

公交线路客流 OD 矩阵推算方法研究

Study on OD Matrix Estimation of Transit Passenger Flows

刘翠¹ 陈洪仁²

(1. 深圳市龙岗规划交通研究中心, 深圳 518172; 2. 哈尔滨工业大学, 哈尔滨 150090)

LIU Cui¹ and CHEN Hongren²

(1. Shenzhen Longgang Urban Planning Institute, Shenzhen 518172; 2. Harbin Institute of Technology, Haerbin 150090)

摘要: 公交线路客流OD矩阵是公共交通管理和控制不可缺少的信息。参照机动车OD矩阵反推技术, 对公交线路客流OD矩阵推算进行了专门研究。在公交线路客流OD调查数据的基础上, 分析得到居民公交出行规律。首次建立了根据公交乘客上下车人数推算公交线路客流OD矩阵的极大熵模型, 分别采用原始资料作为先验OD和重力模型形式的先验OD对模型进行求解, 结果证明该模型精度较高。

Abstract: According to OD matrix estimation of vehicle, OD matrix estimation of transit passenger flow was thoroughly studied. On the basis of the surveyed OD matrix data of transit passenger flow, the residents' transit trip rule was drawn. Maximum entropy model developed based on boarding and alighting counts was adapted to OD matrix estimation of transit passenger flow for the first time. The solutions of the model were derived using a given OD matrix as original data and gravity model respectively. The result proved that the model was accurate enough.

关键词: 公交线路客流OD; OD反推; 居民公交出行规律; 极大熵模型

Keywords: OD matrix of transit passenger flow; OD estimation; residents' transit trip rule; maximum entropy model

中图分类号: U491 文献标识码: A

收稿日期: 2006-06-30

作者简介: 刘翠, 女, 硕士, 深圳市龙岗规划交通研究中心工程师, 主要研究方向: 交通规划。

E-mail: liucui2000_ren@126.com

0 引言

公交线路客流OD矩阵是公共交通管理和控制不可缺少的信息, 但长期进行公交线路客流OD矩阵的观测是相当复杂的, 需花费大量的人力与财力, 且精度难以保证。而利用公交乘客上下车人数推算公交线路客流OD矩阵, 就可以减少人力、物力的投入, 且调查方便, 耗费较少, 准确性较高。

OD反推技术起源于机动车OD矩阵反推, 反推OD是依据交通观测所得的现状路段交通量资料推算出现状出行OD量。目前常用的OD矩阵推算模型有广义最小二乘法、最大似然法、最小信息量法、极大熵法和人工神经网络等方法^[1-4], 但将OD反推技术应用于公交线路客流OD分布的研究仍较少。

传统的公交线路客流OD矩阵推算方法的假设前提是当公交车到达 j 站时, j 站前上车的每一位乘客都以相同的概率在 j 站下车, 而不论他在前方哪一站上车。事实上, 居民公交出行的距离分布有一定规律, 因此传统的公交线路客流OD分布推算方法精度不高。

本文在参照机动车OD矩阵反推技术的基础上, 将极大熵法引入公交线路客流OD矩阵的推算, 在分析居民公交出行规律的基础上, 利用公交乘客上下车人数推算出公交线路客流OD矩阵的分布。

1 居民公交出行规律分析

本次客流OD调查的是哈尔滨市八路公交线路, 线路单程9.2 km, 沿线上行有16个公交站点, 下行有15个公交站点。调查采用票据法, 共进行了5天, 调查了20个运营车次, 上下行分别为10个车次, 得到1363人的OD情况, 其中上行方向743人, 下行

方向620人。

经分析可知，居民公交出行的距离分布有一定的规律，近似服从正态分布。居民的公交出行时耗主要集中在某个区域内，属于中长距离的出行，出行时耗过长或过短的居民，很少采用公交出行方式。此外，居民公交出行也受站点用地性质的影响。附近有大型购物休闲娱乐等设施的站点比普通站点吸引半径更大，而且多是交通枢纽，这类站点上下车的乘客较多。

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；下车人数越多，则该站的吸引力越大。在*i*站发生量一定的情况下，*j*站的吸引力越大，则*i*站到*j*站的OD分布量越大；反之，在*j*站吸引力一定的情况下，*i*站的发生量越大，*i*站到*j*站的OD分布量也越大。因此定义 W_{ij} 为与*i*站、*j*站用地性质有关的系数，称为吸引权系数。用各站的上车人数和下车人数将各站点的用地性质数量化，定义吸引权系数 W_{ij} 为：

$$W_{ij} = e^{-\frac{100 \cdot T_i \cdot U_j}{\sum_i T_i \cdot \sum_j U_j}} \quad (1)$$

式中： T_i 为*i*站的上车人数， U_j 为*j*站的下车人数。

在流量平衡条件下， $\sum_i T_i = \sum_j U_j = M$ ，其中*M*为总运营人数。

将上行和下行客流OD矩阵各项除以吸引权系数，得到各方向和双方向乘客乘行站数概率，见图1~2。

可以看出，去除公交站点用地性质的影响后，公交乘客乘坐一定站数时下车概率为最大，站数过大或过小时下车概率相对较小，即公交乘客乘行站数概率存在一个峰值，并向两侧逐渐减小。从图中可以直观地看出数据近似服从泊松分布。

