

文章编号: 2095-2163(2019)03-0240-04

中图分类号: U491

文献标志码: A

基于统计的轨道交通旅行时间估计方法

陈佳¹, 施海全¹, 孙煜²

(1 上海仪电物联技术股份有限公司, 上海 200233; 2 上海申通地铁集团有限公司技术中心, 上海 201103)

摘要: 本文致力于解决城市轨道交通旅行时间的估算。研究通过对影响换乘路径选择和旅行时间的因素的研究分析, 提出了基于统计的候车时间和换乘时间估计的算法, 并用上海轨道交通的历史数据评价了该算法的可靠性。

关键词: 城市轨道交通; 换乘路径; 旅行时间; K 短路径

A statistics approach to travel time estimation for rail transit

CHEN Jia¹, SHI Haiquan¹, SUN Yu²

(1 Shanghai INESA Network Co., Ltd, Shanghai 200233, China;

2 Technical Center of Shanghai Shentong Metro Group Co., Ltd, Shanghai 201103, China)

[Abstract] This paper addresses the problem of estimating the travel time during urban rail transit. After the analysis of the factors impacting transfer path selection and time cost, the paper proposes a statistics method to estimating the waiting time as well as the transfer time. Based on the above, the paper evaluates the performance of the algorithm on the historical data of Shanghai rail transit, demonstrating its reliability.

[Key words] urban rail transit; transfer path; travel time; K shortest paths

0 引言

轨道交通作为城市公共交通的重要方式之一, 具有运量大、速度快、安全、准时等特点, 已逐渐成为居民出行首选公共交通工具。随着轨道交通路网不断扩大, 换乘站点越来越多, 乘客需跨线乘坐列车的情况日益普遍, 换乘路径选择和旅行时间是乘客普遍关心的问题, 本文研究分析了影响换乘路径选择和旅行时间的因素, 提出一种基于统计的方法来估计路径的旅行时间。

1 影响因素研究

1.1 换乘路径

通常, 乘客出行时并不会考虑从起点到终点之间所有可能的换乘路径, 而仅考虑其中的一部分换乘路径, 称作有效路径。现有清分模型中, 采用 K 短路径法(即求解任意 2 个跨线站点间的 K 条最短路径)来搜索有效路径^[1-4]。主要实现思想如下:

(1) 给定图 $G = (V, E)$, 其中 V 表示所有节点的集合, E 表示所有边的集合, 使用 Dijkstra 算法(或 Floyd 算法)求得任意 2 个节点 v_i, v_j 间的最短路径 $S^{(1)}(v_i, v_j) = \{e_1^{(1)}, \dots, e_{m(1)}^{(1)}\}$, 其中 $e_r^{(1)}$ 是最短路径

中的第 r 条边, 其代价为:

$$d(S^{(1)}(v_i, v_j)) = \sum_{r=1}^{m(1)} d(e_r^{(1)}), \quad (1)$$

(2) 假设当前已经求出了 $k-1$ 短路径 $S^{(1)}(v_i, v_j), \dots, S^{(k-1)}(v_i, v_j)$, 从每一条路径中选出不重复的一条边, 即 $e_{r_l}^{(l)} \in S^{(l)}(v_i, v_j), l=1, \dots, k-1$, 构成了 $k-1$ 条来自不同路径的不同边的配对集 $\bar{E}_{r_1, \dots, r_{k-1}}^{(k)} = \{e_{r_1}^{(1)}, \dots, e_{r_{k-1}}^{(k-1)}\} (e_{r_1}^{(1)} \neq \dots \neq e_{r_{k-1}}^{(k-1)})$ 。从原图 G 中扣除这些边, 构造一个辅助图 $G_{r_1, \dots, r_{k-1}}^{(k)} = (V, E \setminus \bar{E}_{r_1, \dots, r_{k-1}}^{(k)})$, 求出辅助图 $G_{r_1, \dots, r_{k-1}}^{(k)}$ 中 2 个节点 v_i, v_j 间的最短路径 $S_{r_1, \dots, r_{k-1}}^{(k)}$ 。

