

【文章编号】1672-5328(2005)04-0073-04

单条公交线路的IC卡数据分析处理方法

戴 霄 陈学武

(东南大学交通学院, 南京 210096)

【摘 要】及时而准确地掌握城市公共交通运营及客流信息, 是进行科学合理的公交运营决策和公交网络规划的前提。随着城市公交IC卡的广泛运用, 从IC卡信息中提取城市公交客流信息成为又一种公交数据采集方法。通过公交IC卡数据分析处理方法, 可以得到实时准确且连续的公交客流信息数据, 较之传统的人工调查方法经济实用, 数据结果准确度高。以单条公交线路的IC卡信息为例, 提出单条公交线路运营及客流信息的分析处理方法, 得到包括站点客流、线路客流、断面客流等主要客流指标的分析计算方法。

【关键词】公交IC卡; 公交客流; 数据统计

【中图分类号】U491

【文献标识码】A

The Method of Intelligent Card Data Analysis for One Public Transportation Route

DAI Xiao, CHEN Xuewu

(Transportation College, Southeast University, Nanjing 210096, China)

Abstract: Accurate decision for public transport and scientific planning for the network are based on the exact and detailed bus passenger flow information. With the extend of bus Intelligent Card (IC), getting bus passenger flow information data from bus IC becomes one method. It is more precise and continuous than the data getting by traditional method. This paper discusses the IC data of one bus line and puts forward the analysis method of IC data. Then the formulas of calculate station passenger flow, line passenger flow and section passenger flow are put forward.

Keywords: bus Intelligent Card; flow of bus passenger; data statistics

公交客流信息作为城市公共交通运输规划和运营决策的依据, 规划者和管理者一直在致力于获取准确反映公交运营状况的客流信息。公交乘客是公共交通客流组成的基本要素, 公交IC卡(Intelligent Card)数据库记录了每一个公交乘客任何一次刷卡的相关信息, 从这些信息中挖掘出的反映公交运营状况的数据是可行的^[1]。公交IC卡在国内许多城市都有应用, 很多大中城市持卡乘客比例达到60%以上(南京市已接近80%)。IC卡信息量所反映的居民公交出行的数量及分布完全达到调查样本数量和特征的要求, 与人工调查方法相比, IC卡信息不容易受人为因素的影响, 数据准确真实, 得到的客流结果更能真实反映城市公

交运营情况和居民公交出行特征。由于IC卡数据的真实性和准确性, 利用其分析得到的客流信息结果同样真实反映现实状况, 较之目前规划领域使用的人工调查方法更准确、更经济^[2]。

1 IC卡数据库

公交IC卡数据库包含两个不同类型的数据库系统: 公交IC卡操作型数据库和公交IC卡数据仓库。操作型数据库是为完成公交IC卡刷卡系统的日常事务而建立的, 它负责直接读取并存储由刷卡终端传递的刷卡数据。操作型数据库中数据是可变的, 公交IC卡数

收稿日期: 2005-03-23

作者简介: 戴霄(1980—), 男, 东南大学交通学院硕士研究生。E-mail: alufei@163.com

据采集1次则该数据库更新1次。

数据仓库是为进行公交IC卡数据分析而建立的数据库系统。本文为分析单条公交线路的IC卡数据处理方法，建立单条公交线路的公交IC卡数据仓库，将城市公交线路、公交运营基础信息(发车班次、发车间隔等)以及初步筛选的IC卡基础数据库中的刷卡信息提取出来，并保存在这个数据仓库中。IC卡数据仓库在分析过程中是不可变的。建立数据仓库的目的是减少冗余数据和错误数据干扰，以加快访问数据的速度，提高分析过程的效率和分析结果的精度。IC卡操作型数据库是数据源，而IC卡数据仓库则是直接分析的对象。

2 IC卡数据筛选及数据仓库的建立

IC卡记录单个乘客刷卡信息，包括：乘客卡号(ID)、线路号(X)、刷卡日期(DATE)、刷卡时刻(T)、刷卡站点(P)、车辆代号(C)。其中IC卡信息不直接记录刷卡站点，在数据仓库中是利用线路调度数据采用时间匹配的方法来确定的。信息在IC数据库中保存为表的形式，每一行代表

表1 部分公交IC卡刷卡数据(3路)
Tab.1 Part of bus Intelligent Card Data (No.3)

卡编号	卡余额 /元	消费金额 /元	消费日期	消费时间	数据类型	汽车编号	记录号
.....							
00000697	138.00	0.90	04-12-08	09:14:25	01	010187	171017714643
00000823	59.00	1.80	04-12-08	08:02:19	01	018064	151014037361
00000837	113.10	0.90	04-12-08	06:38:39	01	010071	171017758065
00000837	110.40	0.90	04-12-08	16:50:26	01	010036	151014043805
00000862	22.20	1.80	04-12-08	19:40:58	01	018062	151014036575
00000862	20.40	1.80	04-12-08	19:41:00	01	018062	151014036576
00000870	11.30	0.90	04-12-08	14:31:02	01	010227	171017705685
.....							

