

# 基于出行模式的公交 IC 卡乘客下车站点判断方法

陈 君<sup>1</sup>, 吕玉坤<sup>1,2</sup>, 崔美莉<sup>1</sup>

(1. 西安建筑科技大学 土木工程学院, 陕西 西安 710055; 2. 中铁四局集团设计研究院, 安徽 合肥 230023)

**摘要:** 对智能公交系统数据进行分析 and 挖掘, 能够为城市公共交通系统的规划和管理提供科学的决策依据。一票制的公交 IC 卡收费系统仅在上车时刷卡, 乘客下车站点的信息无法直接得到, 而现有方法只能判断出部分乘车记录的下车站点, 这成为公交 IC 卡数据应用的制约因素。根据公交乘客出行的时空规律性, 提出基于通勤出行模式和关联出行模式判断下车站点的思路。利用通勤出行规律判断公交乘客的通勤 OD, 再根据其居住地和工作地站点信息对通勤出行的下车站点进行估计。依据公交乘客长期出行具有一定时空关联性的规律, 通过将当日出行与他日的关联出行数据进行匹配来判断下车站点。对提出的方法进行算法实现, 并应用大规模实际数据对算法进行了实验和精度分析。研究结果表明: 该方法能够有效提高下车站点的判断比例, 准确率达到 87.90%, 能够应用于实际的公交系统规划和管理工作中。

**关键词:** 公共交通; 出行模式; 出行链; 下车站点; IC 卡数据

中图分类号: U491.1

文献标志码: A

文章编号: 1006-7930(2018)01-0023-07

## Estimating alighting stops of smart card public transportation passengers based on travel patterns

CHEN Jun<sup>1</sup>, LÜ Yukun<sup>1,2</sup>, CUI Meili<sup>1</sup>

(1. School of Civil Engineering, Univ. of Arch. & Tech., Xi'an, 710055, China;

2. Design and Research Institute of CTCE Group, Hefei 230023, China)

**Abstract:** Analyzing and mining the data of Advance Public Transportation Systems can provide scientific decision-making basis for urban public transportation planning and management. In flat-rate automated fare collection system, alighting stops information of the passengers is not available directly. The existing method can only determine part of alighting stops, which is a limiting factor for the application of smart card data. According to time-spatial regularity of transit travel, this paper presents the method of estimating alighting stops based on commuting travel patterns and related travel patterns. Commuting OD is judged by commuting regularity, and then the alighting stops of commuting trips are estimated on the basis of commuters' residences and work places. Based on long-term time-spatial correlation of travels, alighting stops are estimated by matching the related travels. The algorithm was implemented, and the experiment and accuracy analysis were carried out using large-scale actual data. The results show that the method can help improve the ratio of estimating alighting stops effectively, and the accuracy ratio reached 87.90%. This can be used in practical work of public transportation planning and management.

**Key words:** public transportation; travel patterns; trip chain; alighting stop; smart card data

公交 IC 卡收费系统在运行过程中产生大量的刷卡交易数据, 其数据具有采集方便、样本量大、连续记录等优点, 对这些数据进行分析 and 挖掘, 能够为城市交通系统的规划和管理提供丰富的决策信息<sup>[1-2]</sup>。公交 IC 卡收费系统有一票制收费和分段收费两类收费制式, 其中一票制收费

是国内、外最普遍的收费方式。一票制系统仅在上车刷卡, 乘客下车站点信息无法直接得到, 成为应用公交 IC 卡数据分析公交出行特征和规律的障碍。

现有方法是利用公交出行链的特征进行下车站点判断。出行链是指以家为起点和终点将居民

收稿日期: 2017-01-02

修改稿日期: 2018-01-10

基金项目: 国家自然科学基金项目(51208408); 陕西省自然科学基金基础研究计划项目(2017JM5121); 陕西省教育厅专项科研计划项目(17JK0461)

