

基于公共交通卡数据的城市轨道交通大事件影响分析模型^{*}

颜彦文¹ 刘 婵² 江 渝³ 徐文洁⁴

(1. 上海久誉软件系统有限公司, 201103, 上海; 2. 复旦大学数学学院, 200433, 上海;

3. 上海财经大学数学学院, 200433, 上海; 4. 上海申通地铁集团有限公司, 201103, 上海//第一作者, 工程师)

摘 要 利用基于公共交通卡数据的城市轨道交通线网客流数据进行反演的模型与算法, 对大事件发生前后线网各站点的客流存量进行了分析和重构。反演得到的客流数据可以客观反映大事件的发生随时间变化对发生站点和邻近站点的影响。在此基础上, 定义了城市轨道交通线网每个站点的大事件影响因子。该影响因子可以比较客观地反映大事件对各个站点的影响程度。实证分析证明了基于公共交通卡数据的城市轨道交通大事件影响分析模型的可靠性和有效性。

关键词 城市轨道交通; 公共交通卡数据; 客流预测; 大事件影响

中图分类号 U293.13; F530.7

DOI: 10.16037/j.1007-869x.2020.07.022

Analysis Model of Urban Rail Transit Important Event Impact Based on Public Transportation Card Data

YAN Yanwen, LIU Chan, JIANG Yu, XU Wenjie

Abstract Model and algorithm deduced from urban rail line network passenger flow data based on public transportation card are adopted to analyze and reconstruct the passenger flow stock at each station before and after important events. The retrieved passenger flow data can reflect objectively how the impact of important event on the station where it occurred and the neighboring stations evolved over time. On this basis, the impact factors of important event at each station in urban rail transit line network are defined, which reflect objectively at what level the important event is influencing each station. The empirical analysis has shown the reliability and effectiveness of the impact analysis model of urban rail transit important event based on public transportation card data.

Key words urban rail transit; public transportation card data; passenger flow prediction; important event analysis

First-author's address Shanghai Jiuyu Software System Co., Ltd., 201103, Shanghai, China

近年来,我国各主要城市的轨道交通线网规模日益庞大,结构日益复杂,线网安全和运营管理面临极大挑战。特别是遇有重要节假日或大型活动等大事件发生时,在特定的车站很容易发生不同程度的大规模客流突发聚集或消散,从而对乘客出行或线网运营产生很大的影响。利用进出站大数据,实时掌握大事件发生前后城市轨道交通线网中各个站点的客流存量和各条线路上客流负载信息,做好大事件影响分析,及时准确对线网超负载的可能位置和时间作出预警,进而通过限流、调整行车间隔等手段有效控制、引导客流,有效防控事故风险,提高乘客乘车舒适度,是城市轨道交通线网精细化管理中最重要的课题。

国内外现有的大事件对城市轨道交通线网影响的分析判断,主要基于经验和感觉。但主观感觉判断的巨大差异性,使得客流分布估计与实际情况相差甚远。解决这一问题的关键在于准确计算各站点的实时客流存量和有效预测客流去向^[1-9]。

文献[10]提出了一种利用乘客进出任意两个站点的公共交通卡记录数据,将路网客流量分配到各条最短路径上,从而得到各个站点客流存量的反演模型和算法。本文利用该算法,对大事件发生前后线网各站点一天中每个时刻的客流存量进行分析和重构,反演得到的客流数据,以客观反映大事件的发生随时间变化对发生站点和邻近站点的影响。在此基础上,定义并计算了线网每个站点的大事件影响因子,可以用于对大事件影响的定量分析,并作为“城市轨道交通运营故障影响分析系

^{*} 国家自然科学基金面上项目(11771270)

统^[10]”的一个有机组成部分。

1 基于交通卡数据的客流分析和反演

1.1 客流存量函数定义

对于每个站点 $S_i (i=1, \dots, M)$, 站点 S_i 通往相邻站点 S_j 的线路为 $e_{i,j} (i \neq j, j=1, \dots, M)$ 。定义以下3类客流量函数: ①从站点 S_i 到达站点 S_j 的客流量函数为 $f_{i,j}=f(e_{i,j})$; ②进入站点 S_i 的进站客流量函数为 I_i ; ③离开站点 S_i 的出站客流量函数为 O_i 。显然, 客流量函数 $f_{i,j}$ (或 I_i, O_i) 与时间相关。

