

道路交通流数据检验与修复方法

蒋 锐 王 均

(交通部公路科学研究院 北京 100088) (北京交通大学 北京 100044)

交通流数据常见问题分为数据缺失、数据无效和数据时间点3类,针对这些问题,提出了一种交通流数据检验与修复方法。以北京市交通流数据为例,应用这一方法进行了分析与验证。结果表明该方法具有很好的实际应用效果。交通流数据质量检验可以保证从数据源所获得数据的正确性和完整性,为数据的管理和应用提供了可靠的数据基础。

交通流数据;交通流理论;数据检验;城市环路

:U 491

:A

0 引 言

交通流数据中蕴含的交通时空分布规律对有关部门在交通控制、路网规划、交通流理论研究等方面具有很大的参考利用价值,对交通流的基础理论研究和现代交通科学管理与决策起到了重要的辅助作用^[1]。然而,由于种种原因^[2](如设备误差、线路故障、处理错误等),这些数据中存在数据错误、缺失值等现象,对交通流量数据的分析和深层次的数据挖掘带来不利影响,因此需要对交通流数据进行质量检验并对问题进行处理。

1 交通流数据质量检验理论

1.1

1) 数据时间点问题。比如理论上一些交通流数据获取频率应为120 s 1次,但实际上间隔通常小于120 s,如119 s,118 s等,而且因为网络传输问题,没有明显的规律可循。经过时间的积累,会带来较大的误差。如6车道1周的数据理论上应为 $6(\text{车道}) \times 30(\text{次/h}) \times 24(\text{h}) \times 7(\text{d}) = 30\,240$ 条,而实际获得30 462条,多了222条,即37组,74 min的数据。

我们还发现,交通流数据记录顺序有时会改变,如本来各道路记录顺序为:1、2、3、11、12、13,会突然改变为:2、3、1、12、13、11。这对于不识别路段编号,直接间隔固定的行数、循环提取数据的方法产生重要的影响。

如果各个监测点记录顺序的差异较大,没有统一的顺序,就造成了对于不同的检测点数据需使用不同的分析程序代码,也会给科研人员带来不必要的工作量。

2) 数据缺失。由于设备老化、线路故障等原因,交通流数据中偶尔会发生数据点缺失的现象,属于比较常见的问题。通常是丢失1个时间点的1组数据。虽然丢失的数据不多,但如果没有及时发现,对于间隔固定行数循环提取数据的交通流分析方法,有着严重的影响。

3) 数据无效。数据无效指的是获得的交通流数据有数据突变点或不符合交通流3要素相互关系,主要依据阈值理论和交通流理论判定数据的有效性,由于无效的概念比较模糊,相对不易判别。另一个比较常见的问题是,在车流量比较少的时候,常测得不符合实际情况的较高速度,显然是错误数据,需要有针对性地予以处理。

1.2

针对以上问题,提出了一种交通流数据检验与修复方法:为了便于检测数据点的缺失情况,先把数据的时间点标准化,使所有数据具有相同的记录顺序;然后修复丢失数据和无效数据。本检验程序已使用Matlab 6.5实现,总体流程如图1所示。

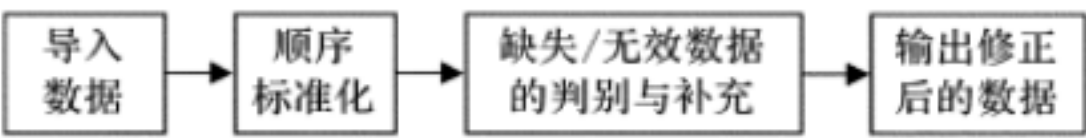


图1 交通流数据检验与修复方法总体流程

1) 数据时间点顺序标准化。统一记录顺序标准为:由外环到内环,路编号由小到大。标准化后,

不但有利于下一步准确发现缺失的数据,而且也
为编写后期分析程序提供了较大便利。算法的详
细流程如下:

- (1) 相关变量初始化。
- (2) 导入交通流数据。
- (3) 初始化交通流数据车道数、编号等特征
信息。
- (4) 由特征信息确定一个标准车道编号排
序,如 1、2、3、11、12、13。
- (5) 从第 1 条数据开始,将每 2 min 内的数据
分为 1 组。
- (6) 将每组内以道路编号为准将数据顺序标
准化。
- (7) 如组内的数据值小于应有车道数,将相
应的数据缺失情况记录到日志文件。
- (8) 依次处理每组数据,直至结束。
- (9) 输出标准化后的数据至文件。