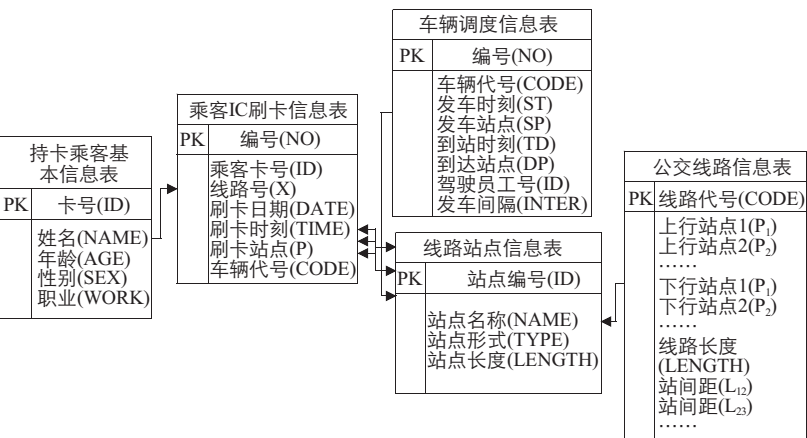


图1 公交IC卡数据仓库数据表结构

Fig.1 Data's framework for bus Intelligent Card data warehouses

一次刷卡记录，每一列代表刷卡记录的一个属性^[3]。表1列出常州市3路车2004年12月8日的部分公交IC卡刷卡数据^[2]。

1) 选择分析所需要的基础数据

对单条公交线路IC卡数据进行分析，除需要IC卡数据之外，该线路车辆及人员配置和每日的线路调度数据也是必须的。如果要对公交乘客的出行特征进行分析，还需公交乘客的基本信息数据。

2) 筛选数据库中错误数据

公交IC卡数据采集过程均由计算机完成，因此，数据出错的机率很小。需要注意的就是如果某日数据采集过程中大量数据丢失，对于该日的数据则不能采用。

3) 建立数据仓库

从IC卡操作型数据库中提取所需乘客刷卡数据，从公交管理数据库中提取所需车辆、线路等数据，利用这些基础数据建立数据仓库。相关基础数据在数据仓库中的关系如图1所示。

3 线路客流数据统计分析

以一条公交线路的公交IC卡数据为例，进行客流统计分析。由于上下行方向客流分析方法基本相同，本文仅给出上行方向客流数据统计方法。对线路基本运营数据作如下假设：线路分上下行两个方向，对向发车；上行方向M个站点(包括首末站)；上行方向全天发车次数为K，首班车发车时刻为T₀、首班车发车时间计为t₀=0；上车人数为S，下车人数为D；公交车代号C；线路上公交IC卡使用率为r^[4-5]。

3.1 站点上车人数统计

3.1.1 判断刷卡站点

已知乘客刷卡站点, 则可以直接统计得到各线路站点任意时段上车人数, 但是目前公交IC系统并不支持记录刷卡站点的功能。本文通过对IC数据进行归类, 利用已知的线路调度资料推测刷卡站点。这种方法准确性较高, 易于计算机实现。

将线路运营看作全天 K 次公交车辆由首站经中途站到末站终止的过程, 一个过程有一部公交车参与, 乘客刷卡行为是各过程中发生的事件, 一次刷卡事件只存在于一个过程之中, 并只发生在一个公交站点上。用 k 和 m 分别代表某一过程与某一站点, 基于这样的定义, 一条刷卡记录可描述为在过程 k 中于 m 站点发生的事件。

首先将刷卡记录分类到各过程。调度基础数据中的发车时刻(ST)和到站时刻(DT)作为各过程的时间界限 $W_{kc} = (ST, DT)$ 。如果刷卡时刻 $T \in W_{kc}$, 且车辆代号与该过程车辆代号相同, 则认为该记录代表的刷卡事件是过程 k 中发生的。这样将一条公交线路的IC卡数据归类为 K 组 $Re(k)$ 。