根据客流守恒 (Flow Conservation): 对于每个站点 S_i , t 时刻前后客流量函数满足:

$$I_i(t_{(-)}) + \sum_{e_{j,i} \in E} f_{j,i}(t_{(-)}) = O_i(t_{(+)}) + \sum_{e_{i,j} \in E} f_{i,j}(t_{(+)}) \quad (1)$$

式(1)等号左(或右)边值即为站点 S_i 在 t 时刻的客流存量, 记为 $C_i(t)$ 。式中集合 E 表示两个相邻站点间所有线路的集合。 $f(t_{(-)})$ 和 $f(t_{(+)})$ 表示函数 $f(t)$ 在时刻 t 左右的极限值。

1.2 算法假设

对时间作离散化, 为描述轨道交通网络中客流流动过程中的算法作如下假设:

1) 城市轨道交通线路上所有列车的发车间隔、运行速度均相同, 发车间隔记为 Δt 。

2) 按发车时刻 $\{t_j\}_{j=1}^N$ 对 $1d$ 进行划分:

$$t_0 < t_1 < \dots < t_N < t_{N+1}, t_{j+1} - t_j \equiv \Delta t \quad (2)$$

其中, 初始时刻记为 t_0 , 结束运营时刻记为 t_{N+1} 。

3) 相邻站点间距离相同, 列车运行时间与发车间隔相同, 均为 Δt 。假设以相邻站点距离中点为界, 即使乘客在列车上, 也等同于在2个站点中的某一个站点上。

4) 忽略列车在车站停留的时间间隔。

5) 在时间段 $[t_{j-1}, t_j)$, 进站的客流必将搭上 t_j 时刻到站列车。

6) 同站换乘时间同发车间隔一样亦为 Δt 。

7) 在时间段 $[t_{j-1}, t_j)$, 从相邻站点坐车抵达的出站乘客, 在 t_j 时刻出站, $[t_0, t_1]$ 内没有出站乘客。

8) 站点间线路客流量的最大承载量为无穷大, 进站出站的客流限制也是无穷大, 即任意两个相邻站点间线路上的容量函数 (即最大客流量函数) 为: $C(e) \equiv \infty, \forall e \in E$ 。

9) 忽略实际运营中的特殊情况, 如坐过站。

10) 乘客倾向于用时少、少换乘的线路。

1.3 公共交通卡数据与客流存量

实际运营中可能获得的原始数据主要是乘客出入站点使用公共交通卡的记录, 每条记录包含的信息仅为: 卡号 (ID)、票价、进出站点 (S_{in}, S_{out}) 相应的进出站点时刻 ($t_{in}(=t_{k1}), t_{out}(=t_{k2})$)。

对单个乘客数据 (ID, $S_{in}, S_{out}, t_{k1}, t_{k2}$), 按照相同出入站点和相同进站时刻 (S_{in}, S_{out}, t_{k1}) 进行分组, 则对于每个离散时刻 $t_k (k=0, \dots, N+1)$, 存在 s_k 组数据:

$$D_k = \begin{pmatrix} \vdots & \vdots & \vdots \\ S_f & S_{in} & S_{out} \\ \vdots & \vdots & \vdots \end{pmatrix}_{s_k \times 3}$$

其中: S_f 指 t_k 时刻在 S_{in} 站点进站乘车, 且在相同终点站 S_{out} 下车的乘客数量; $s_k \leq M \times M$, 即分组数据的组数不超过进出站点组合数的总数。因此矩阵 D_k 的维数是有限的。记时间序列 $D = \{D_k\}_{k=1}^{N+1}$, 称为 OD (起讫点) 数据。

如果途经站点一样, 但出站时刻不同, 假设有 f_r 位乘客选择了路线 Γ :

$$\Gamma = (S_r(t_0), \dots, S_r(t_{k1}), \dots, S_r(t_{k2}), \dots, S_r(t_{n+1}))$$

