

面向浮动车取样偏差修正的数据融合方法

李慧兵¹, 杨晓光²

(同济大学 交通运输工程学院, 上海 201804)

摘要: 考虑到浮动车取样偏差会给路段平均行程时间估计值带来很大误差这一事实, 提出了一个融合模型, 该模型是在真实路段平均行程时间计算模型的基础上推导出来的, 其融合了线圈数据和浮动车数据, 减少了浮动车取样偏差对路段平均行程时间估计值的影响. 最后利用仿真实验对一条具有连续5个交叉口的主干道进行仿真实验, 实验证明, 相对于浮动车法, 融合方法无论在平峰期和高峰期都可以得到精度更高的平均行程时间估计值.

关键词: 路段平均行程时间; 数据融合模型; 取样偏差; 仿真实验

中图分类号: TB114.2

文献标识码: A

Data Fusion Method on Modifying Sampling Bias of Floating Cars

LI Huibing¹, YANG Xiaoguang²

(College of Transportation Engineering, Tongji University, Shanghai 201804, China)

Abstract: On the account of that sampling bias of floating cars may bring a lot of errors to the estimated average link travel time, the paper presents a data fusion model derived from average link travel time model based on the population vehicles. By fusing loop detector data and floating car data (FCD), the effect of biased sampling of floating cars on average link travel time decrease with the model. At last, a simulation of an urban arterial road with five intersections and four links shows that data fusion model can yield a better result than the traditional floating car method.

Key words: average link travel time (ALTT); data fusion model; sampling bias; simulation

路段平均行程时间作为反映道路拥挤程度的关键交通参数之一, 历来受到交通管理与控制部门的

高度重视. 路段平均行程时间的精确度、完整性与时效性, 直接影响动态交通管理措施的应用效果.

不同数据源需要不同的模型算法来估计和预测路段平均行程时间值, 考虑在我国大中城市路网上, 浮动车数据具有数据精度高, 覆盖范围广, 成本低等优点, 因此, 对路段平均行程时间的估计和预测来说, 目前浮动车数据应用最广^[1]. 然而由于浮动车样本量的限制^[2], 以及浮动车参数计算模型的局限性^[3], 单纯利用浮动车数据计算得到的路段平均行程时间 (average link travel time, ALTT) 的精度常常无法满足先进交通管理系统 (advanced traffic management system, ATMS) 系统的需求.

前人的研究表明: 浮动车样本的路段平均行程时间可以准确地反映穿越路段全体车辆的路段平均行程时间^[4-6]. 然而 Sen 等^[7-8] 分析了从实地采集到的浮动车数据, 在对浮动车路段平均行程时间进行统计分析之后, 得到这样的结论: 由于浮动车样本的时空分布具有随机性, 因此, 样本采样总是有偏差的, 所以无论浮动车样本量有多大, 浮动车样本的路段平均行程时间都不会接近全体车辆的路段平均行程时间. Hellinga 等^[9] 近期的文章给出了与 Van Aerde 等^[4] 和 Sen 等^[7] 不同的结论, 但 Van Aerde 等^[4] 和 Sen 等^[7] 的结论本身并没有错误, 它们都只是适用于特定交通和浮动车样本量情况. 这些文章^[7-9] 均表明: 由于浮动车样本的偏差, 不管浮动车数据的样本量有多大, 单纯利用浮动车数据计算出来的路段平均行程时间都可能无法逼近全体车辆的路段平均行程时间.

在参考传统浮动车模型和真实路段平均行程时间模型的基础上, 本文提出一种基于浮动车数据和线圈数据的数据融合模型, 然后对模型提出的小时段进行划分, 最后对模型进行实例验证.