第一作者: 陈 君(1977—), 男, 副教授, 博士, 主要从事城市交通规划与管理研究。E-mail: chenjuntom@126.com

一天当中各种目的出行按发生顺序联接起来所形成的若干闭合链<sup>[3]</sup>。基于出行链的下车站点推算方法最早用来分析轨道交通乘客的出站站点。国外大部分城市轨道交通系统只在进站刷卡,乘客的出站站点没有记录,为了获得轨道交通乘客的出行 OD,纽约市城市公交机构<sup>[4]</sup>利用轨道交通乘客出行形成的出行链来推算乘客的出站站点。推算采用了 2 条假设:(1)同一乘客上一次乘车的终点,为下一次乘车的起点。(2)同一乘客一天最后一次出行的终点,为同一天第一次出行的起点。Jinhua Zhao<sup>[5]</sup>在纽约地铁 2 条假设基础上,对地铁客流的 OD 估计方法进行改进,利用地铁与公交的前后乘车站点之间的空间关系、乘客出行模式的对称关系推算芝加哥市地铁乘客的出站站点。Janine M. Farin<sup>[6]</sup>以巴西圣保罗市为案例,对圣保罗市交通小区间的普通公交 OD 矩阵进行了推算。Martin Trépanier 等<sup>[7]</sup>应用加拿大 Gatineau 市居民出行调查数据(HHT)与相同时段的公交 Smart Card 数据(SC)分析结果进行了比较分析,证明 SC 数据更为精确。Takahiko Kusakabe 等<sup>[8]</sup>从轨道交通中长期大量的数据中,对乘客的上下车站点进行了判断,并对乘客的出行行为进行了分析。Chen Jun<sup>[9]</sup>将出行链方法应用到常规公交系统站点间公交出行 OD 的推算。

这种利用公交出行链特征推算公交卡乘客下车站点的方法一般基于以下 3 个假设:

(1)“下一次出行(Next Trip)”假设:同一乘客在同一天中公交乘车的终点大多接近下一次乘车的起点。

(2)“最后一次出行(Last Trip)”假设:同一乘客在同一天中最后一次乘车的终点接近同一天第一次乘车的起点。

(3)“返程出行(Return Trip)”假设:同一乘客在同一天中如果连续两次乘车的线路相同,方向(上行和下行)相反,则第一次乘车的终点为第二次乘车的起点,第一次乘车的起点为第二次乘车的终点。

以上 3 个假设均需要同一乘客全天公交出行 2 次以上才能够应用。同时,如果乘客的公交出行链不闭合,则可能出现判断错误。出行链方法只能够判断部分公交卡乘车的下车站点,这成为应用公交卡数据分析公交出行行为特征和规律的制

约因素。

本文根据公交乘客长期出行具有时空规律性,提出利用公交乘客长期的出行模式进行公交卡乘客下车站点的判断,对方法进行算法实现,并应用大规模实际数据对算法进行了实验和验证。

## 1 研究数据描述

本文采集了某市 57 条智能公交线路的系统运营数据,包括公交 IC 卡数据、公交 GPS 数据、公交站点坐标数据和公交运行记录数据,公交 IC 卡数据来自公交自动收费系统,其他 3 种数据来自公交智能调度系统,4 种数据源主要字段之间的关联关系,如图 1 所示。

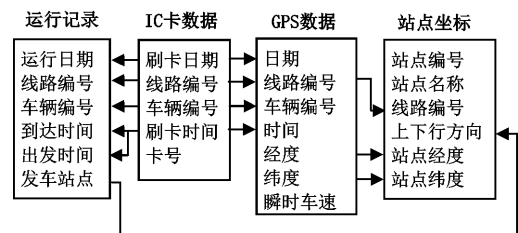


图 1 4 种数据源的关联关系<sup>[10]</sup>

Fig. 1 Association of four data sources<sup>[10]</sup>

在 SQL-Server 数据库导入原始数据并进行数据预处理,将智能调度系统数据与公交 IC 卡数据关联确定出 5 个工作日(周一~周五)公交 IC 卡乘客的上车站点<sup>[10]</sup>,计算过程限于篇幅不再赘述。将 5 个工作日的 899 174 条已知上车站点信息的公交卡数据作为本文的基础研究数据,其字段描述见表 1。

