

基于手机与流量数据的 浙江省高速公路拥堵预测研究

王久辉

(浙江省交通信息中心 浙江 310000)

摘 要 本文以浙江省高速公路网为对象,基于手机上高速公路车速数据和收费站出入口流量数据开展浙江省高速公路拥堵预测研究。对收费站流量数据和手机车速数据质量进行分析,在数据质量可靠基础上,构建基于手机与流量数据的浙江省高速公路拥堵预测模型,并通过历史数据算例验证了该模型的适用性。研究成果可为有效解决浙江省高速公路拥堵问题提供有益参考。

关键词 拥堵预测 高速路网 手机车速数据 收费站流量数据 通行能力

Research on traffic congestion prediction for Zhejiang Expressway Based on mobile phone and traffic data

Jiuhui Wang

(Traffic Information Center of Zhejiang Province, Zhejiang 310000, China)

Abstract In this paper, the expressway network of Zhejiang province is taken as an object and the traffic congestion prediction for Zhejiang Expressway based on mobile phone speed data and toll - gate traffic flow data were investigated. At first, the reliability of toll - gate traffic data and mobile phone speed data were analyzed. Then the prediction algorithm model of traffic congestion for Zhejiang Expressway based on mobile phone and traffic data was studied based on the reliable data quality. With the historical data, the feasibility and effectiveness of the algorithm model was verified. Research results can provide effective reference for solving the expressway congestion problem in Zhejiang province.

Key words congestion prediction; expressway network; Mobile phone speed data; toll - gate traffic flow data; traffic capacity

伴随着经济高速增长和城市规模扩大,浙江省汽车保有量越来越多,高速路网通行能力和交通需求之间的矛盾日益突出。尤其自2012年国庆节开始,浙江省所有收费公路实施在重大节假日免收小型客车通行费,导致在高速公路节点上出现了长时间、大面积拥堵现象,无疑给高速公路管理部门带来巨大压力。

随着当今移动互联网及智能手机终端的普及与技术发展,手机移动网络基本实现城市空间区域的全覆盖。2013年国庆前,浙江省交通运输厅建设了全国第一个省级高速公路路网的手机交通信息采集

系统。截至2015年6月,浙江省交通运输厅利用浙江移动和浙江联通的手机大数据,已覆盖了六条高速公路(包括杭州绕城高速、杭金衢高速、沪杭甬高速、甬台温高速、杭宁高速、杭千高速),共计完成了1260km路网的实时交通状态信息采集、处理、监测等工作,并在实际运行中取得了良好效果。

目前,浙江省交通运输厅除了拥有基于手机大数据的高速公路交通信息,还拥有其他类型的交通大数据,例如收费站数据、交调站数据和视频数据。这些数据虽然能够实时反映高速公路的运行情况,但是对交通管理部门而言依然缺乏交通拥堵预测信

息,无法及时疏导车流,避免交通拥堵的发生。因此,如何利用大数据挖掘技术,以当前手机大数据的交通信息采集系统为基础,引入现有的收费站数据或交调站数据,开展浙江省高速公路交通拥堵的预测预警研究,成为当下交通大数据挖掘的重要发展方向。

1 原始数据质量及特征分析

为实现基于手机和收费数据的高速公路拥堵预测研究,首先需要对收费数据和手机数据的质量进行分析,同时考虑到高速公路收费免费通行政策的影响,数据样本需要包含免费通行日期,因此本次研

究以浙江省收费站信息数据、2015 年 2 月的出入口收费数据为样本(其中 2 月 18 日至 2 月 24 日为免费通行日),结合基于手机信令对应日期的车速数据,对数据质量进行检查和分析,结果表明基于这两类数据进行拥堵分析的数据质量可靠。

1.1 收费站流量数据质量分析

对浙江省高速公路路网中有效的 343 个收费站流量数据进行了质量分析,质量分析维度包括:收费口流量数据日变趋势、收费口流量数据 5min 时变趋势、单个收费口流量数据 5min 时变趋势。