收稿日期: 2012-08-17

基金项目: 国家“八六三”高技术研究发展计划(2008AA11Z205)

第一作者: 李慧兵(1983—), 男, 博士生, 主要研究方向为交通数据挖掘与融合. E-mail: lihuibing_1983@yahoo.com.cn

1 融合数据环境

为了把上述融合模型应用于实际中,路网中的线圈应该采用实时自适应交通控制系统的线圈布设方式,这些线圈可以检测到每 1 s 内交叉口上游出口的断面流量、占有率等数据;运行在路网上的浮动车实时发送(数据采样间隔为 1 s)浮动车数据,该数据包括车辆经纬度、时间、点速度、车辆编号等.由于浮动车每 1 s 发送一次数据,通过比对每辆浮动车的实时经纬度与检测线圈的位置,就可以得到每辆浮动车通过线圈检测器的具体时刻.

2 融合模型建立

本融合模型是基于以下假设建立起来的:在某小时段内到达的每辆浮动车的运行轨迹(路段行程时间)同与该浮动车在同一小时段内到达的其他车辆的运行轨迹(路段行程时间)相似.

2.1 传统浮动车模型

基于浮动车数据的路段平均行程时间传统模型如下:

$$\bar{T}_{ij}(f) = \frac{\sum_{n=1}^m t_{ij,n}}{m}, \quad n = 1, 2, \dots, m \quad (1)$$

式中: $\bar{T}_{ij}(f)$ 是在计算间隔 i (5 min) 内,路段 j 的平均行程时间; $t_{ij,n}$ 是指第 n 辆浮动车在间隔 i 内,路段 j 的行程时间; m 是在间隔 i 内通过的浮动车数.

由于传统模型未考虑浮动车取样偏差对路段平均行程时间计算的影响,因此单纯利用浮动车数据计算得到的路段平均行程时间存在较大误差.

2.2 真实路段平均行程时间模型

假设路段平均行程时间的计算间隔为 5 min(全文同),那么路段上全体车辆的路段平均行程时间(路段平均行程时间真实值)可表示如下:

$$\bar{T}_{ij} = \frac{\sum_{n=1}^m k_{ij,n} \bar{T}_{ij,n}}{\sum_{n=1}^m k_{ij,n}}, \quad n = 1, 2, \dots, m_1 \quad (2)$$

式中: \bar{T}_{ij} 是在计算间隔 i (5 min) 内,路段 j 的真实平均行程时间; $k_{ij,n}$ 是在间隔 i 的第 n 个小时段内,通过路段 j 的车流量; $\bar{T}_{ij,n}$ 是在间隔 i 的第 n 个小时段内,通过路段 j 全体车辆的平均行程时间; m_1 是计算间隔 i 内包含的小时段数.

2.3 融合模型

参照式(2),把其中小时段内真实平均行程时间替换为单辆浮动车的行程时间,即可得到本文的融合模型.该模型可有效融合线圈流量数据和浮动车行程时间,其表达式如下:

$$\bar{T}_{ij} = \frac{\sum_{n=1}^m k_{ij,n} t_{ij,n}}{\sum_{n=1}^m k_{ij,n}}, \quad n = 1, 2, \dots, m \quad (3)$$

其中,根据每个计算间隔内通过某路段的浮动车数 m ,把该间隔划分为 m 个小时段,每辆浮动车都会在其对应的小时段内通过目标路段,比如,第 1 辆浮动车会在第 1 个小时段通过目标路段,第 2 辆浮动车会在第 2 个小时段通过目标路段, ..., 第 m 辆浮动车会在第 m 个小时段通过目标路段.对比公式(2)和(3),可以看出,要使式(3)的融合结果接近于式(2)的真实路段平均行程时间,就必须使 $t_{ij,n}$ 接近 $\bar{T}_{ij,n}$. Hellinga 等^[9]指出,在浮动车行程时间计算间隔内,不同时刻到达的浮动车会产生不同的延误,比如,在有效红灯间隔内到达交叉口进口道的浮动车与在有效绿灯间隔内到达进口道的浮动车所产生的延误差别很大,甚至在同一有效红灯间隔内,不同时间点到达的浮动车所产生的延误也会有较大差别(在过饱和交通状态下,很多车辆会经历二次排队,这些车辆所产生的延误要比只经历过一次排队的车辆要大得多).为了使 $t_{ij,n}$ 接近 $\bar{T}_{ij,n}$,如何划分计算间隔 i 内的 m 个小时段就显得尤为重要.