表 1 研究数据字段描述

Tab. 1 Fields description of research data

字段名	数据类型	字段内容
Card_ID	字符	公交卡编号
IC_Date	日期	消费日期
IC_Time	时间	消费时间
Bus_ID	字符	公交车辆编号
Route_ID	字符	公交线路编号
listno	字符	公交卡数据记录号
Direction_ID	字符	公交运行方向(上行或下行)
Boarding_StopID	字符	上车站点编号
Boarding_StopName	字符	上车站点名称
Stop_Longitude	数字	上车站点经度坐标
Stop_Laititude	数字	上车站点纬度坐标

## 2 基本思想

公交乘客的出行行为具有很强的时空规律性, 即大部分人都是在家和工作单位两点之间展开出行, 而出行时间和地点呈现 24 h 周期性, 公交乘客长期的出行规律是在其活动的空间尺度和时间尺度的制约下形成的<sup>[11]</sup>。根据公交乘客长期出行的时空规律性, 本文分别提出基于通勤出行模式和关联出行模式两种判断下车站点的思路。

### 2.1 基于通勤出行模式的下车站点判断思路

通勤交通是指居民在居住地和工作地(学校)之间的往返过程。在居民的日常公交出行中, 通勤出行占了很大的比重。从居民长期公交出行的上车站点信息可以判断出其早、晚高峰经常性的乘车站点, 从而能够确定出乘客的居住地站点和工作地站点。判断公交卡乘客的通勤 OD, 一般基于以下 3 条假设:

(1) 工作日早、晚高峰经常性乘坐公交的乘客为公交通勤乘客;

(2) 公交通勤乘客早高峰经常性的乘车站点为居住地站点;

(3) 公交通勤乘客晚高峰经常性的乘车站点为工作地站点<sup>[12]</sup>。

根据得到的公交通勤乘客居住地和工作地站点信息, 可能对通勤出行的下车站点进行判断。按照是否为直达通勤出行, 具体分为以下 2 种情况讨论。

#### (1) 直达公交通勤出行的下车站点判断

直达的公交通勤出行指乘客乘坐的是居住地站点和工作地站点之间的直达公交线路, 不需要换乘其它线路或其他交通方式就可以完成通勤出行。对于直达通勤出行, 若乘客在工作日早高峰首次乘坐的站点为居住地站点或附近站点, 且乘坐线路下游存在工作地站点或附近站点, 则可判定该下游站点为本次乘车的下车站点; 若乘客在工作日晚高峰首次乘坐的站点为工作地站点或附近站点, 且乘坐线路下游存在居住地站点或附近站点, 则可判定该下游站点为本次乘车的下车站点。

#### (2) 非直达公交通勤出行的下车站点判断

非直达公交通勤出行是指乘客需要通过换乘其他公交线路或其他交通方式完成通勤, 主要有 4 种模式:

- 1) 公交+公交;
- 2) 其他交通方式+公交;
- 3) 公交+其他交通方式;
- 4) 其他交通方式+公交+其他交通方式。

对于模式 1 和模式 2 的最后一次乘车, 若该次乘车线路下游方向存在某站点为工作地站点或居住地站点或附近站点, 则此可将此站点判断为下车站点。对于模式 3 和模式 4, 其公交乘车的下车站点距离工作地与居住地较远, 根据通勤 OD 信息无法判断。

### 2.2 基于关联出行模式的下车站点判断思路

公交乘客除了在通勤出行上具有高度的规律性以外, 在其他目的公交出行上也表现出一定的时空规律。例如, 在不同日期, 公交乘客经常在相近时间的相近站点上车, 在相近时间的相近站点下车。对于公交乘客当天的出行, 可能从该乘客往日的出行数据中找出具有时空关联性的出行进行匹配来判断当日出行的下车站点。根据对乘客长期公交出行时空规律的观察, 提出以下 4 条假设进行下车站点判断。

假设 1: (同线路同方向匹配) 同一乘客当日和往日的两次公交乘车, 如果乘车线路、乘车方向相同, 上车站点相同或接近, 则当日的下车站点与往日的下车站点相同或接近。

假设 2: (同线路反方向匹配) 同一乘客当日和往日的两次公交乘车, 如果乘车线路相同而乘车方向相反, 那么当日乘车的下车站点是往日乘车的上车站点或附近站点。

假设 3: (不同线路同方向匹配) 同一乘客当日和往日的两次公交乘车, 如果乘车线路不同, 乘车方向相同, 且上车站点相同或接近, 则当日乘车线路下游与往日乘车线路下车站点相同或接近的站点为当日乘车的下车站点。

假设 4: (不同线路反方向匹配) 同一乘客当日和往日的两次公交乘车, 如果乘车线路不同, 乘车方向相反, 则当日乘车线路下游与往日乘车上车站点相同或接近的站点为当日乘车时的下车站点。

### 3 算法实现

集成以上两种判断思路,提出基于出行模式的公交 IC 乘客下车站点判断算法.