2) 缺失/无效数据的判别与补充。将数据时
间点标准化后,再判别缺失或无效的数据,并予以
补充。通过线性插值法对错误数据和丢失数据进
行修正和补齐^[3]。若 t_j 对应的数据组是错误数据,
 t_i 和 t_k 分别是与 t_j 最为接近的左邻和右邻正确数
据组对应的时间点。则

$$x_{t_j} = \frac{(t_j - t_i)x_{t_k} + (t_k - t_j)x_{t_i}}{t_k - t_i} \tag{1}$$

式中: x_{t_j} 为 t_j 所对应的交通量、速度和占有率修正
后的数据值。算法的详细流程如下:

- ① 相关变量初始化; ② 导入标准化后的交
通流数据; ③ 初始化交通流数据车道数、编号等
特征信息; ④ 依据标准数据判断数据是否缺失;
- ⑤ 如缺失则依据线性插值法补充数据; ⑥ 判断
数据的有效性并修正; ⑦ 依次处理每组数据,直
至结束; ⑧ 输出标准化后的数据至文件。

此交通流数据检验与修复方法的整体流程如
图 2 所示。

2 北京市环路数据质量检验实例

现以北京市环路数据为例,详细介绍本文所
提算法的数据处理效果。表 1 所示为未经处理的
2006-03-04 北京市环路交通流原数据的 2 组(每 2
min 为 1 组)数据,来自二环路上编号为 02023 的
检测点,地点为左安门桥西 800 m 处。

2.1

由表 1 可以看出,数据库中的数据是按照记

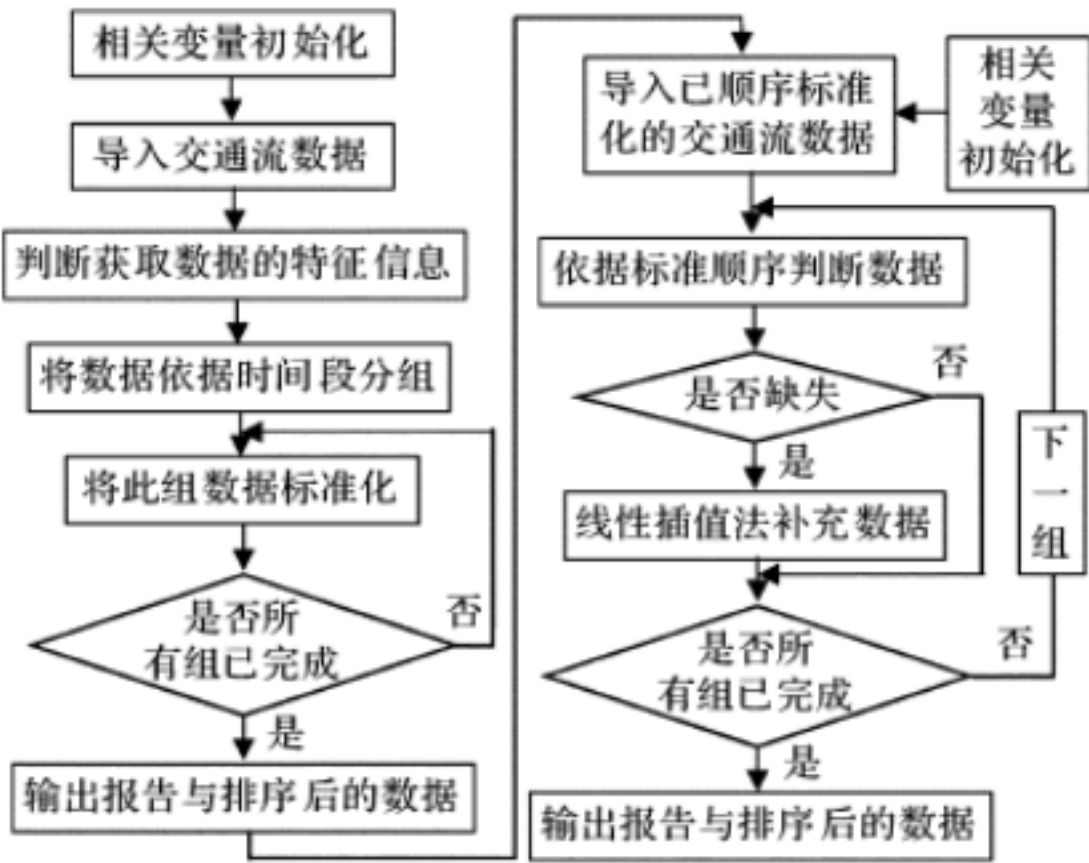


图 2 交通流数据检验与修复方法整体流程图
1 (02023)