3 小时段划分方法

某计算间隔(5 min)内小时段划分的原则在于——使每个小时段内通过的浮动车行程时间 $t_{ij,n}$ 接近于该小时段内通过的全体车辆路段平均行程时间 $\bar{T}_{ij,n}$. 如果 $t_{ij,n}$ 与该小时段内其余车辆的行程时间相接近,那么 $t_{ij,n}$ 就比较接近 $\bar{T}_{ij,n}$ 了.

为了清晰地表述小时段的划分方法,需要设定一条示例路段(图 1),该路段为 1 号交叉口与 2 号交叉口之间,西向东的左转路段.

融合 2 号线圈检测器组检测到的流量数据以及每辆浮动车的行驶轨迹信息(图 2),可以对计算间隔 i 的小时段进行划分.

小时段划分流程如下:

(1)记录在计算间隔 i (5 min) 内,通过示例路段所有浮动车的 ID 号以及浮动车数;

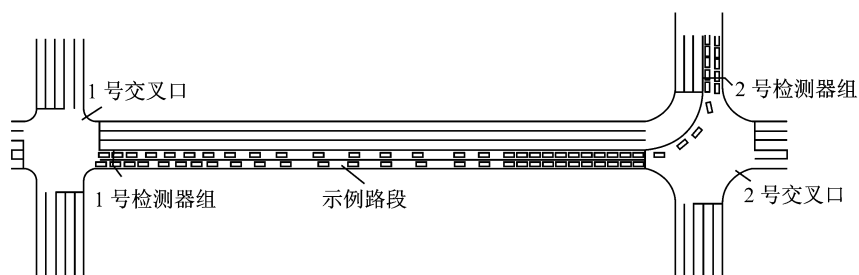


图 1 示例路段示意图

Fig.1 Diagram of object link

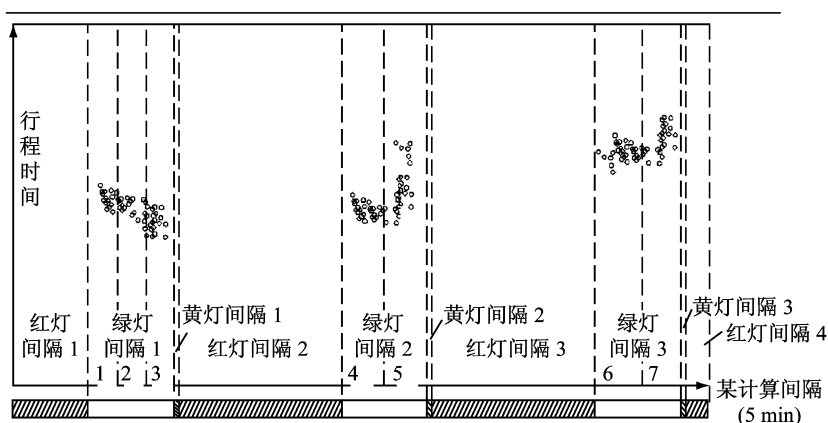


图 2 2号线圈检测器组检测到的流量以及浮动车分布图

Fig.2 Distribution of traffic flow and floating cars detected by the 2nd loop detectors

(2)记录每辆浮动车分别通过交叉口出口断面线圈检测器组 1 和 2 的时刻,它们之间的时间差就是该浮动车穿越示例路段的行程时间;

(3)利用式(1)计算得到基于浮动车数据的路段平均行程时间;

(4)如图 2 所示,横坐标代表某计算间隔的长度(5 min),纵坐标代表行程时间,坐标下方给出了 2 号交叉口左转相位在该 5 min 内的信号配时图.从该图 2 可以看出,有 7 辆浮动车在该时间间隔内通过示例路段.