1.1.1 收费口流量数据日变趋势分析

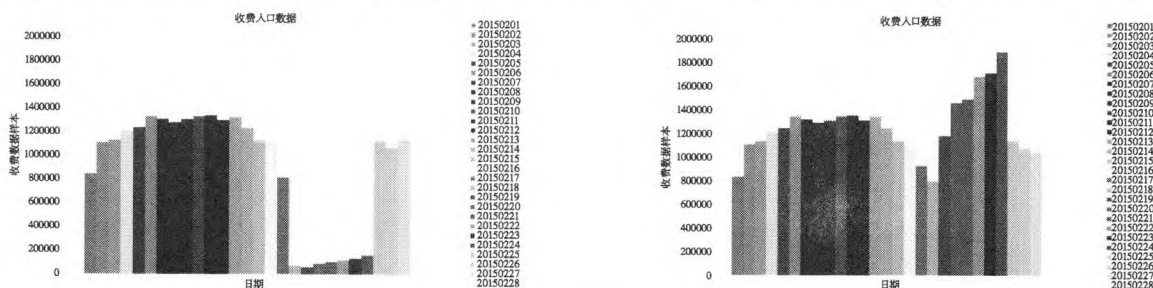


图1 2015 年 2 月入口流量日统计及出口流量日统计

收费入口流量分析:查看各日期数据量分布,除了春节放假期间(02.18-02.24)数据异常减少外,其他日期数据无明显异常情况。

收费出口流量分析:查看各日期数据量分布,在春节期间数据有显著的增长。与春节期间入口数据明显减少相比,主要原因是在春节免收小型客车通行费期间,为提高通行速度在收费 IC 卡上不对小型

车辆、通行时间等信息数据进行录入。

1.1.2 收费口流量数据 5min 时变趋势分析

以浙江高速公路全网作为对象,选取 2 月 3 日(周三)和 2 月 6 日(周六),以 5min 为时间间隔统计收费站的流量,观察工作日与周末收费站流水数据进口与出口的趋势。