Step1: 依次应用出行链方法的“返程出行”、“下一次出行”和“最后一次出行”3 条假设推算同 1 个星期 5 个工作日的公交卡乘客下车站点,算法详见文献[9].

Step2: 应用公交通勤 OD 估计方法识别该星期的通勤乘客卡号,并判断其居住地和工作地站点,将其作为通勤出行模式判断下车站点的匹配数据,算法详见文献[12].

Step3: 将 Step1 无法推算下车站点的公交出行数据提取出来,采用通勤出行模式进行判断.

Step4: 取出其中 1 条数据,如果该条数据的公交卡号为通勤乘客卡号,且刷卡时间为高峰时段,则判断此次出行为通勤出行.

Step5: 若该次通勤出行的公交线路下游方向存在工作地站点或居住地站点或两者附近站点,则判断该站点为下车站点,并且若该次出行上车站点为居住地站点或工作地站点或两者附近站点,则判定该次出行为直达通勤出行,否则为非直达通勤出行.

Step6: 将 Step4、Step5 未能判断出下车站点的数据提取出来,采用关联出行模式判断下车

站点.

Step7: 将 Step1 出行链方法已判断出下车站点的记录作为应用关联出行模式判断下车站点的匹配数据.

Step8: 取出其中 1 条数据,依次应用关联出行模式的假设 1~4 推算下车站点.

以上算法,采用 VB.NET 语言编程实现.

### 4 算法实验

#### 4.1 运算环境

算法程序运行的计算机硬件环境为: 8 核 2.4 GHz CUP; 16GB 内存; 2TB 硬盘. 软件环境为: Windows Server2008 操作系统; Visual Studio 2008 开发环境; SQL Server 2008 数据库.

#### 4.2 出行链方法运算结果

运行算法 Step1, 计算 5 个工作日全天出行 2 次以上的公交乘客的下车站点, 计算前将重复刷卡数据剔除, 并将上车站点不能全部判断的数据剔除. 以周五为例, 处理后的数据为 155 815 条, 对其中全天出行 2 次以上的 118 103 条记录进行运算, 判断出下车站点 104 493 条, 比例为 88.43%, 占全部实验数据 155 815 条的 67.03%, 计算过程详见参考文献[9], 限于篇幅不再赘述, 计算结果详见表 2.

表 2 出行链方法计算结果统计

Tab. 2 Statistics of trip chain method calculation results

项目	周一	周二	周三	周四	周五
“Return Trip”假设	48 220	50 976	53 420	54 066	54 166
推算成功记录/条 “Next Trip”假设	28 816	27 176	29 040	29 584	31 731
“Last Trip”假设	17 517	16 925	17 668	17902	18 542
小计	94 553	95 077	100 128	101 552	104 439
推算失败记录/条	11 139	11 082	12 438	12 204	13 664
合计/条	105 692	106 159	112 566	113 756	118 103
推算成功比例/%	89.46	89.56	88.95	89.25	88.43

#### 4.3 通勤 OD 判断方法运算结果

以周一~周五 5 个工作日的 182 007 个公交卡号, 共计 899 174 条公交 IC 卡记录作为算法 Step2

的计算数据判断通勤乘客卡号及其居住地站点和工作地站点, 计算过程详见文献[12], 运算结果详见表 3.

