

基于公交 IC 卡数据的 OD 推算技术研究

Approaching Bus OD Matrices Based on Data Reduced from Bus IC Cards

周 涛 翟长旭 高志刚

(重庆市城市交通规划研究所 重庆 400020)

ZHOU Tao ZHAI Changxu , and GAO Zhigang

(Chongqing Transport Planning Institute, Chongqing 400020, China)

摘要：公交 OD 矩阵是进行公交调度和公交线网优化的基础资料。随着公交 IC 卡技术的广泛应用，如何利用 IC 卡的数据信息来推算公交 OD 矩阵以及换乘次数、出行次数等公交出行信息，具有很强的现实意义和实用价值。通过对 IC 卡统计数据特点的分析，利用数据挖掘技术，分别对公交线路站点 OD 和区间出行 OD 的推算方法进行了研究，并进一步研究了如何从 IC 卡统计数据中得到公交出行的其他信息。分别给出了由 IC 卡统计数据推算公交线路站点 OD 和区间出行 OD 的方法，并提出了基于 GIS 的公交 IC 卡数据分析处理系统框架，建立了系统分析流程来实现公交 OD 的推算。通过分析发现，公交 IC 卡的数据分析处理最终要建立起能将数据库的处理功能和地理空间的属性特性结合起来的地理信息系统，才能够有效地解决公交 OD 推算、换乘次数以及出行次数等问题。

Abstract：Bus OD matrices are crucial inputs for bus dispatching and bus network optimization. With a wide application of IC (Intelligent card) card technologies, it is both meaningful and valuable to investigate how the data collected from bus IC cards can be fully utilized to approach OD matrices and calculate the numbers of bus transfers and trip frequencies. Based on the characteristics of bus IC card data, this paper develops methods for estimating both stop-based ODs and link-based ODs between two stops using data mining technology. The paper further considers what information else can be obtained from IC card data. The proposed methods aim to approach bus stop OD data and regional trip OD data using a systematic framework based on GIS. Meanwhile, a flow chart is given in the paper to help reduce bus OD data. It was concluded that GIS, which combines data processing function with geographical attributes, is the best way to process IC card data. Using this method, the problems related to bus OD matrices, transfer times, and trip frequencies, can be resolved effectively.

关键词：IC 卡；OD 推算；站点 OD；出行 OD

Keywords：IC cards；approaching OD matrices；station OD；trip OD

中图分类号：U491

文献标识码：A

收稿日期：2006-08-02

作者简介：周涛，男，重庆市城市交通规划研究所副所长，高级工程师，主要研究方向：交通规划与管理。E-mail:taozhou@cq139.com

描述公交出行空间分布的 OD 矩阵是进行日常公交调度和线网优化的基础性资料。传统的调查方法是在乘客上车时发放调查表格，每个人按照自己的出行过程如实填写表格，然后在下车的时候收回表格。这种人工调查方法不仅繁琐，而且耗费人力、物力和财力。在实际操作过程中，难以做到经常性、系统性。公交 IC 卡技术的普遍应用为公交 OD 矩阵调查提供了一种新的方法^[1]。目前，全国已超过 100 个城市启动了公交 IC 卡收费系统，累计发出公交 IC 卡几千万张，IC 卡已经成为公交企业认可的一种理想的收费手段，在我国的市场潜力还很大。IC 卡信息量大且全面，技术简单成熟，其统计数据中蕴含着丰富的信息，但作用尚未得到充分发挥。IC 卡的统计数据仅停留在 IC 卡开发商手中，没能与公交运营者建立直接、及时的联系。本文根据公交 IC 卡数据特点，对线路站点 OD 和区间出行 OD 的推算方法进行研究，并对 IC 卡数据分析处理系统的实现方式进行了探讨。

1 公交 IC 卡数据特点

目前，国内城市公交 IC 卡多采用乘客不分站点的上车刷卡形式^[2]。通过对 IC 卡数据的一些简单汇总分析可以得到不同线路在各时段客流量的大小，但反映的信息只限于线路客

流的总体情况,不能提供准确的各站点各断面的客流变化,得不到完整的乘客出行信息。如果可以通过某种处理方法,使IC卡记录中隐含的信息能够显现,缺失的信息能够还原,将能得到持卡乘客每条记录对应的起讫点,即公交客流最基础的资料——公交OD矩阵。

