

文章编号: 1009-6744 (2008) 04-0030-06

智能交通系统与信息技术

# 一种基于自由流车速的交通数据检验方法

李 江, 孙立军\*

(同济大学 道路与交通工程教育部重点实验室, 上海 201804)

**摘要:** 首先是基于理论上的猜测, 依靠统计学对于固定路段的“自由流车速”进行了稳定性分析和验证, 得到了固定路段的“自由流车速”是一个稳定值的结论; 然后根据这一具有稳定性的交通参数, 采用阈值限定和聚类的方法对疑似问题检测线圈进行了排查, 并结合常理初步制定了判断规则; 为了证实该规则是否有效, 有针对性地进行了实验设计; 并通过多次不同情况下的实验对判定规则进行了检验和调整; 最后对各次的实验结果进行了分析, 归纳出了可靠的问题线圈判定依据。实验结果表明, 在海量交通数据积存的基础上, 利用该方法对可能有问题的检测设备进行甄别具有可行性和科学性。

**关键词:** ITS; 交通数据; 质量控制; 自由流车速

**中图分类号:** U491

**文献标志码:** A

## A Traffic Data Test Method Based on Free-flow Velocity

LI Jiang, SUN Li-jun

(Key Lab of Road and Traffic Engineering, Ministry of Education at Tongji University, Shanghai 201804, China)

**Abstract:** Theory hypothesis is the premise of this paper. The conclusion that “Free Flow Speed” is a stable value in a fixed section of a highway has been drawn depending on the theory hypothesis and analyzing by making use of statistics. This important result can help us move forward to screening questionable loop detectors which have some problems with them and produce distorted data. Threshold values limitation and clustering approaches are used in this paper to exclude and select questionable loop detectors relying on the stable traffic parameter. And initial criterions to screen questionable loop detectors come into being with the help of these approaches and common practice. Experiments have been designed to approve the validity of these initial criterions. Then several necessary experiments in different situations are successfully carried out to validate and adjust the initial criterions. At last, the reliable determinant criterion is concluded after analyzing the results of the experiments. The results of the experiments show that this method is feasible and proper in screening questionable loop detectors based on the mass archived traffic data.

**Key words:** ITS; traffic data; quality control; free flow speed

**CLC number:** U491

**Document code:** A

## 1 引言

在城市的路面上铺设和布设有许多的交通信息采集设备, 这些设备在运行的过程中由于受气

候、自身老化或者设备设置不恰当的影响, 检测的精度会降低, 检测得到的数据可能会发生“漂移”, 这种“漂移”并不是某一项交通参数的“漂移”, 而是

收稿日期: 2008-03-06 修回日期: 2008-04-23 录用日期: 2008-05-07

基金项目: 上海市科委资助项目 (06dz12001)。

作者简介: 李江 (1979 - ), 男, 湖北人, 博士生。

\* 通讯作者: duyuchuantj@126.com

从该检测器获得的全部的交通参数发生的整体性的“漂移”。因此,对于这种情况,无论是逻辑检验、单指标检验还是联合检验都无法检验出数据是错误的<sup>[1-3]</sup>,但事实上,这些数据的确是错误或者是无效的。本文以上海快速路埋地线圈为研究对象展开研究。埋地线圈通常是以20秒为一次采集周期,得到道路断面20秒内的车流量、平均车速和占有率。而一旦数据“漂移”发生,数据的采集处理者无从知晓,这样就会给数据的使用者带来错误的信息,可能会误导交通管理者的判断,给城市交通的运行带来极为不利的影响。通过对自由流车速这一交通参数的分析和研究发现,城市道路中不同断面的自由流车速可能不同,但是固定断面的自由流车速是一个稳定值,依托这一稳定性交通参数,首先对可能存在问题的线圈进行了一些理论上的假设和猜测,然后对有疑问的线圈进行了现场实验,最后对线圈是否工作正常或者说数据是否发生“漂移”作出判断,并给出了基于实验结果的“问题”线圈标准。