(3) 遍历所有的 $k-1$ 条来自不同路径的不同边的配对集 $\bar{E}_{r_1, \dots, r_{k-1}}^{(k)}, r_l = 1, \dots, m(l), l = 1, \dots, k-1, e_{r_1}^{(1)} \neq \dots \neq e_{r_{k-1}}^{(k-1)}$, 则第 k 短路径的数学表达如下: $S^{(k)}(v_i, v_j) =$

$$\operatorname{argmin}_{S_{r_1, \dots, r_{k-1}}^{(k)}, r_l = 1, \dots, m(l), l = 1, \dots, k-1, e_{r_1}^{(1)} \neq \dots \neq e_{r_{k-1}}^{(k-1)}} d(S_{r_1, \dots, r_{k-1}}^{(k)}). \quad (2)$$

(4) 重复步骤(2)和(3), 直到 $k=K$ 。

在实际应用中, 一般取 $K \leq 5$ 。

1.2 旅行时间

本文中跨线站点间的旅行时间特指从进闸机到出闸机总共需花费的时间, 其中包括总行车时间

作者简介: 陈佳(1982-), 女, 硕士, 工程师, 主要研究方向: 大数据分析; 施海全(1982-), 男, 工程师, 主要研究方向: 大数据分析; 孙煜(1985-), 男, 学士, 工程师, 主要研究方向: 轨道交通信息化、物联网应用和大数据应用。

收稿日期: 2019-01-25

(含列车在站台停留时间)、总候车时间、总换乘步行时间、从进站闸机到上车点时间和从下车点到出站闸机时间。用 o 表示某一出发站点, d 表示另一目的站点,则有:

$$t_{o,d} = t_o^{(\text{进站})} + t_{o,d}^{(\text{候车})} + t_{o,d}^{(\text{行车})} + t_d^{(\text{出闸})}, \quad (3)$$

需要指出的是,当本文研究讨论各种时间时,此处则特指这种时间开销的期望值。如果其中经过从某一线路站点 i 到另一线路站点 j 的换乘(含换乘步行及换乘候车时间),即有:

$$t_{o,d} = t_o^{(\text{进站})} + t_{o,i}^{(\text{候车})} + t_{o,i}^{(\text{行车})} + t_{i,j}^{(\text{换乘步行})} + t_{j,d}^{(\text{候车})} + t_{j,d}^{(\text{行车})} + t_d^{(\text{出闸})}. \quad (4)$$

研究中基于简洁的表达考虑,在此仅列举一次换乘的情况,但本文中的推导和计算可以简单推广到多次换乘的情况。同时注意到,即便在同一条线路中,不同运行区间的地铁运营时刻也有所不同,因此候车时间不仅依赖于出发站点 o ,也依赖于目的站点 d 或换乘站点 i 。

在正常运营的条件下,所有班次列车分方向通过相邻站点间的行车时间及在站台停留的时间基本一致。不同出发和目的站点闸机到上下车点所需时间 $t_o^{(\text{进站})}$ 和 $t_d^{(\text{出闸})}$ 虽然存在差异,该差异相对于跨线站点间的换乘步行时间 $t_{o,d}^{(\text{换乘})}$ 基本可忽略不计。而总候车时间和总换乘步行时间则可能受车厢拥挤度、步行速度等各方面因素影响,详细分析见 1.3 节。

1.3 影响因素

本文梳理罗列了影响换乘路径选择的各个因素。对其可做阐释描述如下:

- (1) 途经车站数。
- (2) 换乘次数。
- (3) 是否出站换乘:出站换乘会比站内换乘多出再次安检和过闸机时间。
- (4) 是否首发站。
- (5) 末班车时间:末班车时间可能会影响换乘路径的可达性。