IC卡统计表主要包括四个字段,即卡号、划卡时间、线路号和乘坐车号。不同的卡号代表着不同的乘客,根据卡号信息查找乘客一天的划卡次数,从而确定乘客换乘或者二次出行信息。有条件时,还可将乘客的某次出行定位,按照乘客出行规律确定起讫点。划卡时间即乘客上车的时间,根据线路上某具体车辆的运营调度信息以及线路站点顺序,依据划卡时间来判断乘客的上车站点。线路号表示乘客选择的某条线路,在换乘和第二次出行时会涉及到不同线路之间乘客乘坐信息的搜索。乘坐车号代表了乘客乘坐的具体车辆,线路出行的总OD需要将该线路上所有运营车辆OD进行叠加计算。

2 线路站点OD推算

线路站点OD即以乘客上车站点为O、下车站点为D的矩阵。由于线路有上下行方向,对于某条线路的某个方向站点OD矩阵来说是对角线元素为零的上三角阵,如果将车辆运行的上下行方向OD相叠加就得到车辆运行一个循环所构成的OD矩阵。

在利用IC卡统计数据进行线路站点OD推算中,最难处理的就是下车站点的确定。因为公交IC卡采用的是上车刷卡的形式,虽然刷卡信息中没有明确记录上车的站点,然而根据划卡时间和运营车辆的调度信息,可以推算乘客的上车站点。车辆按照顺序停靠站点,划卡时间为车辆停靠站时间,同一站点乘客划卡时间前后相差不大,再进入路段继续运行,然后又遇到车站停车,车辆在线路上不断重复这个过程,因此可以很容易从乘客的划卡时间和站序推算乘客上车站点。然而,乘客下车站点的确定相对复杂。本文主要研究在采取一次上车刷卡的城市中,如何充分挖掘统计数据信息和居民出行规律,推测乘客下车站点的问题。

2.1 根据 IC 卡统计数据推算

为了准确推算乘客下车站点,应尽量从统计数据中挖掘可靠信息。对乘客一天内出行而言,根据IC卡统计信息推算主要是针对有两次及两次以上刷卡信息的乘客,对于只有一次刷卡信息的乘客,只能采用理论估算的方法推算站点。有两次及两次以上刷卡信息的乘客,下车站点的估算主要依据第二次刷卡的站点与第一次刷卡站点空间位置的接近度。即使这样,出行的最后一次乘车终点仍然必须采用理论估算的方式。根据参考文献[2]中所述方法,可以将居民的出行分为常发性出行和偶发性出行两类,对于乘客的某次出行来说,如果能判断其属于哪类出行,可以以此为依据判断乘客下车站点,特别对于乘客返程的下车站点推算,如果能增加对出行方向等定性的分析,就更能增加下车站点推算的准确度。以上数据的利用建立在长期对乘客出行的数据统计基础之上。

2.2 根据出行理论估算

由于居民出行采用的交通方式和出行目的的多样性等特点,有些情况不能根据统计数据推算下车站点。例如,乘客一天中只有一次刷卡行为或者乘客一天中最后一次出行等。这就需要根据线路运行规律进行估算。此外,虽然根据刷卡站点的关联性,可以比较准确地定位大部分乘客下车站点,但是统计数据量大、过程繁琐,而且包括很多交叉信息和空间信息。为了简化站点统计模型,这里介绍一种比较简单实用的方法——吸引权系数法^[3]。

1) 吸引权系数站点OD推算

根据对公交出行特性的分析,如果*i*站的上车人数一定时,*j*站对*i*站的吸引越大,那么*i*站到*j*站的OD分布量也越大。因此,定义 F_{ij} 为与*i*、*j*站点用地性质有关的系数,称为吸引权系数^[3]。从站点来看,某站点上车的人数越多,就说明该站的发生量越大;下车的人数越多,就说明该站的吸引量越大。因此,可以用各站点的上车人数和下车人数将各站点的用地性质数量化。线路OD矩阵形式应为上三角矩阵,当*i* > *j*时,吸引权系数不必计算。当*j* ≥ *i*时,吸引权系数 F_{ij} 可定义为:

$$F_{ij} = \begin{cases} 10 \log T_i U_j, & T_i = 0, 1 \text{ 且 } U_j = 0, 1 \\ 0, & T_i = 0, 1 \text{ 且 } U_j = 0, 1 \end{cases} \quad (1)$$