表 3 通勤 OD 运算结果统计<sup>[12]</sup>  
Tab. 3 Statistics of commuting OD calculation results<sup>[12]</sup>

项目	周一		周二		周三		周四		周五	
	早高峰	晚高峰	早高峰	晚高峰	早高峰	晚高峰	早高峰	晚高峰	早高峰	晚高峰
总卡号数	37 370	35 319	36 478	35 226	37 828	35 229	38 605	36 347	38 661	39 047
通勤卡号数	18 647	15 758	18 827	16 012	19 799	15 753	20 121	16 513	20 082	15 800
所占比例/%	49.9	44.6	51.6	45.5	52.3	44.7	52.1	45.4	51.9	40.5

4.4 出行模式方法计算结果

运行算法 Step3~Step8 判断周五出行链方法无法判断记录的下车站点, 共计 51 376 条数据, 其中包括全天出行 1 次的记录 37 712 条和 Step1 无法推算的记录 13 664 条. 以算法 Step1 得到的周一~周四的出行链方法计算结果和算法 Step2 得到的通勤 OD 计算结果作为周五判断下车站点的匹配数据.

算法运行结束, 周五全天出行 1 次的记录 37 712 条, 出行模式方法判断出下车站点的记录为

13 279 条, 比例为 35.21%, 运算结果如图 2 所示. 算法 Step1 采用出行链方法推算失败的记录 13 664 条, 应用出行模式方法判断出下车站点的记录 5 627 条, 比例为 41.18%, 运算结果如图 3 所示.

基于出行模式方法推算出周五下车站点的记录 18 906 条, 推算成功的比例为 36.80%. 其中, 关联出行模式的假设 1 判断成功的比例最高, 其次为通勤出行模式中的直达出行, 其他假设条件判断的比例较低, 算法实验结果见表 4.

图 2 刷卡 1 次记录的判断结果

Fig. 2 Calculation results of the records that only used once in a single day

图 3 出行链方法推算失败记录的判断结果

Fig. 3 Calculation results of the records that cannot be estimated by trip chain method

表 4 出行模式方法计算结果统计  
Tab. 4 Statistics of travel pattern method experiment results

项目		刷卡 1 次记录		出行链方法推算失败记录		合计	
		记录数/条	比例/%	记录数/条	比例/%	记录数/条	比例/%
通勤出行模式	直达	4 877	12.93	958	7.01	5 835	11.36
	非直达	177	0.47	326	2.39	503	0.98
关联出行模式	假设 1	5 352	14.19	2 618	19.16	7 970	15.51
	假设 2	1 477	3.92	788	5.77	2 265	4.41
	假设 3	1 110	2.94	676	4.95	1 786	3.48
	假设 4	286	0.76	261	1.91	547	1.06
判断成功数据		13 279	35.21	5 627	41.18	18 906	36.80
实验数据		37 712	100	13 664	100	51 376	100

5 算法验证

5.1 验证思路

根据纽约城市公交机构的研究结果,出行链方法推算下车站点的准确率在 90% 以上<sup>[4-5]</sup>.将周五利用出行链方法推算出下车站点的记录提取出来,再次应用出行模式的方法判断这部分记录的下车站点,然后比较两种方法推算结果的一致性(距离小于 1 000 m 认为一致).假设出行链方法的正确率为 90%,若两种方法推算下车站点的一致性为 A,则可认为出行模式判断方法的精度

为 A/90%.

5.2 验证结果

提取周五由出行链方法判断出下车站点的记录 104 439 条,应用出行模式方法再次判断得到 66 869 条公交出行记录的下车站点,经过比较,两种方法推算结果一致的记录为 52 893 条,一致率为 79.10%,则可出估计出行模式方法的准确率为 87.90%(79.11%/90%).其中,采用通勤出行模式判断的准确率很高,达到了 97.54%,采用关联出行模式判断的准确率达到 81.66%,算法验证的详细结果见表 5.

表 5 算法验证结果  
Tab. 5 Verification results of the algorithm

判断方法		一致样本/条	总样本/条	一致率/%	正确率/%
通勤出行模式	直达	21 342	23 654	90.22	100
	非直达	1 724	2 620	65.80	73.11
	小计	23 066	26 274	87.79	97.54
关联出行模式	假设 1	22 965	28 677	80.08	88.97
	假设 2	4 362	6 952	62.74	69.71
	假设 3	2 053	3 923	52.33	58.15
	假设 4	454	1 043	43.53	48.36
	小计	29 834	40 595	73.49	81.66
总计		52 900	66 869	79.11	87.90

6 结论

现有方法只能判断出部分公交卡乘车的下车站点,成为应用公交卡数据分析公交系统出行行为特征和规律的瓶颈.本文根据公交出行具有的时空规律性,提出了基于出行模式判断公交 IC 卡乘客下车站点的方法.