编号	日期	时间	道路	流量 /veh	速度/ (km · h ⁻¹)	占有 率/%	大车 数量 /veh
02023	2006-03-04	18: 48: 49	1	53	61	18	2
02023	2006-03-04	18: 48: 49	12	27	64	4	0
02023	2006-03-04	18: 48: 49	13	34	47	6	0
02023	2006-03-04	18: 49: 12	11	35	70	4	0
02023	2006-03-04	18: 49: 12	2	39	52	9	0
02023	2006-03-04	18: 49: 12	3	30	48	7	0
02023	2006-03-04	18: 50: 49	1	56	64	17	2
02023	2006-03-04	18: 50: 49	12	36	59	5	2
02023	2006-03-04	18: 50: 49	13	29	60	5	2
02023	2006-03-04	18: 51: 09	2	48	54	10	0
02023	2006-03-04	18: 51: 09	3	33	43	8	0

录到来的时间先后排序的,并不是按照道路编号
的顺序排序。以表 1 所示数据为例,所示数据道路
顺序为: 1、12、13、11、2、3,而且此顺序还可能会因
为线路传输等问题产生变化。这就给数据的使用
带来了许多不便,我们通过 1. 2 所述的方法将其
标准化为 1、2、3、11、12、13 的顺序。效果如表 2 所
列。

2				(02023)			
编号	日 期	时 间	道路	流量 /veh	速度/ (km · h ⁻¹)	占有 率 /%	大车 数量 /veh
02023	2006-03-04	18: 48: 49	1	53	61	18	2
02023	2006-03-04	18: 49: 12	2	39	52	9	0
02023	2006-03-04	18: 49: 12	3	30	48	7	0
02023	2006-03-04	18: 49: 12	11	35	70	4	0
02023	2006-03-04	18: 48: 49	12	27	64	4	0
02023	2006-03-04	18: 48: 49	13	34	47	6	0
02023	2006-03-04	18: 50: 49	1	56	64	17	2
02023	2006-03-04	18: 51: 09	2	48	54	10	0
02023	2006-03-04	18: 51: 09	3	33	43	8	0
02023	2006-03-04	18: 50: 49	12	36	59	5	2
02023	2006-03-04	18: 50: 49	13	29	60	5	2

虽然修正了时间点顺序,但因未检测缺失的
数据,这会导致数据读取时造成了“串行”的现象。



需要在下一步进行解决。

2.2 /

经过上一节的数据时间点标准化后,再对标准化后的数据进行缺失/无效数据的判别与补充,如在表1、表2中第2组缺失的11车道数据使用线性插值法得到了补充,如表3所列。

3			(02023)				
编号	日期	时间	道路	流量 /辆	速度/ (km·h ⁻¹)	占有 率/%	大车 数量 /辆
02023	2006-03-04	18:51:09	11	36	68	4	1
02023	2006-03-04	18:50:49	12	36	59	5	2
02023	2006-03-04	18:50:49	13	29	60	5	2

分别打开原始数据与完成修正的数据,可以发现修正数据的时间点已被标准化,实现了统一的记录顺序,并且之前缺失或无效的数据也得到了补充和修正。处理后的数据即为最终完成修正的数据,将原始数据与最终完成修正的数据进行比较,效果分别如图3、图4所示。