## 2 稳定性验证

自由流车速是指“车辆在道路上行驶时,在环境条件(如晴天、雨天;光线充足,光线不足等)、交通条件(交通控制、限速)等影响因素综合作用下,车辆行驶不受其他车辆干扰(与前后车之间保持足够的车头时距),驾驶员根据自己的意愿自由行驶的期望行驶速度”。该定义有别于严格意义上的自由流车速。对于自由流车速的获取,传统的方法有两种:一种是根据实测数据利用一定的数学模型进行K-V关系回归;一种是试验车进行实地测速。通过借鉴国内外对K-V关系模型的研究,不难发现,早期提出的各模型并不适用于快速路的实际情况,而且最近的研究表明,K-V曲线在密度较小时,即自由流状态时,车速是一个较稳定的值,在K-V曲线上是一个平台,而不是其他形式的曲线。所以,通过模型回归得到的自由流车速值,往往会比实际值要大;另外,各模型回归结果的之间稳定性也是比较差,所以说,现有对自由流车速的研究在自由流车速获取的问题上,准确性和稳定性都存在很大的问题。而试验车实测自由流车速的方法,

由于条件所限,往往只能通过进行小样本的调查得出结果。因为样本数少,所以调查的结果不能全面反映在各种影响因素下的自由流车速的真实情况,各次调查结果之间偏差也很大。

统计学<sup>[4]</sup>是一门研究群体现象数量表现内在规律性的科学。通过对已获取的大量自由流状态下的速度数据,应用统计学原理对这些数据进行分析,从而得出这些海量数据的内在规律性,即自由流车速分布的内在规律性。选取断面上20秒检测周期内流量=1(只有一辆车通过,保证足够的车头时距)且占有率<3%(剔除拥挤状态以及养护维修车辆作业的异常情况)的数据作为分析对象。

### 2.1 正态分布检验

以南北高架桥西侧编号为NBDX12的断面2号车道(车道由内向外,车道编号从1开始增加)为例说明数据分析过程。首先从数据库中选取出该断面满足自由流要求的速度数据,对得到的数据进行分组和车速频率统计,并绘制自由流车速频率直方图(如图1所示)。

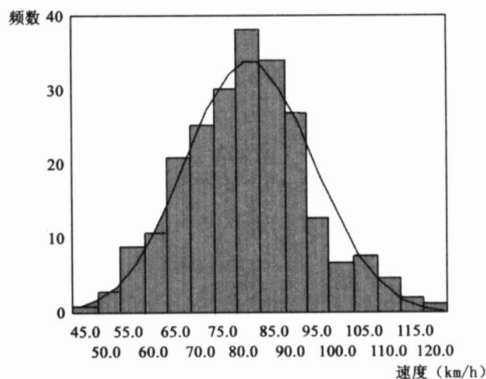


图1 自由流车速频率直方图

Fig. 1 Free flow speed frequency histogram diagram

根据极大似然法,可计算出断面速度频率直方图所对应的正态分布曲线特征值为 $\mu = 80.093\ 62$ ,  $\sigma = 13.877\ 750\ 05$ ,并采用单样本柯尔莫哥罗夫-斯米尔诺夫(Kolmogorov-Smirnov)检验方法,对自由流车速是否服从正态分布进行检验。通过SPSS软件计算可得如表1所示的数据。

双尾检验的显著度水平概率 $P = 0.256$ ,取显著度水平 $\alpha = 0.05$ , $P > \alpha$ ,说明该组自由流车速服从正态分布。

表 1 单样本 Kolmogorov-Smirnov 检验

Table 1 Single sample Kolmogorov-Smirnov test

参数	各指标值
样本数	235
均值	80.093 62
标准差	13.877 750 05
统计量 Z	1.013
双尾检验 P	0.256

注: Test distribution is Normal.

Calculated from data.

FSTR . LOOPG = NBDX12(2) ,group = 1.