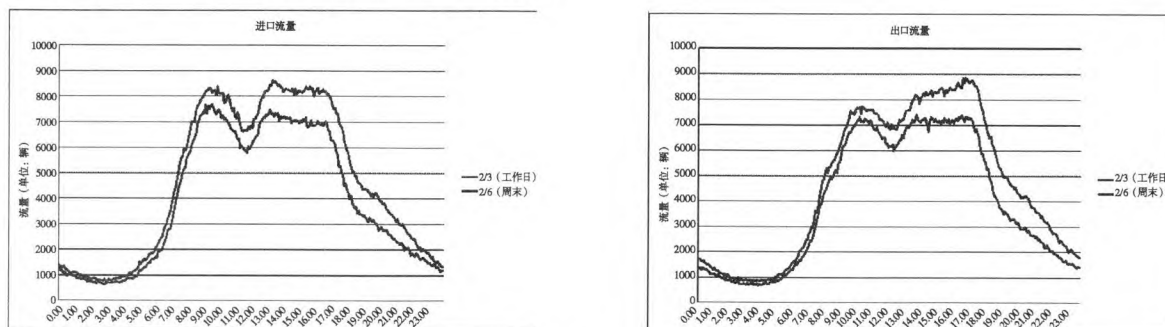


图2 2 月 3 日和 2 月 6 日入口流量 5min 时变统计及出口流量 5min 时变统计

由上图可以看出,收费站进口与出口流量趋势一致,夜间较低,白天较高,中午 12:00 有一个下降趋势;工作日与周末相比,周末的流量较高。

1.1.3 单个收费口流量数据 5min 时变趋势

以王江泾主线收费站(ID 为 1921)为例,分析

单个收费站流量的时间变化规律。

由上图看出,单个收费站与全网的流量增长趋势基本一致,夜间车辆较少,白天车辆较多;周末与工作日相比流量较高。

1.2 基于手机信令的车速数据质量分析

基于移动通信基站信号数据采集道路交通状态特征信息技术是目前新兴的移动采集手段。浙江省交通运输厅于 2013 年建设了基于移动通信基站大数据分

析技术的高速公路交通状态检测系统。本次研究的基于手机信令的车速数据有效覆盖了浙江省高速公路路网单向里程 1263km,覆盖范围参如图 4。

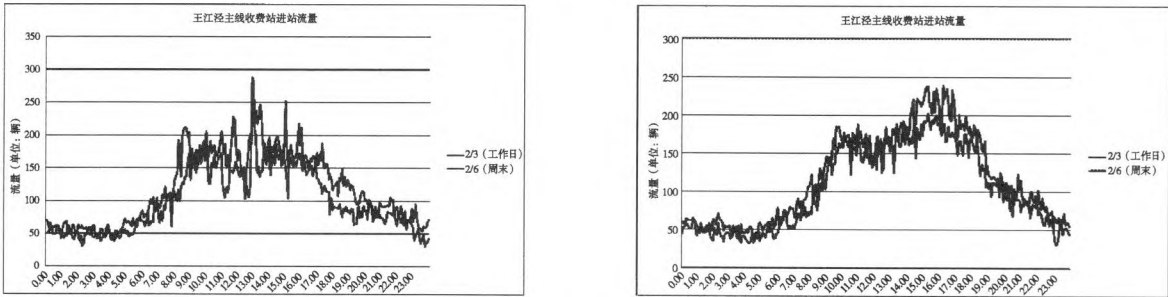


图 3 王江泾主线收费站入口流量 5min 时变统计及出口流量 5min 时变统计

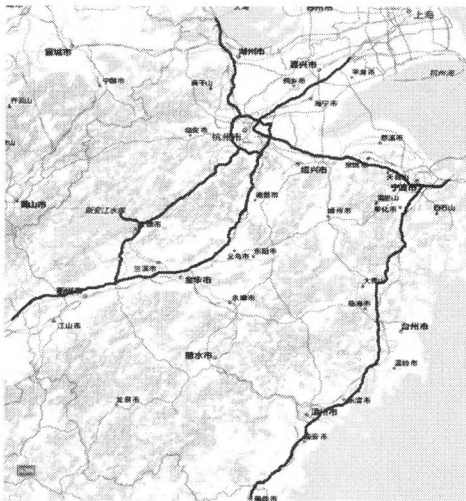


图 4 浙江省基于手机信令的车速数据覆盖路段图

随机选取 2015 年 02 月 08 日至 2015 年 02 月 10 日 3 天的车速数据作为样本,绘制路段杭州绕城高速—绕城西(路段编号为:link_id 为 3301000193)