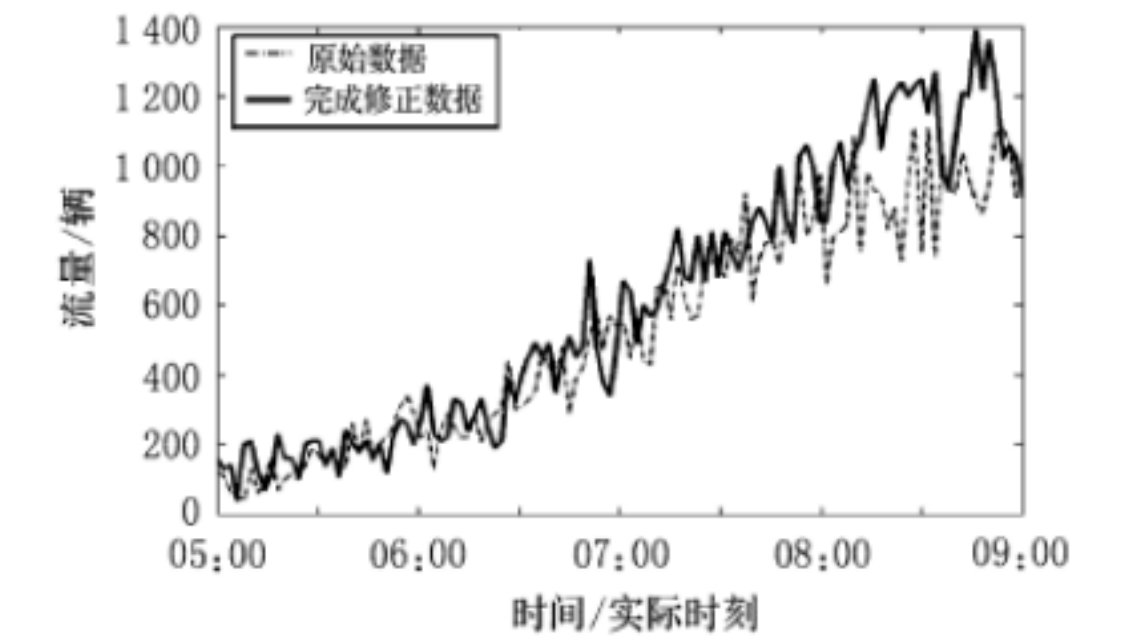


图3 05:00~09:00 原始数据与修正数据对比

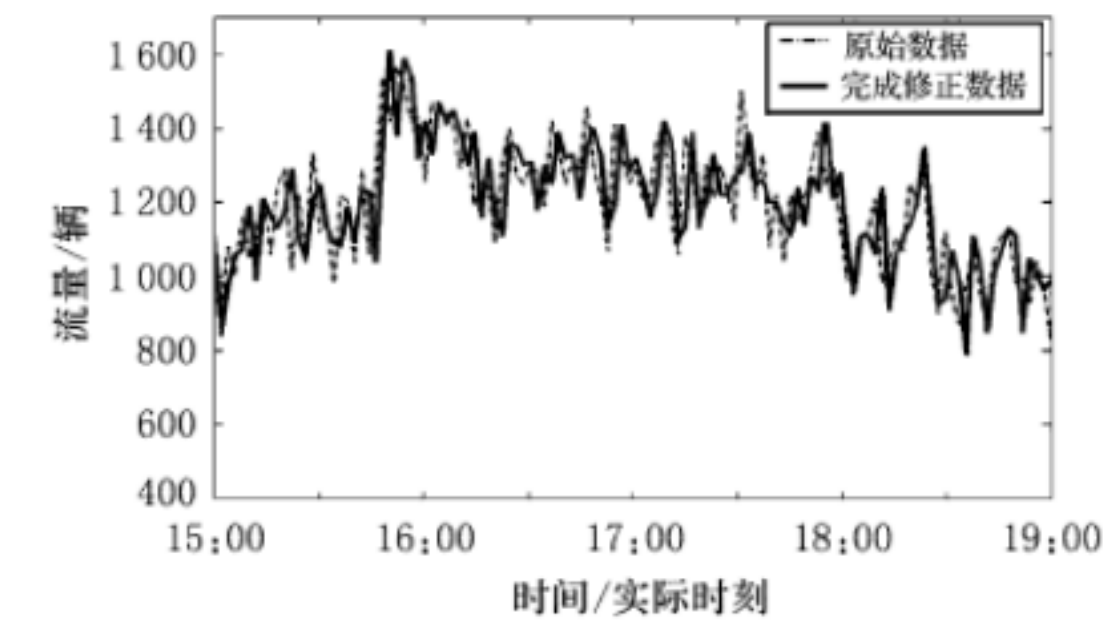


图4 15:00~19:00 原始数据与修正数据对比

3 结束语

本文针对数据缺失、数据无效和数据时间点问题这3类常见问题,提出了一种交通流数据检验与修复方法。最后以北京市交通流数据为例,对该方法进行了分析与验证。结果表明此方法具有很好的检验效果。检验后的数据不但补充修正了缺失、无效数据,还具有统一的标准记录格式,这样的数据可以很方便的直接应用到各种实际研究工作当中。

[1] 裴玉龙,王 富,李宏萍 现代检测技术在城市快速路交通流特性研究中的应用 ITS 通讯, 2004(3): 17-22

[2] 韩卫国,王劲峰,胡建军 交通流量数据缺失值的插补方法 交通与计算机, 2005(1): 39-42

[3] 赵 慧,于 雷,耿彦斌 国内外实时交通流数据质量控制比较与分析 ITS 通讯, 2005(2): 15-19

Method to Verify and Repair Road Traffic Flow Data

JIANG Rui WANG Jun

(Research Institute of Highway of Ministry of Communications, Beijing 100088, China)
(Beijing Jiaotong University, Beijing 100044, China)

Abstract: There are three kinds of common problems in traffic flow data: data loss, data invalidity and data time problem. To solve these problems, this paper has proposed a method of examining and repairing traffic flow data. This method has been analyzed and proved with the traffic flow data in Beijing. The result indicates that this method has splendid ability of practical application. Verifying quality of traffic flow data is to ensure the correctness and completeness of the data obtained from the data sources, providing reliable data for the data management and application.

Key words: traffic flow data; traffic flow theory; data verifying; city ring road