(5)小时段划分的前提是:每一小时段内均有一辆浮动车通过示例路段.由于在红灯间隔内,不会有车辆通过,因此,不考虑红灯间隔,只考虑 3 个绿灯+黄灯的间隔.在每个(绿灯+黄灯)间隔内,假设有 $k(k \geq 2)$ 辆浮动车通过,则第 1 辆浮动车所属小时段始于该计算间隔 i 的起点,终于第 1 辆和第 2 辆浮动车分别到达 2 号检测器组的时间中点,第 k 辆浮动车所属小时段始于第 $(k-1)$ 辆和第 k 辆浮动车分别到达 2 号检测器组的时间中点,终于该(绿灯+黄灯)间隔的终点.第 $i(i=2,3,\dots,k-1)$ 辆浮动车所属小时段均始于第 $(i-1)$ 辆和第 i 辆浮动车分别到达 2 号检测器组的时间中点,终于第 i 辆和第 $(i+1)$ 辆浮动车分别到达 2 号检测器组的时间中点.

这种划分方法可以保证行驶轨迹相似(或在 2 号交叉口处所经历的延误相近)的车辆被划分在同一小时段.同时,计算间隔 i 内的浮动车样本量越大,利用该融合模型计算得到的路段平均行程时间准确度就会越高,因为此时计算间隔所划分的小时段数变多,某小时段内浮动车样本的行驶轨迹(路行程时间)就会更接近于该小时段内其余车辆的行驶轨迹(路段行程时间).

4 实例分析

4.1 数据来源

本文以南京市中山路南北向的道路(介于建宁路与汉中路之间)作为实际参照路网,利用 Vissim 生成该路网上的各种原始数据,包括浮动车数据、线圈数据、各路段的实际行程时间等.如图 3 所示,仿真路网为一条主干路,包括 5 个交叉口和 4 条路段,所有交叉口的信号控制方式都为定时控制,信号配时是典型的 4 相位配时方案,信号周期均为 120 s.最后,该路网生成一个平峰和一个晚高峰(10:00~10:45 和 17:40~18:25)两个时段的数据,其中浮动车占全体车辆的百分比介于 1%和 10%之间.本文的研究对象为路段 1 左转相位所在的路段(介于广

州路与汉中路之间,长度为 400 m),研究的目标变量是该路段的平均行程时间(图 3).

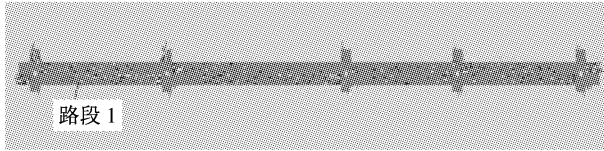


图 3 仿真路网

Fig.3 Road network based on Vissim simulation

4.2 实验结果分析

本文使用相对误差(relative error, RE)和平均相对误差(mean relative error, MRE)来对融合模型的有效性进行评价,表示如下:

$$RE_i = \frac{|A_i - E_i|}{A_i} \times 100\% \quad (4)$$

$$MRE = \frac{\sum_{i=1}^n RE_i}{n}, \quad i = 1, 2, \dots, n \quad (5)$$

式中:RE_i 为第 i 个计算间隔内的路段行程时间相对误差;A_i 为第 i 个计算间隔下的实际路段行程时间;

E_i 为第 i 个计算间隔下的路段行程时间估计值;n 为评价时段内的计算间隔数.

根据 Vissim 获取的路网原始数据,通过传统浮动车模型和融合模型分别计算研究路段的平均行程时间,得到一个平峰和一个晚高峰(10:00~10:45 和 17:40~18:25 两个时段)下的路段平均行程时间,见表 1—2;分别计算这两种模型下的行程时间相对误差(表 3—4),同时绘制相应的曲线图(图 4—7).其中,基于传统浮动车模型的路段平均行程时间估计值用 ATF 表示,基于融合模型的路段平均行程时间估计值用 ATD 表示,路段平均行程时间真实值用 RAT 表示;基于传统浮动车模型的路段平均行程时间相对误差用 REF 表示,基于融合模型的路段平均行程时间相对误差用 RED 表示.

在平峰期,基于传统浮动车模型计算得到的路段平均行程时间 MRE 为 14.75%,基于融合模型计算得到的路段平均行程时间 MRE 为 5.01%.