其中, $S_r(t)$ 是 t 时刻该乘客途经的站点。在进出站时间段 (t_{k1}, t_{k2}) 以外的时刻, 乘客途经站点设为 0, 即 $S_r(t) = 0 (t \in \{t_k\}_{k=k_1}^{k_2})$ 。可以得到基于线路的客流量函数表示为:

$$I_i(t_k) = \sum_{\Gamma} f_r \delta_{S_{in}, \Gamma(t_k), S_i} \quad (3)$$

$$f_{j,i}(t_k) = \sum_{\Gamma} f_r \delta_{S_r(t_{k-1}), S_j} \delta_{S_r(t_k), S_i} \quad (4)$$

$\delta_{x,y} = \{1(x=y), \text{ or } 0(x \neq y)\}$ 是 Kronecker 函数。 Γ' 表示包括所有不同 OD 但都经过 S_i 站点的线路。相应 S_i 站点 t_k 时刻的客流存量为:

$$C_i(t_k) = I_i(t_k) + \sum_j f_{j,i}(t_k) \quad (5)$$

1.4 客流存量分析和反演算法

具体选择路线 Γ 的乘客数量 f_r 无法直接得到, 但是乘客选择路线时, 大概率会选择最短路径。通过基于概率统计的大数据分析技术, 定义乘客对给定起始站 S_{in} 和终点站 S_{out} 间前 K 条最短路线的倾向度 (即线路选择概率) P_l 满足: $0 \leq P_l \leq 1; \sum_l P_l = 1 (l=1, \dots, K)$ 。

这样, 在时刻 t_k 进站、以 S_{in} 和 S_{out} 为进出站点的共 S_f 位乘客中, 可以按出站时间前后分为 K 组, 分别对应前 K 条最短路径。或者可以理解为有 $S_f \times P_l$

位乘客选择了第 l 短路线。

综上分析,得出客流存量分析和反演算法(算法 1)如下:

- 1) 对于每个时刻 t_k ,统计分组数据 D_k ;
- 2) 对于每一条分组数据,即矩阵 D_k 中的每一行,计算站点 S_{in} 与站点 S_{out} 之间的前 K 条最短路径;
- 3) 对 S_j 位乘客按出站时间前后分为 K 组,分别对应前 K 条最短路径,计算前 K 条最短路线的倾向度(即线路选择概率) P_i ;
- 4) 由式(3)、式(4)计算客流量 $I_i(t_k)$ 和 $f_{i,j}(t_k)$;
- 5) 由式(5)计算客流存量 $C_i(t_k)$ 。

在实际应用中,基于 2018 年上半年上海轨道交通的 OD 数据:

- 1) 采用贪心算法计算任两个站点 S_i 与 S_j 之间的前 K 条最短路径(参见文献[14]),路线数 $K=3$ (最短路径算法一般假设 $K \leq 3$,参见文献[11])。
- 2) 基于一个较长时期工作日/节假日的海量大数据,利用 KNN 方法(参见文献[15]),可以得到一个全局性的倾向度(即线路选择概率) $(p_1, p_2, p_3) = (0.7, 0.2, 0.1)$ 。

2 大事件影响分析模型

利用前述算法,可以反演得到大事件发生日与平日(若干个与发生日相同星期几的平均数据)的各地铁站客流存量,建立起大事件影响分析模型,确定因大事件发生而引起的客流量及影响区域。如果某日有大事件发生,对路网中某站点 A ,可进行大事件影响分析。大事件影响分析模型(算法 2)如下:

- 1) 利用算法 1,计算大事件发生日 A 站点每个时刻 t 的客流存量 $C_A(t)$ 。
- 2) 利用算法 1,计算平日 A 站点每个时刻 t 的客流存量 $T_A(t)$ 。
- 3) 计算大事件发生日与平日 A 站点每个时刻 t 的客流存量差: $E_A(t) = C_A(t) - T_A(t)$ 。
- 4) 计算每个 t 时刻的如下指标: $f_A(t) = E_A(t) / C_A(t)$ 。
- 5) 计算站点 A 的大事件影响因子: $f_A = \frac{\text{mean } f_A(t)}{(\forall t, E_A(t) \neq 0)}$ 。
- 6) 利用式(4),计算大事件发生日和平日每个时刻 t 邻近站点 S_j 进入站点 S_i 的客流 $f_{i,j}(t)$ 。