### 2.2 稳定性分析

研究某个因素对多个样本均值的影响可以用方差分析的方法实现. 方差分析是用于两个及两个以上样本均数差别的显著性检验. 由于各种因素的影响, 研究所得的数据呈现波动状. 造成波动的原因可分成两类, 一是不可控的随机因素, 另一是研究中施加的对结果造成影响的可控因素. 方差分析的基本思想是: 通过分析研究不同来源的变异对总变异的贡献大小, 从而确定可控因素对研究结果影响力的大小. 采用日期作为影响因素, 研究目的就是要看日期这个影响因素对自由流车速会造成多大程度的波动, 影响是否显著.

表 3 各组数据方差分析结果

Table 3 Result of analysis of variance

方差来源	平方和	自由度	均方和	F 值	F 值对应的显著性水平 P
组间和	166.482	2	83.241	0.452	0.637
组内和	119 502.945	649	184.134		
总和	119 669.428	651			

由于 F 值对应的显著性水平  $p = 0.637 > > 0.05$ , 可以认为各组均值之间无显著差异, 即不同日期分析得到的自由流车速是稳定的. 通过这种方法可以得到每个断面各个线圈的自由流车速.

### 3 判别“问题”线圈

检测得到的线圈数据可能会发生某种“漂移”, 这种“漂移”并不是某一项交通参数的“漂移”, 而是从该检测器获得的所有的交通参数发生的整体性的“漂移”. 对于这种情况, 无论是逻辑检验还是单指标检验都无法检验出数据错误, 对于这种线圈, 称它为“问题”线圈.

进行方差分析一般需符合以下 3 个条件:

被检验的各样本空间均服从正态分布;

各样本空间的方差都相等;

从每一样本空间中抽出的样本是随机且独立的.

由于篇幅所限, 方差分析的详细讲解就不在此赘述. 按照方差分析的过程, 首先检验样本空间的正态性, 这在前面已经验证过. 接下来是验证各组数据的方差齐次性, 验证方法可参考相关文献, 将验证结果汇总于表 2 中.

表 2 各组数据方差齐次性检验结果

Table 2 Result of test of homogeneity

Levene 统计	自由度	自由度	显著度水平
值	1	2	P
0.056	2	649	0.946

由表 2 数据可以看出, 方差齐次检验的显著度水平  $P = 0.946 > > 0.05$ , 即可以认为这几组数据之间方差没有显著差异, 方差是齐性的.

最后, 用 F 检验的方法计算上述各指标值, 并将计算方差分析结果汇于表 3 中.

### 3.1 判别流程的确立

通常认为在多车道断面中, 内车道的自由流速比外车道的自由流车速普遍要稍高, 其实同一断面不同车道的行使速度与道路的线性有着较大的关系, 因此不能通过车道所在的位置来判断哪个车道的自由流车速应该高一些, 哪个车道的自由流车速应该低一些.

通过对自由流车速历史数据进行的全面分析, 发现同一断面的自由流车速差距不大, 由统计得到同一断面中自由流车速最大值与最小值的差值所占的百分比如图 2 所示.

根据图 2 显示的比例, 认为同一断面内的自由流车速的最大值与最小值差值大于 10 时才有可能

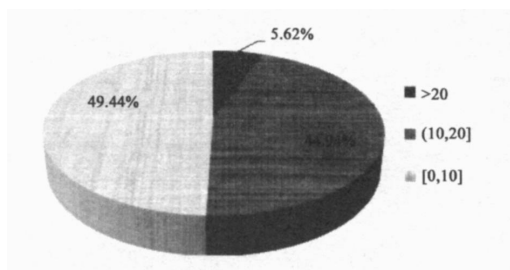


图2 断面内自由流车速最大值与最小值差值所占百分比

Fig.2 Percentage of difference between max and min of free flow speed in one section

是“问题”线圈,因为就一般性而言,出现问题的线圈往往是少数;其次,为了尽可能地把可能发生问题的线圈都包括进去,取差值10作为初始判断阈值是比较合理的。

另外,对同一断面中,中央车道自由流车速小于两侧车道自由流车速的比例也进行了统计分析,结果如图3所示。

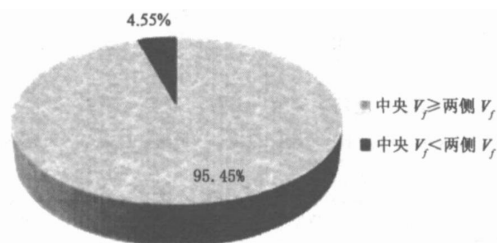


图3 断面内中央车道自由流车速与两侧车道自由流车速大小比较所占比例

Fig.3 Percentage of difference between center lane and two side lanes of free flow speed in one section