以上因素(1)~(3)均与路径的时间开销相关,由此可推,时间开销对乘客的换乘路径选择有重要的影响。而时间开销中除了候车时间和换乘步行时间,其它时间基本没有弹性。

影响花费时间中的总候车时间的因素有:

(1) 行车间隔时间:假设只要列车到站就可以上车,单次候车时间取值范围为 0~发车间隔时间。可见,行车间隔时间越短,可能的最长候车时间越短。

(2) 车厢拥挤度、候车人数和排队位置:这一组客观因素会影响乘客能否乘上车,与因素(1)及乘客是否赶路这一主观因素综合,亦可能影响乘客是否有意愿再等候一辆列车,从而影响等候的列车数。若没能乘上第一辆列车或情愿再等候一辆列车,则将延长候车时间,延长时间等于行车间隔时间乘以多等候的列车数。

(3) 是否首发站:综合因素(1)、(2)及乘客是否赶路可能影响乘客是否有意愿再等候一辆列车,从而影响等候的列车数。若情愿再等候一辆列车,则将延长候车时间,延长时间等于行车间隔时间乘以多等候的列车数。

影响总换乘步行时间的因素主要是乘客步行速度,而乘客步行速度与下述因素相关,可将其整合概述为:

- (1) 乘客自身步行速度。
- (2) 乘客是否赶路。
- (3) 换乘通道拥挤度。

2 旅行时间估算

2.1 同线进出站旅行时间的估算

如 1.2 节所述,同线进出站的旅行时间可表示为:

$$t_{o,d} = t_o^{(\text{进站})} + t_{o,d}^{(\text{候车})} + t_{o,d}^{(\text{行车})} + t_d^{(\text{出闸})}. \quad (5)$$

本文采用上海申通地铁提供的相邻站点间上下行行车时间(含站台停留时间)来得到总行车时间的估算 $\hat{t}_{o,d}^{(\text{行车})}$ 。仍需说明的是,由于上下行的区分,在本文中,所有的实际站点都被区分为带有上下行标记的 2 个不同的逻辑站点。

对于闸机到上下车点所需时间 $t_o^{(\text{进站})}$ 和 $t_d^{(\text{出闸})}$,研究假设进出闸的步行时间为与站点 o 和 d 无关的常量,即:

$$t_o^{(\text{进站})} \approx t^{(\text{进站})}, t_d^{(\text{出闸})} \approx t^{(\text{出闸})}. \quad (6)$$

由 1.3 节可知,影响候车时间的主要因素是行车间隔时间,而从上海申通地铁提供的路网各线行车间隔时分表可得每条线路(区段)的同一方向在一个时间段内的发车间隔时间是一个定值,该定值直接影响了期望候车时间。过程中,由于逻辑站点本身带有上下行的标记,因此一个实际区段也被区分为带有上下行标记的 2 个不同的逻辑区段。用 $R(x)$ 表示站点 x 所属的区段,则针对某一出发站点所在的区段 $R(o)$ 和另一目标站点所在的区段 $R(d)$,有:

$$t_{o,d}^{(\text{候车})} = t_{R(o),R(d)}^{(\text{候车})}, \quad (7)$$

于是,研究推导得到:

$$t_{o,d} = t_o^{(\text{进闸})} + t_{o,d}^{(\text{候车})} + t_{o,d}^{(\text{行车})} + t_d^{(\text{出闸})} \approx t_{o,d}^{(\text{行车})} + [t_{R(o),R(d)}^{(\text{候车})} + t^{(\text{进闸})} + t^{(\text{出闸})}], \quad (8)$$

令:

$$t_{R(o),R(d)}^{(\text{非行车})} = [t_{R(o),R(d)}^{(\text{候车})} + t^{(\text{进闸})} + t^{(\text{出闸})}], \quad (9)$$