的速度曲线图(图 5)。由图中数据可知,该 3 天速度曲线的趋势基本一样,夜间速度较高,白天速度较低甚至出现拥堵,符合实际情况。

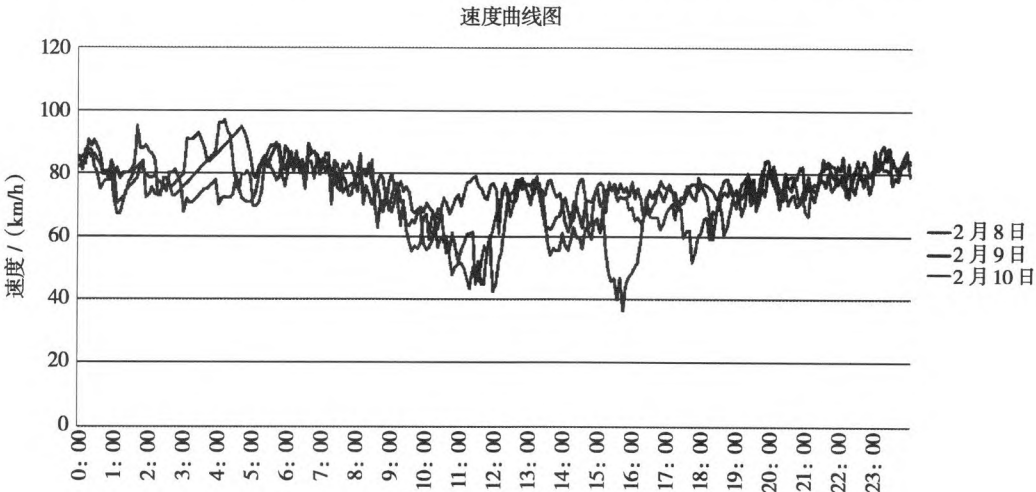


图 5 基于手机信令数据的车速时序分布规律示意

上述两种数据质量分析的结果表明,数据完整性较好、收费站信息相对齐全,车速数据与实际高速公路运营规律一致,表明基于这两类数据进行拥堵分析的数据质量可靠。

2 拥堵预测算法与参数标定

2.1 拥堵预测关键技术及算法

本研究采用的算法主要思路为:通过收费站进口数据短期预测收费站瓶颈断面的流量,若未来可能超过通行能力,则有很大概率发生拥堵,研究的主体流程如图 6 所示。

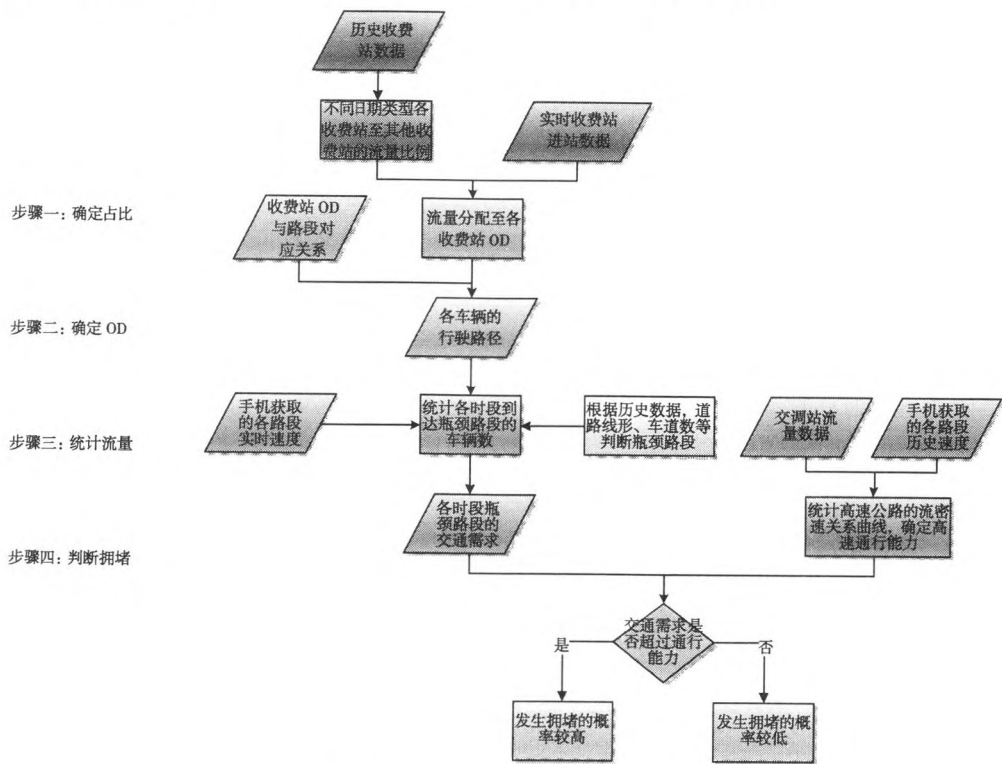


图 6 基于收费站和车速数据的拥堵路段预测方法流程图

2.2 经过常发拥堵路段的 OD 矩阵确定方法

本次采用的流量预测方法主要分为两步骤:首先是获得各收费站 OD 矩阵的流量,其次是根据路段路阻函数确定该 OD 所经过的路径。

计算模型如下:

步骤一,根据收费数据表,计算各收费站 OD 矩阵:

所计算的 OD 矩阵应当满足条件:

$\sum_{j=1}^n q_{ij} = P_i, \sum_{i=1}^n q_{ij} = A_j$
 q_{ij} ——据收费 i 至收费站 j 区间的交通量;
 P_i ——收费站 i 入口交通量;
 A_j ——收费站 j 出口交通量;

步骤二,计算各 OD 对 q_{ij} 所有可能的 m 条路径的路阻系数 $R_l(l=1,2,3,\dots,m)$;

$$R_l = \sum_{k=1}^n Length_k$$

其中 K 为路径 l 中的所有 $link$, R_l 为路径总长度;

步骤三,根据最小路阻系数,确定最优路径 R_s ;

$$R_s = \min(R_l = \sum_{j=1}^n Length_k)$$

当各 OD 对所对应的唯一最优路径确定后,可获得浙江省高速公路网收费站 OD 矩阵和路径表。

2.3 各入口收费站通过常发拥堵路段的占比

在进行收费站各入口进入某路段的流量占比计算前,首先需要获得浙江省高速公路各入口及各出口间的 OD 路径表。浙江省高速路网有效收费站数量为 343,因此理论 OD 对数量应当为 $343 * 343$ 。2015 年 2 月收费数据计算所得的 OD 对及流量数据,如表 1 所示。

通过常发拥堵路段的流量占比计算过程如下:

步骤一:首先获得入口收费站 i 某时间范围内的入口总流量 Q_{all} ;

步骤二:计算通过拥堵路段的入口收费站为 i 的 OD 对,并计算相同时间范围内的流量 Q_i ;

$$P = \frac{Q_i}{Q_{all}}$$

步骤三:流量占比可按照下面的公式进行计算:

OD 对及 2015 年 2 月的流量数据 表 1

流量		SFZ_D 收费站终点					
		1011	1013	1014	1015	4219
SFZ_O 收 费 站 起 点	1011	0	0	979	1050	0
	1013	967	0	0	0	0
	1014	0	1	0	111	0
	1015	1023	99	0	3	0

	4219	0	0	0	0		0

2.4 常发拥堵路段的流量预测和拥堵判断方法

当获得收费站 OD 矩阵和路径后,可结合路段车速,对目标路段断面进行流量预测,具体步骤为:

步骤一:根据各 OD 路径,确定通过目标路段 j 的路径集合为 Q ;假设通过路段 j 的第 i 条路径 Q_i 所包含的路段为 $L_{ik} (k = 1, 2, 3, \dots, j)$,其中 L_{ij} 为目标路段所在的位置;根据路段 L_{ik} 长度 $Length_{ik}$ 和 L_{ik} 所对应的发布段车速 v_{ik} ,计算通过该路段所需要的时间 t_{ik} ;求和 t_{ik} 获得路径 Q_i 的车辆从收费站驶入通过路段 j 的时间总 t_{ij} ;

$$t_{ik} = Length_{ik} / v_{ik}$$

$$t_{ij} = \sum_{k=1}^j t_{ik}$$

步骤二:通过车辆经过入口站的时刻和步骤一计算所得的时间,得到车辆行驶到检测断面的时刻;

步骤三:统计单位时间通过该检测断面的车辆

数。将所有车辆行驶至检测断面的时刻信息存储后,根据需要,统计单位时间内车流量,当流量超过路段通行能力时,判断为拥堵。

2.5 常发拥堵路段的通行能力标定方法

常发拥堵路段的通行能力标定是判断拥堵与否的关键,本研究提出的方法是结合历史数据,对常发拥堵路段的流量和车速之间的关联性进行分析,一般认为,当交通流量超过了通行能力时,会出现显著的通行车速的下降,因此,通过观测车速变化的趋势,来判断通行能力与流量间的关联性,进行标定。