在高峰期,基于传统浮动车模型计算得到的路段平均行程时间 MRE 为 16.62%,基于融合模型计算得到的路段平均行程时间 MRE 为 7.23%.

表 1 在平峰期,不同模型下计算得到的路段平均行程时间

Tab.1 ALTTs based on different calculation methods during the off-peak period

指标	第 1 时段	第 2 时段	第 3 时段	第 4 时段	第 5 时段	第 6 时段	第 7 时段	第 8 时段	第 9 时段
ATF	80.48	51.09	62.70	80.48	62.79	65.25	71.43	74.35	68.38
ATD	70.42	63.49	75.19	66.33	70.92	71.94	83.16	68.73	59.26
RAT	71.43	60.06	69.57	71.17	67.68	69.08	78.13	71.68	61.73

表 2 在高峰期,不同模型下计算得到的路段平均行程时间

Tab.2 ALTTs based on different calculation methods during the peak period

指标	第 1 时段	第 2 时段	第 3 时段	第 4 时段	第 5 时段	第 6 时段	第 7 时段	第 8 时段	第 9 时段
ATF	98.77	69.93	129.87	89.69	76.34	98.52	106.38	82.47	98.15
ATD	92.81	76.78	110.80	125.79	85.65	81.30	88.11	105.54	71.30
RAT	86.58	82.47	106.10	119.40	81.14	86.20	94.79	99.01	78.74

表 3 在平峰期,不同模型下计算得到的路段平均行程时间相对误差

Tab.3 RE of ALTTs based on different calculation methods during the off-peak period

指标	第 1 时段	第 2 时段	第 3 时段	第 4 时段	第 5 时段	第 6 时段	第 7 时段	第 8 时段	第 9 时段
REF	12.48	24.28	19.93	17.57	12.84	10.24	16.38	7.52	13.30
RED	1.27	5.64	8.03	6.83	4.76	4.09	6.45	4.08	3.93

表 4 在高峰期,不同模型下计算得到的路段平均行程时间相对误差

Tab.4 RE of ALTTs based on different calculation methods during the peak period

指标	第 1 时段	第 2 时段	第 3 时段	第 4 时段	第 5 时段	第 6 时段	第 7 时段	第 8 时段	第 9 时段
REF	12.34	17.94	18.30	33.13	6.29	12.5	10.90	20.05	18.11
RED	6.71	7.42	4.24	5.07	6.42	6.03	7.58	6.19	10.43