式中: T_i 为 i 站上车总人数; U_j 为 j 站下车总人数。

根据单约束重力模型中的乌尔希斯重力模型的思想^[4], 建立推求公交站点OD的结构化模型为:

$$x_{ij} = T_i U_j F_{ij} / \sum_j U_j F_{ij}, \quad (2)$$

式中 x_{ij} 为 i 站到 j 站的OD分布量。

根据 x_{ij} , 可求出理论上各站上车人数 T_i' 和各站下车人数 U_j' 。由于 $T_i' = \sum_{j=1}^n x_{ij} = \sum_{j=1}^n T_i \frac{U_j F_{ij}}{\sum_j U_j F_{ij}} = T_i$, 而 $U_j' =$

$$\sum_{i=1}^n x_{ij} = \sum_{i=1}^n T_i \frac{U_j F_{ij}}{\sum_j U_j F_{ij}} = U_j, \text{ 引入修正系数 } k_i = T_i / T_i', k_j =$$

U_j / U_j' , 对分布量 x_{ij} 进行修正: $x'_{ij} = x_{ij} k_i k_j$ 。如此进行迭代计算, 直至第 k 次计算的修正系数 k_i^k 、 k_j^k 接近 1, 这时的 x'_{ij} 即为推算出的站点OD分布量。

2) 各站点下车人数估算

站点上车人数是已知数, 要根据上述公式推算站点OD, 还需进行站点下车人数的估算。通过对同一线路上下行两方向的同一站上下车人数进行统计, 发现其上行某站的下车人数占该上行客运总人数的比例与下行该站点的上车人数占该下行客运总人数的比例相当。用公式表示为:

$$\frac{\text{上行某站的下车人数}}{\text{上行上车总人数或下车总人数}} \cdot 100\% = \frac{\text{下行某站的上车人数}}{\text{下行上车总人数或下车总人数}} \cdot 100\% \quad (3)$$

因此, 某站点的下车人数可以利用式(4)求解:

某站点下车人数 =

$$\text{该方向上车总人数} \cdot \frac{\text{对向该站点上车人数}}{\text{对向上车总人数}} \quad (4)$$

3 区间公交出行OD推算

3.1 根据 IC 卡统计数据推算

在确定线路站点OD的基础上, 对区间公交出行OD进行推算。对于只有一次刷卡信息或者乘客的最后一次出行等不能根据IC卡统计信息进行推断的情况, 其站点O点或者D点即为出行的O点或者D点。然而对

于有两次及两次以上刷卡信息的出行OD与站点OD则有所不同, 站点OD只关心乘客在某条线路上下车站点的分布情况, 而乘客的上下车站点不一定是乘客出行的起讫点, 部分乘客需要换乘, 换乘可能发生在该线路上车前或者下车后, 因此, 区间出行OD推算显得更为复杂。

乘客从第一次刷卡到第二次刷卡之间, 可能是完成了一次出行任务, 也可能是完成从一条线路到另一条线路的换乘(见图1)。为了确定有二次刷卡信息的乘客到底是换乘还是第二次出行, 可以根据居民出行规律, 分析第一次和第二次刷卡时间间隔的长短来判断是哪种出行。时间间隔可能为(车辆运行时间+到站+出行)的时间, 也可能为(车辆运行时间+到站+换乘)的时间。可以采用参考文献[1]提供的方法对乘客出行行为进行推算。

当推算出各乘客的出行OD之后, 还要涉及到站点合并的问题。因为, 一个大城市中可能有几百条公交线路、上千个站点, 必须对站点进行合并, 重新划分交通小区。与普通交通需求预测的交通小区划分原理有所不同, 以站点合并为主的交通小区划分, 更注重公交站点覆盖范围对交通小区划分的影响。因此, 在划分时应尽量保持同一站点的覆盖区位于同一交通小区。

3.2 根据公交 OD 反推理论推算

根据站点OD的推算, 能够得出各站点的上下车人数和站点间的断面客流量。如果又已知部分通过调查或者推算得到的公交出行OD, 则可以根据公交OD反推理论进行推算。OD反推技术起源于机动车OD矩阵反推, 是依据交通观测所得的现状路网流量反推出交通分区现状出行的OD量。常用的OD反推模型有极大熵模型、广义最小二乘模型、信息极小模型、极大似然模型等等。