通过该大事件影响分析模型,可得如下结论:

1) 将计算得到的实时客流存量 $S_A(t)$ 、 $T_A(t)$ 和客流存量差 $E_A(t)$ 作可视化展示,能直观地观察到大事件的发生对发生站点和邻近站点客流的影响,以及站点客流量随时间的分布和变化情况。

2) 将计算得到的线网中每个站点的大事件影响因子作可视化展示,能直观地观察到大事件发生的影响区域。

3 实证分析

以 2018 年 3 月 2 日(元宵节)的上海轨道交通的交通卡数据为例,考察“豫园灯会”这一大事件对轨道交通线网的影响。

3.1 大事件发生对发生站点的影响

利用算法 1,反演得到 2018 年元宵节当天与平日豫园站一天的客流存量分布,如图 1 所示。

由图 1 可见,14:00 之前,元宵节与平日的客流存量基本一致,08:00—10:00 时出现了早高峰;由于 3 月 2 日举办“豫园灯会”,从 16:00 起客流存量逐渐增加,至 20:00 达到顶峰,后逐渐减少;3 月 2 日约 20:00 晚高峰的客流存量约为平日晚高峰客流存量的 1.8 倍,站内非常拥挤。

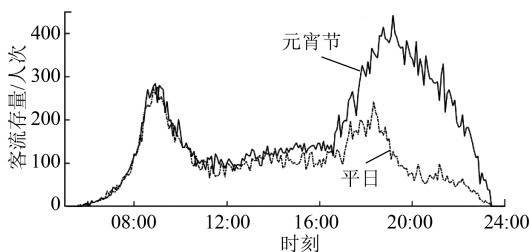


图 1 元宵节与平日豫园站客流存量分布图

3.2 大事件发生对邻近站点的影响

利用算法 2,可以得到豫园站及其附近站点元宵节与平日 10:00 后的客流存量变化差,如图 2 所示。

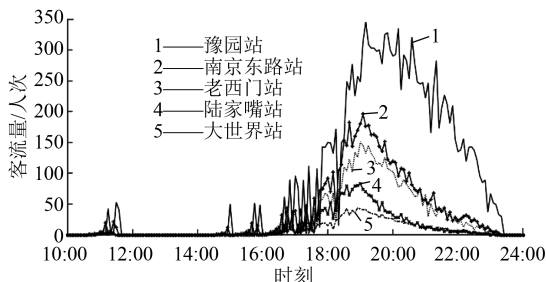


图 2 元宵节与平日豫园站及其附近站点的客流存量变化差

由图2可见,16:00之后,“豫园灯会”这一大事件对豫园站及其附近站点的客流造成了显著的影响。越邻近南京东路站和老西门站,影响越大;客流量从远(大世界站、陆家嘴站)逐渐向近(南京东路站、老西门站)汇集,并最终汇集到豫园站;客流高峰出现时刻也按照从远(大世界站、陆家嘴站,19:00前)到近(南京东路站、老西门站,约19:00)的顺序后移,最终豫园站客流量在接近20:00时达到顶峰。

进一步利用式(3),得到邻近站点 S_j 进入站点 S_i 的客流 $f_{j,i}$,可以考察 S_i 站内客流存量的组成成分,特别是来自各个邻接站点的客流以及进站客流。

图3为南京东路站在元宵节与平日16:00—23:00时,不同客流成分的百分比占比面积图,其中空白部分表示在该站进站的客流百分比。

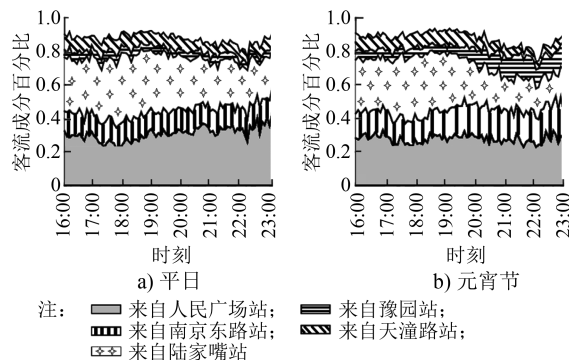


图3 元宵节与平日16:00—23:00南京东路站不同来向的客流成分百分比

元宵节19:00之后来自豫园站的客流占比上升,并在20:30—22:00达到高峰。这充分表明大事件高潮的结束驱动客流从豫园站经由南京东路站逐步分散到线网的其它区域。