表示从某一区段的出发站点到某一区段的的目的站点的非行车时间为:

$$t_{R(o),R(d)}^{(\text{非行车})} = t_{o,d} - t_{o,d}^{(\text{行车})}. \quad (10)$$

这为研究中运用统计的方式求得 $t_{R(o),R(d)}^{(\text{非行车})}$ 和 $t_{o,d}$ 提供了便利。令 $\{t_{o,d}^{(1)}, \dots, t_{o,d}^{(n)}\}$ 表示在较长的一段时间(例如一个月)内,从站点 o 到 d 的所有历史交易记录中所获得的实际乘客从进闸到出闸的时间开销。针对第 k 次交易记录,就可以计算得出其非行车时间为 $t_{o,d}^{(k)} - \hat{t}_{o,d}^{(\text{行车})}$ 。综合所有的交易记录,研究可以取计算得到的非行车时间集合的中位数 $\bar{t}_{o,d}^{(\text{非行车})}$ 作为从该数据集中获得的统计值。这里采用中位数而非平均数的原因是,平均数较易受少数实际时间开销特别大的异常交易的干扰。

基于相同区段的候车时间的共性,从充分利用历史数据的角度出发,可以将相同出发区段内的所有出发站点和目的区段内所有目的站点构成各种组合,针对每一种组合计算非行车时间的中位数,并将所有组合的中位数综合在一起,计算平均数,获得对 $t_{R(o),R(d)}^{(\text{非行车})}$ 的估计。具体的算法如下:

(1) 遍历线路上所有的出发区段 O 和目标区段 D 。

(2) 获取该出发区段和目标区段的出发站点、目标站点组合所构成的集合 $O \times D = \{(o, d) : R(o) = O \wedge R(d) = D\}$ 。

(3) 针对每一种组合 (o, d) , 从其历史交易记录中获得实际时间开销 $\{t_{o,d}^{(1)}, \dots, t_{o,d}^{(n)}\}$, 并据此计算非行车时间 $\{t_{o,d}^{(1)} - \hat{t}_{o,d}^{(\text{行车})}, \dots, t_{o,d}^{(n)} - \hat{t}_{o,d}^{(\text{行车})}\}$ 。最终获得中位数 $\bar{t}_{o,d}^{(\text{非行车})}$ 。

(4) 将各种组合的中位数综合在一起,计算出该出发区段到目的区段的非行车时间平均数,推得该数学计算公式如下:

$$\hat{t}_{o,d}^{(\text{非行车})} = \frac{1}{\|O \times D\|} \sum_{(o,d) \in O \times D} \bar{t}_{o,d}^{(\text{非行车})}, \quad (11)$$

(5) 输入任一出发站点 o 和目标站点 d , 找到站点所在区段 $R(o)$ 和 $R(d)$, 估计其旅行时间为:

$$\hat{t}_{o,d} = \hat{t}_{o,d}^{(\text{行车})} + \hat{t}_{R(o),R(d)}^{(\text{非行车})}. \quad (12)$$

2.2 跨线进出站旅行时间的估算

如 1.2 节所述,经历一次换乘的旅行时间可表示为:

$$t_{o,d} = t_o^{(\text{进闸})} + t_{o,i}^{(\text{候车})} + t_{o,i}^{(\text{行车})} + t_{i,j}^{(\text{换乘步行})} + t_{j,d}^{(\text{候车})} + t_{j,d}^{(\text{行车})} + t_d^{(\text{出闸})}, \quad (13)$$

由于换乘站点 i 和出发站点 o 同线,研究可以利用 2.1 节中的分析得到的结果,从而得到:

$$t_{o,d} \approx t_{o,i}^{(\text{行车})} + [t_{R(o),R(i)}^{(\text{非行车})} - t^{(\text{出闸})}] + t_{i,j}^{(\text{换乘步行})} + t_{j,d}^{(\text{候车})} + t_{j,d}^{(\text{行车})} + t^{(\text{出闸})} = t_{o,i}^{(\text{行车})} + t_{j,d}^{(\text{行车})} + t_{R(o),R(i)}^{(\text{非行车})} + [t_{i,j}^{(\text{换乘步行})} + t_{j,d}^{(\text{候车})}], \quad (14)$$