该过程也可以在美国 Caliper 公司开发的

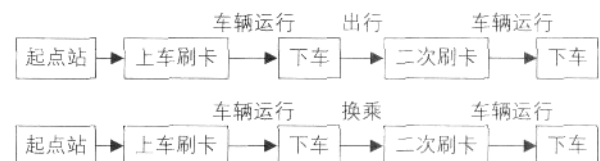


图1 公交乘客可能的出行方式

Fig.1 The possible trip form of bus passengers

TransCAD 软件上利用其中的“Transit O-D Matrix Estimation”模块来实现^[9]。在进行OD反推前首先需要在TransCAD中建立底层基础道路网、公交线网以及录入相关的属性信息,在此基础上再输入与公交反推相关的数据:如已知的部分公交出行OD矩阵以及各站点上下车人数或者线路断面流量,并选择合适的分配模型,从而实现对区间OD矩阵的推算。具体实现过程可见参考文献[5]中关于公交OD反推实现章节。

4 公交IC卡数据分析系统设计

4.1 数据分析处理系统设计框图

公交IC卡数据信息庞大,分析过程复杂,依靠人工的数据处理方式是不现实的,即使采用普通的数据处理软件也会比较困难。公交IC卡数据分析处理软件既要具有很强的数据统计功能,也要能进行站点线路搜索等与空间相关的功能。图2给出了利用IC卡统计数据进行分析的理论框架。图3给出了基于IC卡数据分析处理软件的系统设计流程图^[9]。地理信息系统(GIS)对现实世界、地理实体及相互关系进行抽象,不仅能管理海量的空间数据,也能管理地理要素的属性数据,并且能建立这两种数据间的联系^[7]。特别是在对公交出行OD矩阵进行推算时,地理空间属性信息显得更为重要。因此,可充分应用GIS的地理图形与数据库相结合的特点,对公交IC卡的统计信息进行分析处理。

4.2 线路站点 OD 推算简例

要对多条线路IC卡数据进行分析,需根据上述IC卡数据分析处理过程建立一套完整的基于GIS的数据处理软件。为说明问题,下面选一简例说明数据分析处理过程,利用统计软件完成线路OD的推算,线路的地理空间属性信息采用对应关系表的形式间接表示。

本例采用Access强大的数据统计特性,利用软件的查询功能,通过对表中的数据进行更新查询、选择查询、追加查询等操作,实现对线路OD矩阵的推算。若还需对OD矩阵的进一步操作,如OD合并等,通过该软件的查询功能也可实现。可先将原IC卡统计表进行扩充,如表1,除了包括原来统计信息外,又增加了

几个字段,主要用于对上下车站点和出行起讫点等的推算;然后再对其进行查询等操作,对于表中新增的字段,将采用数据更新的方式得到所需信息,如公交线路站点OD、区间出行OD、换乘次数以及刷卡次数等信息。同时,为了使IC卡统计数据与站点线路等空间属性相联系,还需建立对应关系表。可以建立如表2所示的时间、站点和车辆之间的对应关系表,也可建立站点和出行起讫点之间的对应关系表。最后,在原统计数据扩充表和对对应关系表之间采用公共字段