3 拥堵预测方法实施效果评价

基于手机的车速统计和调研,确定了浙江高速公路中的典型常发拥堵路段如表 2 所示。可见,拥堵主要集中在杭金衢高速、杭宁高速及杭州绕城高速上。

浙江省常发拥堵高速公路及路段信息 表 2

序号	路段编号	高速名称	方向	拥堵路段位置
1	3301000057	杭金衢高速	SB	杭金衢和绕城交接段
2	3301000422	杭金衢高速	NB	杭金衢衢州段
3	3301000577	杭金衢高速	SB	杭金衢衢州段
4	3301001341	杭宁高速	SB	杭宁湖州段
5	3301001355	杭宁高速	SB	杭宁德清段
6	3301000040	杭州绕城高速	WX	绕城西
7	3301000081	杭州绕城高速	WX	杭金衢和绕城交接段
8	3301000104	杭州绕城高速	WX	杭金衢和绕城交接段
9	3301000192	杭州绕城高速	NX	绕城西
10	3301000196	杭州绕城高速	WX	绕城西

3.1 常发拥堵路段 - 收费入口流量占比分析

针对典型常发拥堵路段,结合前述拥堵预测模型,确定了各拥堵路段关联的收费入口流量占比。

表3 列出各常发拥堵路段部分主要关联收费站占比。

各常发拥堵路段主要关联收费站占比 表3

序号	路段编号	入口编号	收费站名称	占比	排名
1	3301000057	1011	浙沪主线收费站	0.231846585	1
	3301000057	1921	王江泾主线收费站	0.088572466	2
	3301000057	1031	萧山机场收费站	0.087973936	3
2	3301000422	1153	窑上收费站	0.478424464	1
	3301000422	1143	沈家收费站	0.136789098	2
	3301000422	1149	常山收费站	0.098211733	3
3	3301000577	1011	浙沪主线收费站	0.073699202	1
	3301000577	1145	衢州收费站	0.052315283	2
	3301000577	1921	王江泾主线收费站	0.048249983	3
4	3301001341	1327	父子岭(江浙省界)收费站	0.68478031	1
	3301001341	2125	浙皖主线收费站	0.091008887	2
	3301001341	1323	长兴收费站	0.078905835	3
5	3301001355	1327	父子岭(江浙省界)收费站	0.529412251	1
	3301001355	1315	德清收费站	0.166089307	2
	3301001355	1319	湖州收费站	0.102946499	3
6	3301000040	2620	三墩收费站	0.208766302	1
	3301000040	1327	父子岭(江浙省界)收费站	0.093762776	2
	3301000040	1921	王江泾主线收费站	0.087905621	3
7	3301000081	3111	杭州北收费站	0.122354754	1
	3301000081	2011	昱岭关收费站	0.067685358	2
	3301000081	2637	义桥收费站	0.066206138	3
8	3301000104	2637	义桥收费站	0.150929347	1
	3301000104	1327	父子岭(江浙省界)收费站	0.076051385	2
	3301000104	3111	杭州北收费站	0.069152128	3
9	3301000192	2611	转塘收费站	0.182086616	1
	3301000192	1153	窑上收费站	0.08569057	2
	3301000192	2613	龙坞收费站	0.077258129	3
10	3301000196	2616	留下付收费站	0.15676203	1
	3301000196	2620	三墩收费站	0.114303424	2
	3301000196	1327	父子岭(江浙省界)收费站	0.094445884	3

3.2 常发拥堵路段—流量车速关联分析

针对常发拥堵路段,根据流量预测模型,构建2015年2月份每天的流量时序模型,并与对应的车速统计结果进行比较,研究流量预测模型的准确性和有效性,根据流量—速度变化曲线的规律,进一步标定拥堵路段的通行能力。

各拥堵路段的统计曲线主要有4条,分别为:

根据各入口流量预测28天的拥堵路段流量时序分布曲线图;

根据手机信令数据估计的28天的拥堵路段车速时序分布曲线图;

根据预测的28天拥堵路段流量,绘制流量时序箱图;

根据估计的28天拥堵路段车速,绘制车速时序箱图。

针对每个拥堵路段的流量、车速曲线,分析流量的时空特性和稳定性;构建流量与车速的关联性。一般认为高速公路平均车速低于80km/h可认为进

入拥堵状态。因此,在本次分析过程中,以该车速为阈值标定道路的通行能力。

现以杭州绕城高速—绕城西—3301000040常发拥堵路段为例,2015年2月各日预测流量分时折线图如图7(a)所示,各日分时车速如图7(b)所示,基于各日流量和车速绘制的流量箱图和车速箱图分别如图7(c)以及7(d)所示。