令:

$$t_{i,j,R(d)}^{(\text{换乘})} = t_{i,j}^{(\text{换乘步行})} + t_{j,d}^{(\text{候车})}, \quad (15)$$

表示换乘时间,则:

$$t_{o,d} \approx t_{o,i}^{(\text{行车})} + t_{j,d}^{(\text{行车})} + t_{R(o),R(i)}^{(\text{非行车})} + t_{i,j,R(d)}^{(\text{换乘})}. \quad (16)$$

同样,研究也可以利用历史数据来估算 $t_{i,j,R(d)}^{(\text{换乘})}$ 。令 (o, d) 为某一出发目标站点组合,组合内的出发目标站点均只存在一条有效路径,且该路径经过 (i, j) 一次换乘。用 $\{t_{o,d}^{(1)}, \dots, t_{o,d}^{(n)}\}$ 表示在较长的一段时间(例如一个月)内,从站点 o 到 d 的所有历史交易记录中所获得的实际乘客从进闸到出闸的时间开销。针对第 k 次交易记录,研究就可以计算得出其换乘时间为 $t_{o,d}^{(k)} - [\hat{t}_{o,i}^{(\text{行车})} + \hat{t}_{j,d}^{(\text{行车})} + \hat{t}_{R(o),R(i)}^{(\text{非行车})}]$ 。综合所有的交易记录,可以取计算得到的换乘时间集合的中位数 $\bar{t}_{i,j,R(d)}^{(\text{换乘})}$ 作为从该数据集中获得的统计值。

与 2.1 节类似,研究可以将相同出发区段内的所有出发站点和目的区段内所有目的站点构成各种组合,并甄选出只存在一条有效路径,且该路径经过 (i, j) 一次换乘的组合,针对每一种组合计算非行车时间的中位数,并将所有组合的中位数综合在一起,计算平均数,获得对 $t_{i,j,R(d)}^{(\text{换乘})}$ 的估计。具体的算法如下:

(1) 遍历换乘线路上所有的出发区段 O 和目标区段 D 。

(2) 获取该出发区段和目标区段的出发站点、目标站点组合所构成的集合,可将其写作如下数学形式:

$$\overline{O \times D} = \{(o, d) : R(o) = O \wedge R(d) = D \wedge (o, d) \text{ 只存在经 } (i, j) \text{ 一次换乘的一条有效路径}\}, \quad (17)$$

(3) 针对每一种组合 (o, d) , 从其历史交易记录中获得实际时间开销 $\{t_{o,d}^{(1)}, \dots, t_{o,d}^{(n)}\}$, 并据此计算

换乘时间, 即:

$$\{ t_{o,d}^{(1)} - [\hat{t}_{o,i}^{(行车)} + \hat{t}_{j,d}^{(行车)} + \hat{t}_{R(o),R(i)}^{(非行车)}], \dots, t_{o,d}^{(n)} - [\hat{t}_{o,i}^{(行车)} + \hat{t}_{j,d}^{(行车)} + \hat{t}_{R(o),R(i)}^{(非行车)}] \}, \text{ 最终获得中位数 } \bar{t}_{i,j,(o,d)}^{(换乘)}$$

(4) 将各种组合的中位数综合在一起, 计算出该出发区段到目的区段的换乘时间平均数, 可使用如下公式计算求出:

$$\hat{t}_{i,j,R(d)}^{(换乘)} = \frac{1}{\|O \times D\|} \sum_{(o,d) \in O \times D} \bar{t}_{i,j,(o,d)}^{(换乘)}, \quad (18)$$