3.3 大事件影响因子

利用算法2算出的受“豫园灯会”这一大事件影响的主要区域范围内各站点影响因子,如图4所示。由图4可知:影响因子以豫园站为中心向外扩散减少;距离豫园站较近的老西门站、新天地站和大世界站受到的影响较大;南京东路站虽然直接相邻豫园站,但因为平时客流基数非常大,大事件影响的增量效应并不十分明显。

综上分析可知,大事件影响分析模型可以很好地反映大事件发生对城市轨道交通线网的影响,相关性分析结论和现实相符。同时,模型的量化、可视化结果能进一步为大事件响应提供科学

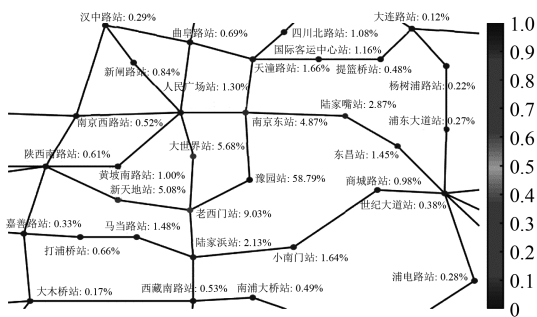


图4 “豫园灯会”大事件影响的主要区域范围内各站点影响因子

可靠的决策依据。

参考文献

- [1] KATO H, KOSUDA K. Elimination-by-aspects in urban rail demand analysis [C]//Sixth International Conference of Traffic and Transportation Studies Congress (ICTTS). Reston, USA: American Society of Civil Engineers, 2008.
- [2] HIBINO N, HYODO T, UCHIYAMA H. A study on characteristics of non-IIA route choice models on high density railway network [J]. Doboku Gakkai Ronbunshu, 2004, 765: 131.
- [3] HIBINO N, UCHIYAMA H, YAMASHITA Y. A study on evaluation of level of railway services in Tokyo metropolitan area based on railway network assignment analysis [J]. Journal of the Eastern Asia Society for Transportation Studies, 2005(6): 342.
- [4] 四兵锋, 毛保华, 刘智丽. 无缝换乘条件下城市轨道交通网络客流分配模型及算法 [J]. 铁道学报, 2007(6): 12.
- [5] 四兵锋, 钟鸣, 高自友. 城市混合交通条件下路段阻抗函数的研究 [J]. 交通运输系统工程与信息, 2008(1): 68.
- [6] 吴建军, 高自友, 孙会君. 城市交通网络上个体选择行为的统计动力学特性研究 [J]. 交通运输系统工程与信息, 2008(2): 69.
- [7] 吴祥云, 刘灿齐. 轨道交通客流量均衡分配模型与算法 [J]. 同济大学学报(自然科学版), 2004(9): 1158.
- [8] 刘剑锋, 孙福亮, 柏赞, 等. 城市轨道交通乘客路径选择模型及算法 [J]. 交通运输系统工程与信息, 2009(2): 81.
- [9] 刘剑锋. 基于IC卡数据的城市轨道交通网络配流模型研究 [J]. 物流技术, 2010(8): 224.
- [10] 吴志强, 黄天印, 颜彦文, 等. 基于大数据的城市轨道交通运营故障影响分析系统对客流影响的分析 [J]. 城市轨道交通研究, 2019(4): 31.
- [11] GOLDBERG A V, TARDOS É, TARJAN R E. Network flow algorithms [Z]. Stanford: Stanford University CS Department, 1989.
- [12] BISHOP C M. Pattern recognition and machine learning (Information Science and Statistics) [M]. New York: Springer-Verlag New York, Inc, 2006.

(收稿日期: 2019-08-06)