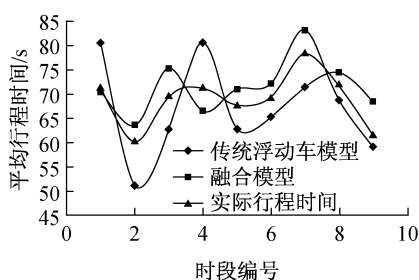


图 4 在平峰期,不同模型下计算得到的路段平均行程时间曲线图

Fig. 4 Curves of ALTT based on different models during off-peak period

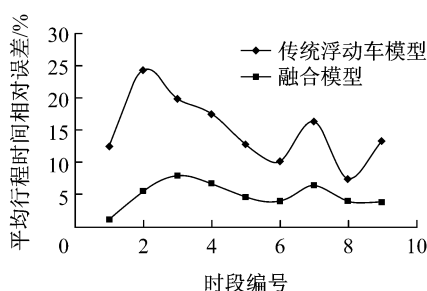


图 6 在平峰期,不同模型下计算得到的路段平均行程时间 RE 曲线图

Fig. 6 Curves of RE of ALTT based on different models during off-peak period

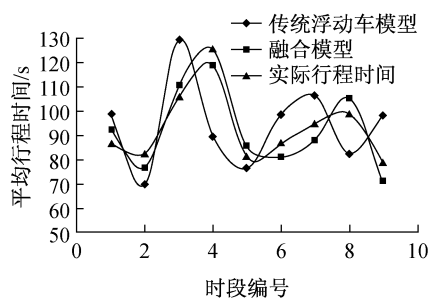


图 5 在高峰期,不同模型下计算得到的路段平均行程时间曲线图

Fig. 5 Curves for ALTT based on different models during peak period

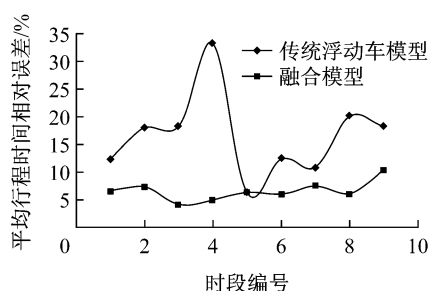


图 7 在高峰期,不同模型下计算得到的路段平均行程时间 RE 曲线图

Fig. 7 Curves of RE of ALTT based on different models during peak period

5 结论

(1) 本文首先基于传统浮动车模型对路段平均行程时间进行估计,其相对误差较大(平峰期 MRE 为 14.75%;高峰期 MRE 为 16.62%);建立了一个融合模型,利用该模型分别在平峰时段和高峰时段对路段平均行程时间进行估计,与传统浮动车模型相比,其行程时间相对误差显著减小(平峰时段 MRE 为 5.01%;高峰时段 MRE 为 7.23%),基本满足出行者对实时信息精度的要求,因此可以作为改进的浮动车计算模型应用于 ITS 智能交通系统中。

(2) 该模型可以很大程度上消除浮动车的取样偏差,因此具有较大的理论价值和实际应用价值;但是出于理论研究的必要性,仿真路网中的交叉口采用的是固定信号配时,实时自适应交通控制系统内的交叉口信号配时实行实时微调,因此该模型在该系统中的实际应用效益还需要进行实地验证^[10]。

(3) 研究中浮动车数据的采样间隔为 1 s,但实际中浮动车数据的采样间隔一般为 10—30 s,因此,需要修正该模型以便将其更好地应用于实践。当浮

动车数据发送间隔比较大(10—30 s)时,可通过融合交叉口信号配时信息及每个浮动车点的状态信息,结合动态的路段划分,来获取每辆浮动车通过线圈检测器的时刻^[11-12]。

(4) 目前许多城市采用的不是实时自适应交通控制系统,并且线圈检测器很容易毁坏,因此如何利用该融合模型对浮动车数据和其他类型检测器(视频检测器,雷达检测器等)检测到的车流量进行融合是一个值得研究的课题。

参考文献:

- [1] 姜桂艳,常安德,吴超腾. 基于 GPS 浮动车的交通信息采集方法[J]. 吉林大学学报:工学版,2010,40(4):971.
JIANG Guiyan, CHANG Ande, WU Chaoteng. Traffic information collection method based on GPS equipped floating car[J]. Journal of Jilin University: Engineering and Technology Edition, 2010, 40(4): 971.
- [2] 张存保,杨晓光,严新平. 基于浮动车的交通信息采集系统研究[J]. 交通与计机,2006,24(5):31.
ZHANG Cunbao, YANG Xiaoguang, YAN Xinping. Traffic data collection system based on floating cars[J]. Computer and Communications, 2006, 24(5): 31.
- [3] 胡明伟. 基于 GPS 的实时交通信息采集方法的研究[J]. 公路