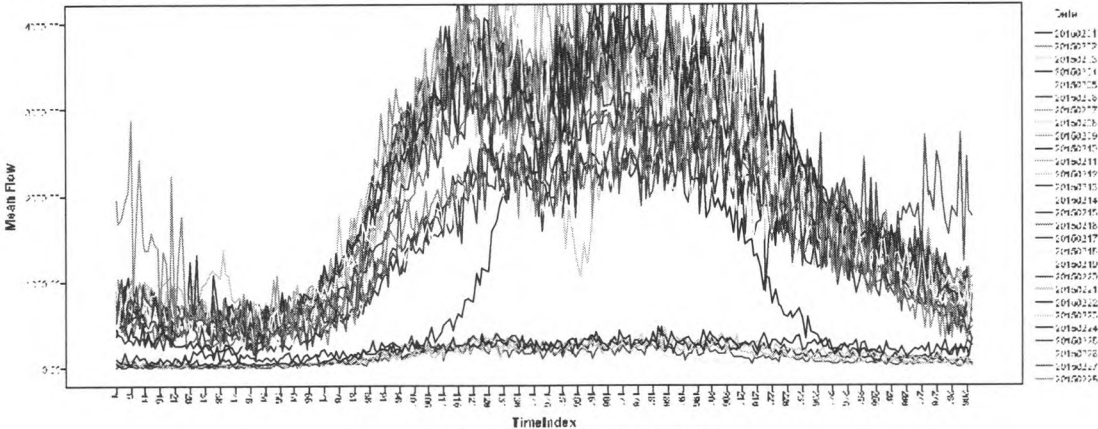
研究结果表明:

该路段流量峰值总体出现在10:30am左右,峰值流量在3800pcu/h;

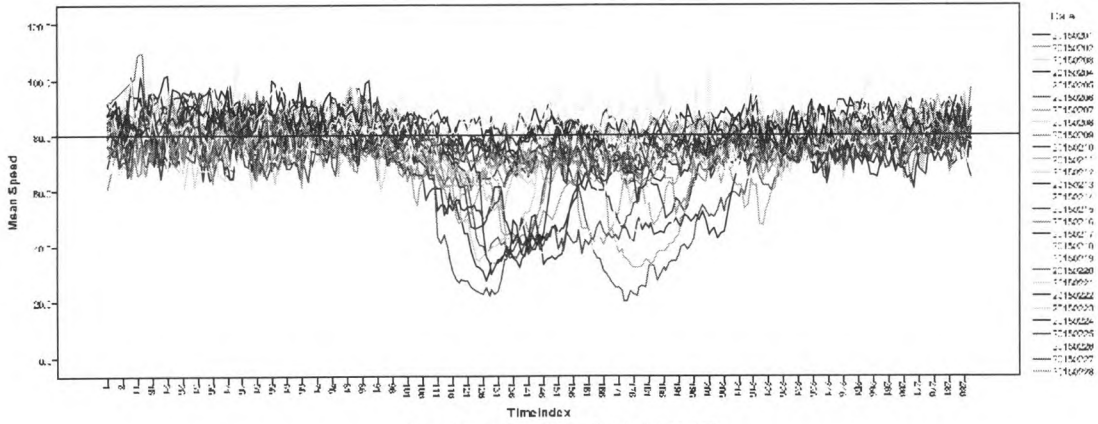
除去节假日数据外,其他日期未见流量出现异常;

平均车速随着流量的增加,呈现总体的显著下降,最低下降至50km/h以下,表明该路段发生常发性严重拥堵;