然后将每组刷卡记录分类到各站点。公交车在站点停留的时间较短, 则认为 $Re(k)$ 中刷卡时刻相差时间在2 min以内的记录为在同一站点发生的事件。这些记录归类为该过程数据集的子数据集, 最多可有 $m-1$ 个子记录集, 可计算子记录集的平均刷卡时间。使用发车时刻和站点间平均行驶时间估算车辆到达各站点时刻的范围区间, 平均刷卡时间在某站点公交到达区间内, 则认为该子集的事件发生在该站点, 即子集中各记录对应刷卡站点为 p , 无子集对应的站点认为没有刷卡事件发生。各子记录集可表示为 $subRe(k, m-1)$ 。由以上分析, 可找到刷卡记录对应的发车车次及刷卡站点。

3.1.2 统计站点上车人数

已知乘客的刷卡站点和刷卡时间信息, 可运用简单的统计方法得到任意时间段各站点上车的人数。子集 $subRe(k, m)$ 包含的记录条数 $SUM(subRe(k, m))$ 即为 k 过程 m 站点的上车人数。则全天各站点上车人数可表示为:

$$S_i = \sum_{k=1}^K SUM(subRe(k, i)). \quad (1)$$

线路总上车人数即客流总量为:

$$S = \sum_{i=1}^m S_i \quad (S_i \text{ 为线路在 } i \text{ 站点上车人数}). \quad (2)$$

根据各子集 $subRe(k, m)$ 的刷卡时刻可得到各时

段站点上车人数和线路总客流量的数据。表示为“COUNT(subRe(k, i)) WHERE $T \geq t_0$ AND $T \leq t_1$ ”(t_0, t_1 为分析时段的时间上限和下限)。

考虑公交IC卡使用率, 需要对统计得到的上车人数作进一步调整。根据该线路调查日运营的现金收入推测得到全天投币乘车的人数 S_0 , 则该线路IC使用率 $r = \frac{S}{S+S_0} \times 100\%$ 。调整后全天各站点上车人数与线路总上车人数分别为: $S' = S/r, S'_i = S_i$ 。

3.2 推算站点下车人数

公交乘客下车不刷卡, 因此, 根据IC卡数据不能直接统计得到站点的下车人数。通常公交线路客流的流量流向是比较稳定的, 乘客由某站点上车到某站点下车的概率也是较稳定的。找到这个概率, 就可以推算出站点的下车人数。用 p_{ij} 表示在 i 站上车乘客在 j 站下车的概率, 建立下车概率矩阵 $P = (p_{ij})_{m \times m}$, 这里利用站点吸引率 x_i 来确定 p_{ij} 。

吸引率反映站点对乘客出行的吸引强度。考虑到上下行线路客流的对称性, 即上(下)行线路各站点上(下)车人数占总上(下)车人数比例与下(上)行线路相应站点下(上)车人数占总下(上)车人数比例接近。这主要是因为工作日公交乘客出行目的大部分为通勤及回程出行, 上班上学与其回程的公交出行线路方向相反、路线一致。根据下行线路各站点上车人数 S'_{bk} 运算得到线路各站点的吸引权重 x_i :

$$x_i = \frac{S'_{b(m-i+1)}}{\sum_{k=1}^m S'_{bk}}. \quad (3)$$

显然 i 站点上车的乘客可能在 $i+1, i+2, \dots, m$ 站下车, 得到站点下车矩阵的确定公式:

$$P_{ij} = \begin{cases} \frac{x_j}{\sum_{k=i+1}^m x_k} \times 100\%, & i < j. \\ 0, & i \geq j. \end{cases} \quad (4)$$

线路单向运行, 因此, 起点站下车乘客数为0, 即 $D_1 = 0$; 第2站下车人数来自于站点1, 可得到 $D_2 = S'_1 \times p_{12}$ 。类推得到各站全天下车人数:

$$D_i = \sum_{k=1}^{i-1} S'_k \times p_{ki}, \quad i = 1, 2, \dots, m. \quad (5)$$

3.3 线路断面流量推算

已知站点上车人数 S'_i 、 S'_k (车次 k 的各站点上车人数)及相应上车时间 t_{ki} , 利用式(5)推算下车人数 D_i 、 D_k (车次 k 的各站点下车人数)及相应下车时间 $t'_{ki} = t_{ki}$,

由于站点*i*与站点*i+1*间的断面流量为前*i*个站点的上车人数与下车人数之差,因此得到断面流量计算公式:

站点*i*与站点*i+1*间车次*k*的断面流量:

$$Q_{k(i+1)} = \sum_{x=1}^i S'_{kx} - \sum_{x=1}^i D_{kx}, \quad i = 1, 2, \dots, m-1. \quad (6)$$