从图3可以看出,一般情况下,中央车道的自由流车速都大于或等于两侧车道的自由流车速,因此,暂且认为如果中央车道的自由流车速小于两侧车道的自由流车速,可能是线圈有问题发生。因此,能够得到可以作为判别疑似“问题”线圈的两条初始原则:

同一断面的自由流车速应该大致相当,当最大值与最小值的差值大于10时,该线圈疑为“问题”线圈;

在具有3车道或3车道以上的断面中,中央车道的自由流车速为最小时,该线圈疑为“问题”线圈。

发生数据“漂移”的线圈的稳定性参数值一定有别于其它正常运行的线圈的稳定性参数值,换句话说,如果把正常运行的线圈当成一个集合或者聚类的话,那么发生数据“漂移”的线圈则属于其他的集合或者聚类。引入K-WMeans聚类方法<sup>[5]</sup>,K-WMeans是KMeans的改进型方法,该方法表述如下:

给定一个有 $N$ 个元组或者纪录的数据集,构造 $K$ 个分组, $K < N$ ,每一个分组就代表一个聚类或簇,用 $c_j$ 表示, $1 \leq j \leq K$ 。 $K$ 个分组必须满足下列条件:

每一个分组至少包含一个数据纪录;

每一个数据纪录属于且仅属于一个分组。

随机选 $K$ 个对象作为初始的簇中心,对于每一个簇 $c_j$ 可计算出簇中对象的加权平均值:

$$AWM_j = \frac{1}{t} \sum_{i=1}^t w_i p_i \quad (1)$$

其中 $AWM_j$  ( $1 \leq j \leq K$ )表示簇 $c_j$ 的加权平均值; $t$ 是簇 $c_j$ 中对象的个数,不同的簇 $t$ 值不同; $p_i$ 是空间中的点,表示给定的簇 $c_j$ 中 $t$ 个对象之一; $w_i$ 是簇 $c_j$ 中数据对象的权重。将每个对象赋给最类似的簇,然后对每一类,计算其“质心”(即求每一维坐标的平均值得到的点),用它作为新的代表点,再对所有点分类。经过迭代,直至准则函数 $E$ 收敛使平方误差函数值最小。

$$E = \sum_{j=1}^k \sum_{p_i \in c_j} |p_i - AWM_j|^2 \quad (2)$$

其中 $p_i$ 为 $c_j$ 中的每一个数据。

根据K-WMeans聚类方法将数据集分为两类:一类为正常线圈;另一类为“问题”线圈,由于 $K < N$ ,因此该方法只适用于线圈数大于等于3的断面。另外,依据K-WMeans聚类方法可以进行线聚类判别和空间聚类判别的多重聚类判别。

线聚类判别是对同一断面内的线圈进行聚类,通过对疑似“问题”线圈在断面内的唯一存在性进行判别来进一步确定出该线圈是否为“问题”线圈。如某断面内的3个线圈,通过初始判断原则判断后发现中央自由流车速最小,通过线聚类发现并非只有中央自由流车速最小,因此,不能确定该中央线圈是否真的为“问题”线圈,或者自由流车速最大的线圈可能才是“问题”线圈,需要进一步判别。

空间聚类判别是将疑似“问题”线圈与相临断面相同车道的线圈进行聚类,如果可以聚类,则说明这两个线圈具备某些类似的属性(如线圈所在位置受道路线形、路面平整度等因素的影响一致),则不能确定是否为“问题”线圈,但如果不能与相临断面的线圈进行聚类,则说明该线圈具备“独特”的属性,很可能就是“问题”线圈。判别逻辑流程如图4所示。