- 交通技,2007,24(5):121.
- HU Mingwei. Study on GPS-based real-time traffic information collection method[J]. Journal of Highway and Transportation Research and Development,2007,24(5):121.
- [4] Van Aerde M, Hellinga B, Yu L, et al. Vehicle probes as real-time ATMS sources of dynamic O-D and travel time data: large urban systems [C]//Proceedings of the ATMS Conference. Petersburg:[s. n.],1993:207-230.
- [5] Srinivasan K K, Jovanis P P. Oversaturation delay estimates with consideration of peaking [C]//Transportation Research Record 1537, TRB. Washington D C:[s. n.],1996:15-22.
- [6] Turner M S, Holdener D J. Probe vehicle sample sizes for real-time information: the Houston experience [C]//Proceedings of Vehicle Navigation and Information Systems (VNIS) Conference. Seattle:[s. n.],1995:3-9.
- [7] Sen A, Thakuriah P, Zhu X, et al. Frequency of probe reports and variance of travel time estimates [J]. Journal of Transportation Engineering,1997,123(4):285.
- [8] Sen A, Soot S, Ligas J, et al. Arterial link travel time estimation: probes, detectors and assignment-type models [C]//Preprint No. 970943, Transportation Research Board, 76th Annual Meeting. Washington D C: Transportation Research Board,1997:1256-1272.
- [9] Hellinga B, Fu L. Assessing expected accuracy of probe vehicle travel time reports [J]. Journal of Transportation Engineering, American Society of Civil Engineers,1999,125(6):1.
- [10] 王明军,李志恒,张毅. 基于 UTC/SCOOT 的城市交通控制管理平台设计应用 [J]. 公路交通科技,2006,23(6):37.
- WANG Mingjun, LI Zhiheng, ZHANG Yi. Design and application of traffic control management platform based on UTC/SCOOT [J]. Journal of Highway and Transportation Research and Development,2006,23(6):37.
- [11] 姜桂艳,常安德,张玮. 基于 GPS 浮动车的自然路段行程时间估计方法 [J]. 公路,2009,2(5):87.
- JIANG Guiyan, CHANG Ande, ZHANG Wei. Physical link travel-time estimation method based on GPS equipped floating Car [J]. Highway, 2009,2(5):87.
- [12] 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.
- YU Dexin, GAO Xueying, YANG Zhaosheng. Individual vehicle travel-time estimation based on GPS data and analysis of vehicle running characteristics [J]. Journal of Jilin University: Engineering and Technology Edition,2010,40(4):965.
-
- (上接第 1490 页)
- MENG Shangjiu, CAO Wenhai, YUAN Xiaoming. A summary of research on residual strain and pore pressure of soils under earthquake loads [J]. World Information on Earthquake Engineering,2001,17(3):49.
- [6] 于洪治,何广纳,杨斌. 软土震陷特性的实验研究 [J]. 大连理工大学学报,1996,36(1):76.
- YU Hongzhi, HE Guangna, YANG Bin. Experimental investigation of soil seismic deformation property [J]. Journal of Dalian University of Technology, 1996,36(1):76.
- [7] 中华人民共和国水利部. SL237—1999 土工试验规范 [S]. 北京:中国水利水电出版社,1999.
- Ministry of Water Resources of the People's Republic of China. SL237—1999 Geotechnical test specification [S]. Beijing: China Waterpower Press,1999.
- [8] Nagase H, Ishihara K. Effects of load irregularity on the cyclic behavior of sand dynamics and earthquake engineering [J]. Soil Dynamics and Earthquake Engineering. 1987,6(4):239.
- [9] 曹继东,陈正汉,王权民. 厦门软粘土的振陷特性研究 [J]. 岩土力学,2004,25(1):160.
- CAO Jidong, CHEN Zhenhan, WANG Quanmin. Study on characteristics of seismic depression of soft clay in Xiamen [J]. Rock and Soil Mechanics, 2004,25(1):160.
- [10] 朱定华,施建军,陈国兴. 南京新近沉积淤泥质土震陷特性试验研究 [J]. 防灾减灾工程学报,2008,28(1):26.
- ZHU Dinghua, SHI Jianjun, CHEN Guoxing. Experimental study on characteristics of seismic settlement of recently deposited muddy soil in Nanjing City [J]. Journal of Disaster Prevention and Mitigation Engineering, 2008,28(1):26.
- [11] 骆亚生,谢定义,董为民,等. 不同地区黄土振陷变形特性的对比分析 [J]. 陕西水力发电,2001,17(1):4.
- LUO Yasheng, XIE Dingyi, DONG Weimin, et al. Comparative analysis of the vibration deformation behavior of the loess from the different regions [J]. Journal of Shaanxi Water Power, 2001,17(1):4.