站点*i*与站点*i+1*间全天断面流量:

$$Q_{i(i+1)} = \sum_{x=1}^i S'_x - \sum_{x=1}^i D_x, \quad i = 1, 2, \dots, m-1. \quad (7)$$

将高峰时段断面客流量累加得到高峰小时各断面流量 $Q_{i(i+1)\max}$, $i = 1, 2, \dots, m-1$ 。

3.4 相邻站点公交车辆平均运行时间

公交车辆在站点间的运行时间,直接决定了线路运营的效率及车辆调度决策。道路条件的变化经常会影响公交车在站点间的运行时间,不同站点间的道路条件不同,运行时间也不相同。

在统计分析各站上车人数时,得到各站上车刷卡记录子数据集。将子数据集的平均刷卡时刻作为该次车到达站点的时刻,表示为 $t_{ki} = \text{AVERAGE}(T_{\text{From subRe}(k, i)})$ (*k*次车到达站点*i*的时刻)。由此,相邻站点车辆平均运行时间可由公式(8)计算:

$$\bar{t}_{i(i+1)} = \frac{\sum_{k=1}^K (t_{k(i+1)} - t_{ki})}{K}, \quad i = 1, 2, \dots, m-1. \quad (8)$$

已知站点间平均站距 $l_{i(i+1)}$,也可以得到站点间车辆平均运营速度 $\bar{v} = l_{i(i+1)} / \bar{t}_{i(i+1)}$ 。

4 结语

对一条线路的公交IC卡信息统计分析,得到线路客流及运营信息包括:全天客流总量、各站点全天上车人数、高峰小时各站上车人数、高峰小时全线路客流量、全天线路各断面流量、高峰小时各断面流量、站点间车辆平均运行时间、站点间车辆平均运营速度、全天各站点的平均等车时间。这些分析结果比较全面准确,反映了该条公交线路现状运营及客流情况,可直接用于线路运营评价和调整。由于公交IC卡信息记录线路每天的客流数据,可以用同样的方法得到线路许多天的客流数据,掌握线路客流的变化趋势,进而预测未来短期内(如一个月)该线路客流情况,为决策者及时制定可靠的线路运行方案提供准确依据^[6]。

利用公交IC卡数据分析单条线路运营及客流情况方法简单,全过程可由计算机完成;根据需要可以分析过去任何一日的运营情况,所得到的数据结果准确可靠;可利用多天数据及相似运营日的客流数据结果,较准确地预测未来某时间客流趋势;实现线路客流信息实时掌握,分析决策周期大大缩短。但是仅仅利用公交IC卡数据分析得到单条线路客流数据,还不能反映整个公交系统的客流情况。进一步分析得到全市的公交OD是公交IC卡数据分析处理的下一步目标。

参考文献

- 1 陆永宁. IC卡应用系统[M]. 南京: 东南大学出版社, 2002. 53
- 2 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, 37(2): 105~110
- 3 夏火松. 数据仓库与数据挖掘技术[M]. 北京: 科学出版社, 2004. 26~35
- 4 戴名强, 李卫军, 李新鹏. 一类公交车调度问题的数学模型及解法[J]. 运筹与管理, 2004, 13(1): 73~74
- 5 王炜, 杨新苗, 陈学武. 城市公共交通系统规划方法与管理[M]. 北京: 科学出版社, 2002. 210
- 6 牛学勤. 城市公交车调度优化理论与方法[D]. 南京: 东南大学交通学院, 2004

《城市道路——施工图设计深度图样》05MR101

《城市道路——施工图设计深度图样》05MR101依据《道路工程制图标准》(GB 50162—92)、《城市道路设计规范》(CJJ 37—90)等规范,按照市政公用工程施工图设计文件编制深度的相关规定,借鉴工程实例,采取图文并茂,以图为主的形式,为城市道路工程施工图的绘制提供一种示范画法,以求对有关深度规定和制图标准予以细化和图样化。内容包括城市道路工程施工图绘制说明,图面绘制示范,施工图图样设计等。道路图样提供封面、目录、施工图设计说明、道路地理位置图、平面总体设计图、道路平面线位设计图、道路平面设计图、纵断面设计图、土方横断面、交叉口竖向设计等施工图设计图样和必要的绘制说明,以供设计人员在工程中参照。

本图集适用于我国城市各类新建、扩建和改建的快速路、主干路、次干路、支路的施工图设计,居住区道路可参照使用。

中国建筑标准设计研究院