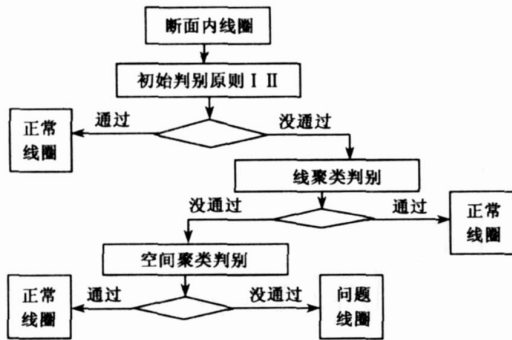


图4 判别逻辑流程图

Fig. 4 Flow chart of estimation logic

### 3.2 实验方案设计

依据上述两个初始原则及判别逻辑流程,对各个断面里不同线圈(车道)的自由流车速进行了对比分析。根据数据分析的结果,结合实际的工程实验条件,选取 NBXX22 断面和 NBXX44 断面之间的路段作为实验路段。该路段包含了 5 个疑似“问题”线圈,见表 4。对该路段进行封道,试验车在“问题”线圈上以某一车速行驶得到  $v_{gi}$  ( $i = 1, 2, 3, n$ ),与此同时采用 PDA 载 GPS(GPS 单点测速精度可达到 mm/s 级<sup>[6]</sup>,完全满足实验要求)每一坐标点的车速进行记录,把线圈的坐标与 GPS 坐标进行匹配,可以得到每一个“问题”线圈的 GPS 测定车速  $v_{gi}$  ( $i = 1, 2, 3, n$ )。将 GPS 测定车速和线圈采集车速进行对比分析,就可以确定出疑似线圈是否就是“问题”线圈。确定原理如图 5 所示。

表 4 为实验路段通过线聚类判别后的断面表,表 5 为实验路段通过空间聚类判别后的断面表,对比表 4 和表 5 可以发现缺少了 NBXX32 和 NBXX33 这两个断面,原因在于 NBXX32 和 NBXX33 这两个前后断面中央线圈的自由流车速可以聚类,如表 4 中黑色方框所框部分所示。

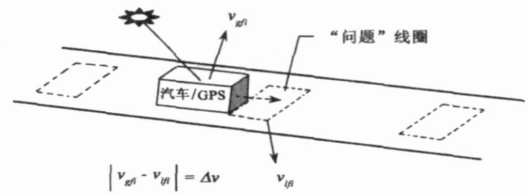


图5 “问题”线圈确定示意图

Fig. 5 Ascertain questionable loop

为了全方位地了解线圈工作的情况和对速度的敏感性,对线圈分别进行高速和低速两次实验,高速行使基本处于自由流状态下,低速行驶车速基本恒定在 30km/h 左右。

表4 断面中疑似“问题”线圈自由流车速

Table 4 Free flow speed of suspected questionable loop in one section

线圈断面	NBXX24	NBXX32	NBXX33	NBXX35	NBXX36
1号线圈	88	85	86	84	92
2号线圈	80	79	81	71	80
3号线圈	90	88	86	86	95
极值差	10	9	5	15	15

表5 空间中疑似“问题”线圈自由流车速

Table 5 Free flow speed of suspected questionable loop in space

线圈断面	NBXX24	NBXX35	NBXX36
1号线圈	88	84	92
2号线圈	80	71	80
3号线圈	90	86	95
极值差	10	15	15

### 3.3 实验结果分析

#### (1) 低速测试实验结果及分析。

表 6 是低速测试的实验结果。从实验结果可以看出线圈 NBXX24(2)、NBXX35(2)和 NBXX36(2)这三处中央线圈的两种测试速度存在差距但差距较小。其他部分线圈的两种速度完全一致。另外,相临的线圈 NBXX32(2)与 NBXX33(2)空间聚类判别通过,线圈不存在问题,这与实验结果也完全一致。因此,低速实验可以说明前文论述和假设的存在性和正确性。