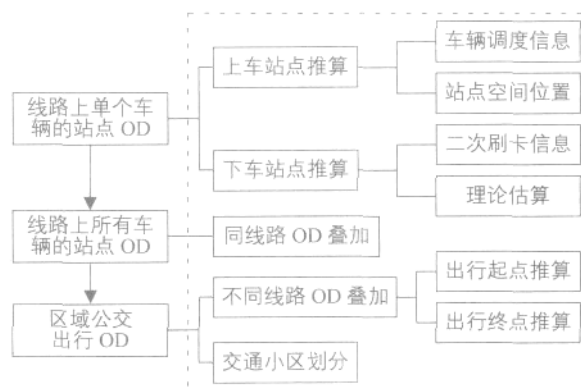


图2 由IC卡数据推算出行OD的分析过程

Fig.2 The process of reducing ODs from IC card data

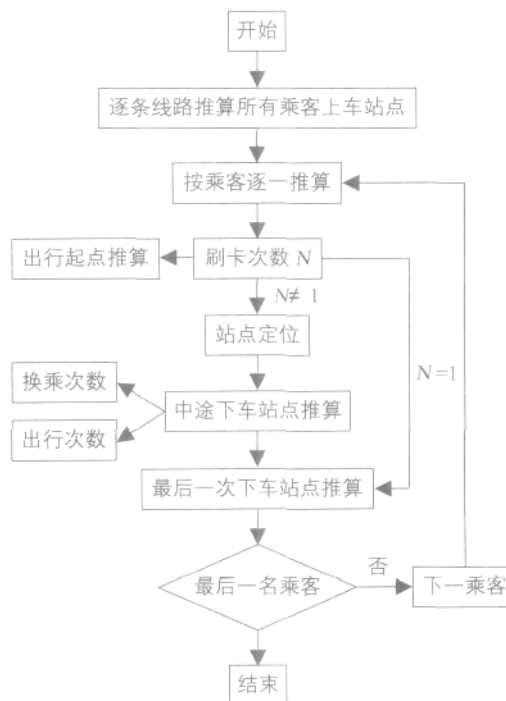


图3 由IC卡数据推算线路OD流程图

Fig.3 The flow chart for reducing link-based ODs from IC card data

表1 公交IC卡数据结构扩充表
Tab.1 The extended data structure of bus IC cards

卡号	划卡时间	线路号	乘坐车号	刷卡次数	出行次数	上车站点	下车站点	出行起点	出行终点
00185178	7:22	102	03455	"	"	"	"	"	"
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

表2 102线路划卡时间和站点对应关系表
Tab.2 IC card activation time by stops along Bus Route 102

均值时间	站点编号	乘坐车号
7:20	1	03455
7:30	2	03455
⋮	⋮	⋮

注：7：20 为公交车到站时间与离站时间的平均时间，即均值时间 = (到站时间 + 离站时间)/2。

索引的方式进行原统计表的查询等操作。

例如，对表1中的上车站点进行分析时，首先要建立如表2所示的对应关系表，再利用如下SQL查询语句：

UPDATE 表1 SET 表1.上车站点=表2.站点编号，
WHERE 表1.划卡时间 between (表2.均值时间 - 0.05) and (表2.均值时间+0.05)。

按照上述方法对表1进行更新，则可得到乘客的上车站点。同理，根据其他对应关系表可以对表1中的其他字段进行更新等操作，可推算出公交线路的站点OD。

5 结语

公交IC卡技术的广泛应用，为公交OD的调查提供了一种全新的手段。本文讨论了如何根据IC卡统计数据

和区间出行OD的推算方法。通过分析发现，公交IC卡的数据分析处理最终要建立起能将数据库的处理功能和地理空间的属

性特性结合起来的地理信息系统，才能够有效地解决公交OD推算、换乘次数以及出行次数等问题。

参考文献

1 陈学武, 戴霄, 陈茜. 公交 IC 卡信息采集、分析与应用研究 [J]. 土木工程学报, 2004, (2): 105~110
2 师富民. 基于 IC 卡数据的公交 OD 矩阵构造方法研究 [D]. 长春: 吉林大学, 2004
3 朱从坤, 丁建霆, 陈瑜. 公交线路 OD 反推的结构化模型研究 [J]. 哈尔滨工业大学学报, 2005, (6): 851~853
4 王伟, 徐吉谦, 杨涛, 李旭宏, 等. 城市交通规划理论及其应用 [M]. 南京: 东南大学出版社, 1998
5 Caliter Corporation. TransCAD Transportation GIS Software (Travel Demand Modeling with TransCAD 4.0) [Z]. America: Caliter Corporation, 2000
6 O'Neill W..Linear Linear Location Referencing Methods: Pasts, Present, and Future [M]. America:Utah State University, 1996
7 贾世贤. 基于 GIS 的公交规划管理信息系统的研究与开发 [D]. 北京: 中国科学院研究生院, 2004

(上接第57页)

3 Abdel-Aty M, Kitamura R, Jovanis P. Investigating effect of travel time variability on route choice using repeated measurement stated preference data [J]. Transportation Research B, 1995, (1493): 39~45
4 Lo HK, Luo XW, Siu BWY. Degradable transport network: travel time budget of travelers with heterogeneous risk aversion [J]. Transportation Research B, 2006, (40): 792~806
5 黄海军. 城市交通网络平衡分析理论与实践 [M]. 北京:

人民交通出版社, 1994
6 刘海旭, 蒲云. 基于行程质量的随即用户平衡分配模型 [J]. 中国公路学报, 2004, (4): 93~96
7 Willian H. K. Lam and G.Xu. A Traffic flow simulator for network reliability assessment [J]. Journal of Advanced Transportation, 1999, 32 (2): 159~182
8 Hu Shao, William H. K, Lam, Mei Lam Tam. A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand [J]. Networks and Spatial Economics, 2006, (6): 173~204