该路段出现的几次车速有规律显著下降,开始发生在8:45min左右;对应的流量值为3000pcu/h。



(a) 2015年2月流量折线图



(b) 2015年2月车速折线图

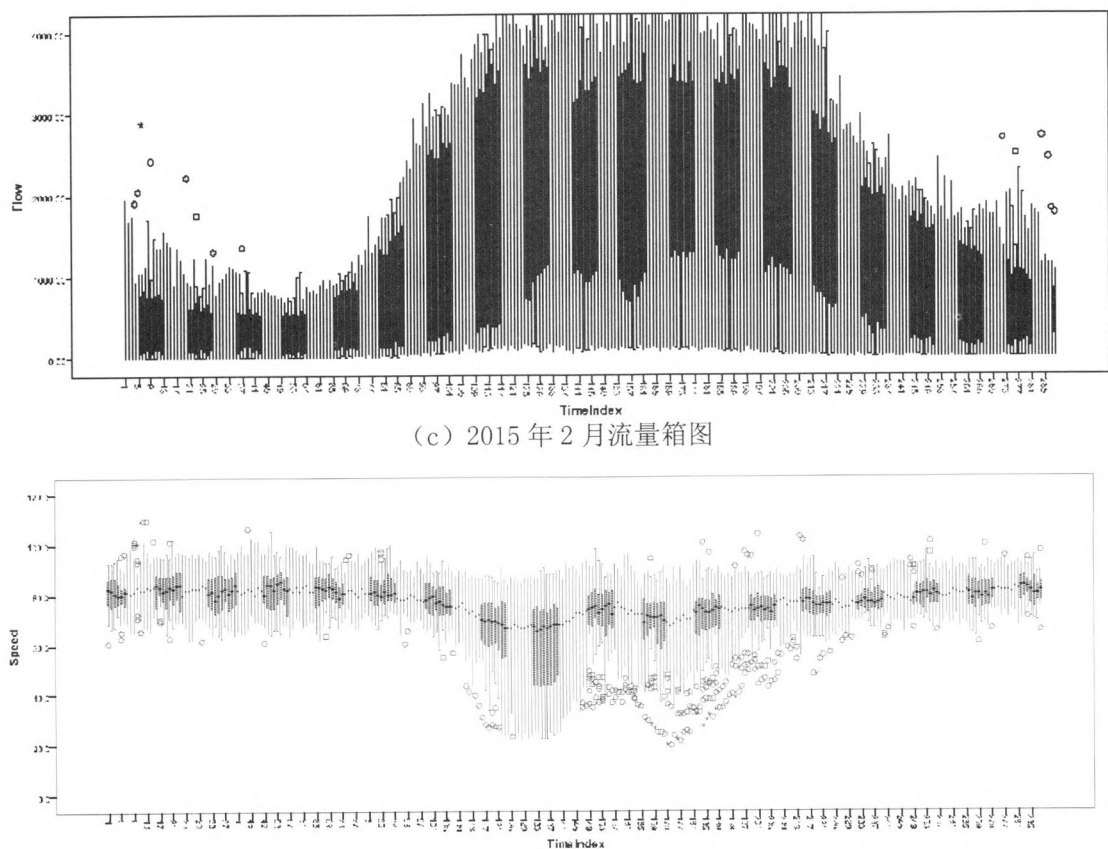


图7 杭州绕城高速——绕城西—3301000040 流量-车速图

3.3 常发拥堵路段—收费入口流量阈值分析

根据上节分析可知,杭州绕城高速—绕城西—3301000040 路段出现了典型的常发拥堵,因此,根据确定的路段通行能力,确定各入口的阈值,计算过程如下:

步骤一:根据标定的通行能力以及各主要入口的占比,确定各主要入口的分配流量阈值。案例计

算过程如表4-3所示。本路段对应的通行能力为3000pcu/h;计算得出各入口分配至该路段的流量阈值分别为:以三墩收费站为例,由于该收费站驶来经过该路段的车辆,占该路段总流量的20.87%,因此计算可知一旦该收费站经过该路段的流量超过 $3000 \times 20.87\% = 626(\text{pcu/h})$,则会发生拥堵。

分配流量阈值求解过程 表4

通行能力	入口编号	收费站名称	路段占比系数	分配流量阈值
3000	2620	三墩收费站	0.2088	626
3000	1327	父子岭(江浙省界)收费站	0.0938	281
3000	1921	王江泾主线收费站	0.0879	264
3000	2622	勾庄收费站	0.0593	178
3000	3111	杭州北收费站	0.0580	174
3000	2624	半山收费站	0.0514	154
3000	1311	南庄兜收费站	0.0380	114
3000	2125	浙皖主线收费站	0.0354	106
3000	1011	浙沪主线收费站	0.0314	94
3000	2632	下沙收费站	0.0297	89

步骤二:根据各入口分配至该路段的占比系数,确定出各入口的流量阈值;

根据第三章的公式计算可知,各入口经过该路段的流量占比如表 4-4 所示,根据占比,可通过分

配流量阈值/入口流经流量占比得到入口流量 1h 流量阈值,如表格 5 最后一列。以三墩收费站为例,计算过程为 $626/0.7460=840(\text{pcu/h})$

入口流量阈值折算过程 表 5

分配流量阈值	入口编号	收费站名称	入口占比系数	入口流量阈值(1h)
626	2620	三墩收费站	0.7460	840
281	1327	父子岭(江浙省界)收费站	0.2022	1391
264	1921	王江泾主线收费站	0.1191	2214
178	2622	勾庄收费站	0.3414	521
174	3111	杭州北收费站	0.2796	622
154	2624	半山收费站	0.2886	534
114	1311	南庄兜收费站	0.0785	1452
106	2125	浙皖主线收费站	0.0933	1139
94	1011	浙沪主线收费站	0.0429	2195
89	2632	下沙收费站	0.0958	931

运用此方法对所有的入口和常发拥堵路段之间的阈值关联起来,标定出面向各常发拥堵路段的入口收费流量阈值。

3.4 常发拥堵路段 – 基于收费入口流量的拥堵预测有效性评价

分布、与基于各入口预测的流量分布以及与车速间的关联性。选取入口 2620,1327,1921 以及 2622 这四个主要入口的分时流量数据,预测拥堵路段的流量,并基于全入口的数据进行比较,结果如图 8 所示。