#### (2) 高速测试实验结果及分析。

表 7 是高速测试的实验结果。由于在高速测试实验中断面 NBXX24 没有在封闭路段内,因此

NBXX24(2)在高速测试实验中舍弃。从实验结果可以看出,在高速测试的情况下,线圈的误差对车速具有敏感性,误差的幅度增大。但也有不是线圈NBXX32(2)和NBXX33(2)不符合表4“问题”线圈的规律,但与表5的规律一致,原因在于:对“问题”线圈的选择上约束不够。

表6 低速测试的实验结果

Table 6 Result of low speed test experiment

线圈编号	NBXX24(2)	NBXX32(2)	NBXX33(2)
线圈	27	30	29
GPS	31	30	29
线圈编号	NBXX35(2)	NBXX36(2)	
线圈	24	31	
GPS	30	32	

表7 高速测试的实验结果

Table 7 Result of low speed test experiment

线圈编号	NBXX32(2)	NBXX33(2)
线圈	78	73
GPS	78	73
线圈编号	NBXX35(2)	NBXX36(2)
线圈	63	65
GPS	76	74

## 4 判别标准

通过高低速测试实验可以看出,两次实验的结论一致,但有疑似“问题”线圈并不真是“问题”线圈,问题在于对“问题”线圈的选择上约束不够。由表5和表6可以看出,当断面中最大自由流速对应的线圈与最小自由流速对应的线圈的差值大于等于10并且在空间中独立成类时才是问题线圈。因此,“问题”线圈判别标准为:在具有3车道或3车道以上的断面中,中央线圈的自由流速为最小,并与该断面中最大自由流速的差值大于等于10,且空间独立成类时,该线圈为“问题”线圈。在海量数据积存的基础上,采用稳定性交通参数检验方法对可能有“问题”的检测设备进行甄别是有效的。

## 5 研究结论

检测设备的工作状态直接决定了交通数据质

量的高低,通过数据对检测设备是否存在问题进行甄别对ITS系统的运行具有不言而喻的重要意义。本文从稳定性交通参数自由流速入手,结合实验研究了通过采集到的交通数据来对数据质量进行有效控制的方法,从实验的结果来看,该方法能够有效地指出存在“问题”的检测设备,在数据与采集设备之间建立了有效的联系,对工程应用具有一定的指导意义和实用价值。

但由于实验条件的限制和实验数据的缺乏,同一断面内超过一个以上的线圈发生“飘移”,或者全部线圈都发生“飘移”的情况下该如何进行判别,在文中没有涉及到,这也是下一步要开展的研究工作。

### 参考文献:

- [1] 裴玉龙, 马冀. 实时交通数据的筛选与恢复研究[J]. 土木工程学报, 2003, 36(7): 78 - 83. [PEI Yur-long, MA Ji. Real-time traffic data screening and reconstruction [J]. China Civil Engineering Journal, 2003, 36(7): 78 - 83.]
- [2] Turner S. A simple approach to archiving operations data: case study in austin, texas[C]. TRB 2002 Annual Meeting (CD - ROM) Jan 2002.
- [3] Turner S. Guidelines for Developing ITS Data Archiving Systems[R]. Report 2127 - 3. FHWA, U. S. Department of Transportation, 2001.
- [4] 潘承毅, 何迎晖. 数理统计的原理与方法[M]. 上海: 同济大学出版社, 1992. [PAN Chen-yi, HE Ying-hui. Theory and Method of Mathematical Statistics [M]. ShangHai: Tongji University Press, 1992.]
- [5] 孙士保, 秦克云. 改进的 K- 平均聚类算法研究[J]. 计算机工程, 2007, 33(13): 200 - 209. [SUN Shi-bao, QIN Ke-yun. Research on modified K-means data cluster algorithm[J]. Computer Engineering, 2007, 33(13): 200 - 209.]
- [6] 王甫红, 张小红, 黄劲松. GPS 单点测速的误差分析及精度评估[J]. 武汉大学学报(信息科学版), 2007, 32(6): 515 - 519. [WANG Fu-hong, ZHANG Xiao-hong, HUANG Jin-song. Error analysis and accuracy assessment of GPS absolute velocity determination with SA off[J]. Geomatics and Information Science of Wunan University, 2007, 32(6): 515 - 519.]