(1)本文研究证明能够根据历史出行模式对公交乘客的下车站点进行估计.对于出行链方法不能判断的 32.97%的公交卡乘车记录,采用本文提出的出行模式方法能够判断出其中 36.80%的下车站点,出行模式方法使得能够判断公交卡乘车下车站点的比例提高了 12.13%,达到 79.16%.

(2)采用通勤出行模式判断的准确率为

97.54%, 采用关联出行模式判断的准确率为 81.66%, 出行模式方法总体准确率达到 87.90%, 精度能够满足实际公交系统规划和管理工作的需要。

(3) 本文限于数据条件, 仅采用了 5 个工作日、部分公交线路的数据进行了算法实验。随着智能公交系统的普及, 应用更长时段、城市全部公交线路数据进行计算时, 出行模式方法能够判断的比例和准确率还能够得到进一步提高。

## 参考文献 References

- [1] BAGCHI M, WHITE P. The potential of public transport smart card data[J]. *Transport Policy*, 2005 (12):464-474.
- [2] PELLETIER M, TREPANIER M, MORENCY C. Smart card data use in public transit: a literature review[J]. *Transportation Research Part C*, 2011, 19 : 557-568.
- [3] 褚浩然, 郑猛, 杨晓光, 等. 出行链特征指标的提出及应用研究[J]. *城市交通*, 2006, 4(2):64-67.  
CHU Haoran, ZHEN Meng, YANG Xiaoguang, et al. A study on trip-chain indices and their application [J]. *Urban Transport of China*, 2006, 4(2):64-67.
- [4] BARRY J, NEWHOUSER R, RAHBEE A, et al. Origin and destination estimation in New York city using automated fare system data[J]. *Transportation Research Record*, 2002, (1871):183-187.
- [5] ZHAO J H, RAHBEE A, WILSON N. Estimation a passenger trip origin-destination matrix using automatic data collection systems [J]. *Computer-Aided Civil and Infrastructure Engineering*, 2007, 22(5):376-387.
- [6] FARZIN J M. Constructing an automated bus origin-destination matrix using farecard and GPS data in Sao Paulo, Brazil [J]. *Transportation Research Record*, 2008, 2072: 30-37.
- [7] CHU K K, CHAPLEAU R, TREPANIER M. Driver-assisted bus interview (DABI): passive transit travel survey using smart card automatic fare collection system and its applications [J]. *Transportation Research Record*, 2009, 2105: 1-10.
- [8] KUSAKABE T, IRYO T, ASAKURA Y. Estimation method for railway passengers' train choice behavior with smart card transaction data[J]. *Transportation*, 2010, 37(5):731-749.
- [9] CHEN J, WANG Z F. Algorithm of estimating alighting bus stops of smart card passengers based on trip-chain[C]. 2nd International Conference on Civil Engineering and Transportation. Guilin: TTP, 2012: 1918-1921.
- [10] 陈君, 杨东援. 基于智能调度数据的公交 IC 卡乘客上车站点判断方法[J]. *交通运输系统工程与信息*, 2013, 13(1):76-80.  
CHEN Jun, YANG Dongyuan. Method of estimating smart card commuters OD distribution based on APTS data[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2013, 13(1):76-80.
- [11] 刘靛. 普适计算环境下居民交通行为特征研究[D]. 上海: 同济大学, 2008.  
LIU Liang. Research on inhabitant travel behavior under pervasive computing environment [D]. Shanghai: Tongji University, 2008.
- [12] 陈君, 杨东援. 基于 APTS 数据的公交卡乘客通勤 OD 分布估计方法[J]. *交通运输系统工程与信息*, 2013, 13(4):47-53.  
CHEN Jun, YANG Dongyuan. Method of estimating smart card commuters OD distribution based on APTS data[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2013, 13(4):47-53.

(编辑 桂智刚)