5 min 为计算间隔,对路段行程速度进行计算,并与实测值进行对比,结果见表 2、3。实测数据由同周期的多辆高频高精度的浮动车数据计算得到,可认为是实际数据。

表 2 低峰时段路段平均行程速度

周期	车辆样本数	路段估计行程速度/(km/h)	路段实测行程速度/(km/h)
周期 1	8	39.18	35.1
周期 2	8	39.89	36.4
周期 3	9	41.28	36.7
周期 4	8	39.86	35.8
周期 5	10	37.44	35.9
周期 6	5	44.24	39.0
周期 7	7	36.29	34.6
周期 8	6	35.89	39.8

表 3 高峰时段路段平均行程速度

周期	车辆样本数	路段估计行程速度/(km/h)	路段实测行程速度/(km/h)
周期 1	13	16.78	19.4
周期 2	15	25.21	22.9
周期 3	17	20.17	22.8
周期 4	13	26.52	28.1
周期 5	11	25.81	23.2
周期 6	13	20.86	23.1
周期 7	11	17.43	20.7
周期 8	10	18.13	21.3

表 2、3 的结果表明,利用文中所提出的方法计算得到的估计值整体上接近于实测值,可以适应于道路状态的判断。当然也存在一些误差较大的数据,经分析主要原因有两个:一方面与路段上出租车样本数过少有关,导致结果有所偏差;另一方面与出租车特点相关,一般在较畅通的路段上,出租车速度普遍大于车辆平均行驶速度。但从整体上看,估计结果可信度较高。

5 结束语

出租车 GPS 数据是现实中常见的浮动车数据,但

其采样频率普遍较低,且存在诸多无效数据。文中根据出租车运营特性,提出了在定位误差、低频采样、数据量少等因素影响下的路段行程速度的估计方法,并通过实例证明该方法能得出与实际值较为接近的估计值,切实可行。文中所提方法充分利用了现有的数据资源,具有实际意义。但由于出租车行驶情况特殊,计算结果有所偏离,还需要做进一步研究和改进。

参考文献:

[1] Bishop R. Floating car data projects worldwide;a selective review[C]//Proceeding of ITS America annual meeting. [s. l.]:[s. n.],2004:192-197.

[2] Emam E B, Ai-Deek H. Using real-life dual-loop detector data to develop new methodology for estimating freeway travel time reliability[J]. Transportation Research Record:Journal of the Transportation Research Board,2006,1959(1):140-150.

[3] Quiroga C A. An integrated GPS-GIS methodology for performing travel time studies[R]. [s. l.]:[s. n.],2002.

[4] 陈青. 基于GPS浮动车的城市道路交通状态判别技术研究[D]. 西安:长安大学,2009.

[5] Zhu Yanmin,Li Zhi,Zhu Hongzi,et al. A compressive sensing approach to urban traffic estimation with probe vehicles[J]. IEEE Transactions on Mobile Computing,2013,12(11):2289-2302.

[6] 张和生,张毅,胡东成. 路段平均行程时间估计方法[J].

交通运输工程学报,2008,8(1):89-96.

[7] 姜桂艳,常安德,张玮,等. 基于GPS浮动车的自然路段行程时间估计方法[J]. 公路,2009(11):87-90.

[8] 朱鲤,杨东援. 基于低采样频率浮动车的行程车速信息实时采集技术[J]. 交通运输系统工程与信息,2008,8(4):42-48.

[9] Thiagarajan A,Ravindranath L,LaCurtis K,et al. VTrack:accurate,energy-aware road traffic delay estimation using mobile phones[C]//Proceedings of the 7th ACM conference on embedded networked sensor systems. Berkeley:ACM,2009:85-98.

[10] Caveney D S. Method and system for predicting a future position of a vehicle using numerical integration; U. S. ,7865299(B2)[P]. 2011-01-04.

[11] 李筱菁,孟庆春,魏振钢,等. GPS技术在城市交通状况实时检测技术中的应用[J]. 青岛海洋大学学报:自然科学版,2002,32(3):475-481.

[12] 董红召,吴方国. 基于FCM的路段平均行程时间估计[J]. 科技通报,2011,27(3):426-430.

[13] Quiroga C A,Bullock D. Travel time studies with global positioning and geographic information systems; an integrated methodology[J]. Transportation Research Part C Emerging Technologies,1998,6(1):101-127.

[14] 于德新,高学英,杨兆升. 基于GPS数据及车辆运行特性分析的单车路段行程时间估计[J]. 吉林大学学报:工学版,2010,40(4):965-970.

(上接第14页)

始终不变的,只能用于短期预测;

(2)通常对历史统计数据(即原始数据序列)的要求在50个以上;

(3)实际工作中,应获取更多的数据,对已建立的预测模型用新的真实值进行检验,并且要不断向序列中加入新的真实值,来修正或重新拟合更能反映真实情况的黄金走势预测模型。

参考文献:

[1] Box G,Jonkins G. Time series analysis forecasting and control[M]. San Francisco:Holden-Day,1970.

[2] Helfenstein U. Box-Jenkins modelling in medical research[J]. Stat Methods Med Res,1996,5(1):3-22.

[3] 杨叔子. 时间序列分析的工程应用(上下册)[M]. 第2版. 武汉:华中科技大学出版社,2007.

[4] Clements M P,Hendry D F,陆懋祖. 预测经济时间序列[M]. 北京:北京大学出版社,2008.

[5] 时照华,苏虹,秦风云,等. ARIMA模型在常见呼吸道传染病疫情预测中的应用[J]. 安徽医科大学学报,2013,48(7):783-786.

[6] Cardinal M,Roy R,Lamber J. On the application of integer

valued time series models for the analysis of disease incidence[J]. Statistics in Medicine,1999,18(15):2025-2039.

[7] 徐国祥. 统计预测和决策[M]. 上海:上海财经大学出版社,1998:150-177.

[8] VO THI THANH VAN. 时间序列分析技术的研究[D]. 长沙:湖南大学,2013.

[9] 王振龙. 时间序列分析[M]. 北京:中国统计出版社,1999.

[10] 王锐. 多种变量选择方法在ARMA阶数确定中的比较[D]. 济南:山东大学,2013.

[11] Sun Zhangzhen,Xu Tianhe,Mo Yijun,et al. Prediction of polar motion based on combination of weighted least-squares and autoregressive moving average[C]//第五届中国卫星导航学术年会论文集-S8 卫星导航模型与方法. 出版地不详:出版者不详,2014.

[12] Chen Jing,He Guozhu,Yu Ping. Prediction of retail price index of Sichuan Province based on ARIMA model[D]. Ya'an:Sichuan Agriculture University,2014.

[13] 罗伯特·卡弗. 探索性数据分析—基于JMP软件[M]. 上海财经大学统计与管理学院,译. 引进版. 上海:上海财经大学出版社有限公司,2013.

[14] Tsay R S. 金融时间序列分析[M]. 第3版. 北京:人民邮电出版社,2014.

基于出租车GPS数据的道路平均速度估计方法

作者：[丁涛杰](#)，[史殿习](#)，[李永谋](#)，[DING Tao-jie](#)，[SHI Dian-xi](#)，[LI Yong-mou](#)

作者单位：[国防科学技术大学 计算机学院, 湖南 长沙, 410073](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

年，卷(期)：2015(7)

引用本文格式：[丁涛杰](#). [史殿习](#). [李永谋](#). [DING Tao-jie](#). [SHI Dian-xi](#). [LI Yong-mou](#) [基于出租车GPS数据的道路平均速度估计方法](#)[期刊论文]-[计算机技术与发展](#) 2015(7)