针对常发拥堵路段,分析 2 月 10 日的总体流量

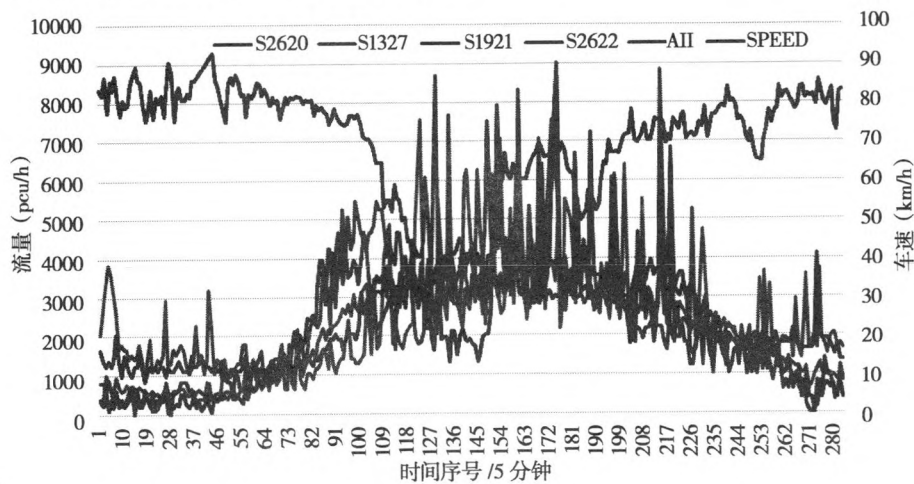


图 8 基于各入口数据的拥堵路段分时流量估计结果

由图 8 可知,基于各入口的流量占比,计算得到 的预测总体流量时间分布 (下转第 59 页)

程中产生的场坪区边缘角点设计标高便于施工阶段的标高控制。为保证计算精度,在使用该方法时,应注意结合公路工程特点和地形特征,采取相应措施加以控制。

参 考 文 献

- [1] 李鹏,张瑞森.南方CASS在道路工程断面法土方计算中的应用[J].城市建设理论研究:电子版,2015,5(12):28-29.
- [2] 刘美娥,乐晶,夏昆,等.土地整治项目土方工程量计算方法研究[J].广东土地科学,2011(5):33-39.

(上接第42页)与实际回归得到的总体流量分布,总体趋势一致,同时预测流量峰值出现的时刻,对应的车速出现下降,说明基于收费入口的常发拥堵路段流量预测结果有效。

4 结 论

本文以收费数据和手机车速数据相结合的方式开展研究,以浙江省高速公路为对象,在确保两种数据质量可靠的基础上,利用最短路径方法,确定了各收费入、出口间的OD路径;并结合收费数据,确定了各入口经过各路段的流量分配百分比,并根据各入口距常发拥堵点的距离计算车辆达到时间,最终预测基于相关收费入口流量的常发拥堵流量。

本文提出的算法能够作为浙江省大范围推广、应用拥堵预测系统的技术基础和支撑。具体在高速公路拥堵预测判断和流量预警过程中,可将所标定的通行能力,根据各入口与常发拥堵路段间的两个关联系数:各入口占拥堵路段流量比以及各入口分配至拥堵路段流量占比系数折算出各收费站面向各常发拥堵路段的流量阈值,构建各收费入口与常发拥堵路段的关联性,当入口流量值超过常发拥堵路段通行能力折算至本入口流量阈值时,触发常发拥堵路段的拥堵预警。

- [3] 何晗芝.GPS-RTK、全站仪及CASS7.0在平场中的应用分析—以两江新区水土高新技术产业园6标平场为例[D].湖南:湖南大学,2012.
- [4] 张海印.土方计算精度与方格网边长之关系的研究[J].东华理工大学学报(自然科学版),2000,23(1):70-73.
- [5] 邓铁军,何晗芝.误差传播定律在土方计算偏差分析中的研究[J].铁道科学与工程学报,2009,6(1):68-73.

参 考 文 献

- [1] LINDLEY J A. Quantification of urban freeway congestion and analysis of remedial measures [R]. Federal Highway Administration, Washington, DC, 1986.
- [2] 姜桂艳,温慧敏,杨兆升.高速公路交通事件自动检测系统与算法设计[J].交通运输工程学报,2001,(1):77-81.
- [3] 徐云飞.高速公路拥堵事件检测中的背景建模及状态判别方法研究[D].重庆大学,2014.
- [4] Haritaoglu I, Harwood D, Davis LS. Hydra: multiple people detection and tracking using silhouettes [C]. Second IEEE Workshop on Visual Surveillance, 1999, 210-217.
- [5] 杨兆升,杨庆芳,冯金巧.基于模糊综合推理的道路交通事件识别算法[J].公路交通科技,2003,20(4):92-94.
- [6] 吕梦蛟.基于移动通信基站大数据的高速公路交通状态采集研究与应用[J].公路,2016,(08):157-163.