(5) 输入任一出发站点 o 和目标站点 d , 换乘站 (i, j) , 站点所在区段 $R(o), R(i)$ 和 $R(d)$, 估计其旅行时间为:

$$\hat{t}_{o,d} = \hat{t}_{o,i}^{(行车)} + \hat{t}_{j,d}^{(行车)} + \hat{t}_{R(o),R(i)}^{(非行车)} + \hat{t}_{i,j,R(d)}^{(换乘)}. \quad (19)$$

2.3 实验结果

本文中实验用 2018 年 4 月 2 日~13 日间工作日早高峰(进站时间在 7:00~9:00 间)进出站点在 1 号线主线(莘庄站-上海火车站)上的所有 OD 对的历史数据, 以及进站站点在 1 号线主线上一次换乘、且只存在一条换乘路径的所有 OD 对的历史数据, 采用 2.2 节提出的算法估算了 2018 年 4 月 16 日~25 日间工作日早高峰进站站点在 1 号线主线上一次换乘、且只存在一条换乘路径的所有 OD 对(共 525 个 OD 对)的旅行时间。然后分别将本文算法估算值和上海地铁官网查询值与实际数据所得时间开销的统计值进行比较, 最终运行结果见表 1。

表 1 上海地铁官网查询值与本文算法估算值偏差比较

Tab. 1 Comparison of the offset between the query value of Shanghai Metro official network and the estimated value of this algorithm %

| 比较值 | 偏差平均值 | 偏差大于 10% 占比 | 偏差大于 20% 占比 |
|-------|-------|-------------|-------------|
| 官网查询值 | 5.3 | 17.0 | 1.9 |
| 算法估算值 | 3.4 | 4.8 | 0.6 |

由表 1 可知, 本文算法估算值在偏差平均值、偏差大于 10% 占比和偏差大于 20% 占比这 3 项评价指标上均优于官网查询值。研究针对算法估算值偏差超过 10% 的 OD 对分析发现均存在以下一种或全部现象: 周期内历史数据量非常少, 均小于 10 笔; OD 间站点数量非常少, 只有 2~4 站。

3 结束语

本文探讨了上海轨道交通清分模型的 K 短路径算法, 研究分析了影响城市轨道交通换乘路径选择和旅行时间的各类因素, 提出了一种基于统计的旅行时间估计的方法, 最后采用上海轨道交通的历史数据评价了该算法, 所得结果要优于上海地铁官

网提供的查询结果。

参考文献

- [1] 王明中. 城市轨道交通一票(卡)通换乘票务清分系统的研究与实现[D]. 上海: 上海交通大学, 2004.
- [2] 吴敏, 苏厚勤, 王明中. $k(k \leq 3)$ 条渐次最短路径搜索算法的研究及其实现技术[J]. 计算机应用与软件, 2004, 21(8): 81-83.
- [3] 黄胜, 孟世聪, 胡幼华. 轨道交通系统票务清分算法[J]. 计算机应用研究, 2004, 21(6): 104-106.
- [4] EPPSTEIN D. Finding the k shortest paths[J]. SIAM Journal on Computing, 1998, 28(2): 652-673.
- [5] ZHAN F B. Three fastest shortest path algorithms on real road networks[J]. Journal of Geographic Information and Decision Analysis, 1997, 1(1): 69-82.

(上接第 239 页)

- [2] 卞艺杰, 马玲玲. 云环境下移动视频监控系统安全性研究[J]. 计算机技术与发展, 2013, 23(9): 119-122, 127.
- [3] 魏利. 移动视频系统关键技术研究[D]. 长春: 吉林大学, 2010.
- [4] 杨琦. 基于极端环境下监控视频的图像去噪、增强与检索的研

- 究[D]. 西安: 西安电子科技大学, 2017.
- [5] 陈亮, 汪景福, 王娜, 等. 基于 DNN 算法的移动视频推荐策略[J]. 计算机学报, 2016, 39(8): 1626-1638.