通过 χ^2 检验法检验上行、下行乘客公交乘行站数是否符合泊松分布，见表1。检验结果可知，在显著性水平为0.05的条件下，上行和下行方向的数据都能接受假设。

通过以上分析可知，居民公交出行规律主要受公交出行距离和站点用地性质的影响，即：

$$X_{ij} \propto T_i, X_{ij} \propto U_j, X_{ij} \propto W_{ij}, X_{ij} \propto P_{ij},$$

其中： X_{ij} 为*i*站到*j*站的公交线路客流OD分布量； P_{ij} 为公交乘客乘行站数概率，表示为 $P_{ij} = \frac{e^{-\lambda} \lambda^{(j-i)}}{(j-i)!}$ ，其中 λ 为平均站距，即泊松分布系数。

2 公交线路客流OD矩阵推算的极大熵模型

极大熵法(Maximum entropy method, ME)是Willumsen于1981年提出的，其依据是极大熵原理。该方法模型结构简单，求解过程较为简单，估计精度也较高。

设某条公交线路上行方向有*n*个公交站点，则其线路客流OD矩阵形式为上三角矩阵。对某一公交站点*i*而言，其客流分别流向第*i*+1, *i*+2, ..., *n*个站点，即共有(*n*-*i*)个OD量，则总的OD量为 $\frac{n(n-1)}{2}$ 。

令 X_{ij} 为*i*站到*j*站的公交线路客流OD分布量，*i*=1, 2, ..., *n*-1; *j*=*i*+1, *i*+2, ..., *n*。由流量守恒知：

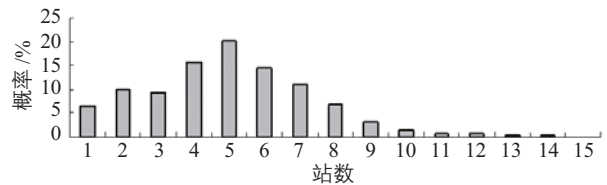


图1 上行乘客乘行站数概率统计直方图

Fig.1 Statistical probability chart for the number of bus stops passed by upward passengers

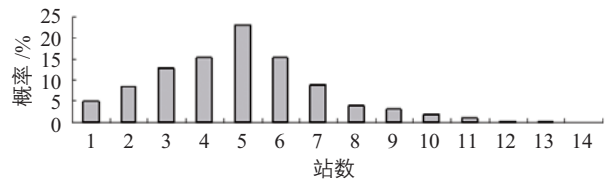


图2 下行乘客乘行站数概率统计直方图

Fig.2 Statistical probability chart for the number of bus stops passed by down passengers

表1 χ^2 检验结果表

Tab.1 χ^2 test result

参数和统计量	上行	下行
λ	5.02	4.94
χ^2	16.34	9.63
$\chi^2 \alpha$	$\alpha = 0.05$	16.919
	$\alpha = 0.025$	19.023
	$\alpha = 0.01$	21.666
		15.507
		17.535
		20.090

$$\begin{cases} \sum_{j=i+1}^n X_{ij} = T_i, i=1, 2, \dots, n-1 \\ \sum_{i=1}^{j-1} X_{ij} = U_j, j=2, 3, \dots, n \end{cases}, \quad (2)$$

将上述两个方程组合并为一个方程组，定义 V_k ，当 $k=1, 2, \dots, n-1$ 时， $V_k = T_i, i=1, 2, \dots, n-1$ ；当 $k=n, n+1, \dots, 2(n-1)$ 时， $V_k = U_j, j=2, 3, \dots, n$ 。则式(2)变为：

$$\begin{cases} \sum_{j=i+1}^n X_{ij} = V_k, k=1, 2, \dots, n-1 \\ \sum_{i=1}^{j-1} X_{ij} = V_k, k=n, n+1, \dots, 2(n-1) \end{cases}。 \quad (3)$$

设 P 表示方程组(3)的系数矩阵，则式(3)变为：

$$P X = V。 \quad (4)$$

极大熵模型的基本思想是：所求的交通分布是OD表中各OD交通量同时发生的概率达到最大时的交通分布，同时发生概率 W 根据多项式概率分布应为：

$$W(X_{ij}) = \frac{X!}{X_{12}! X_{13}! \dots X_{n-1,n}!} X_{12}^{x_{12}} X_{13}^{x_{13}} \dots X_{n-1,n}^{x_{n-1,n}}, \quad (5)$$

式中： X 为出行总量； X_{ij} 为 i 站到 j 站的OD分布量； x_{ij} 为 i 站到 j 站的先验OD分布量。

$\max W(X_{ij})$ 等价于 $\max \ln W(X_{ij})$ ，令 $S = \ln W(X_{ij})$ ，借助于stirling逼近公式： $\ln X! = X \ln X - X$ ，则

$$S = \ln(X!) - \sum_i \sum_j (X_{ij} \ln \frac{X_{ij}}{X_j} - X_{ij})。 \quad (6)$$

忽略常数项 $\ln(X!)$ ，式(6)等价于以下极大值问题：

$$\begin{cases} \max S' = - \sum_i \sum_j (X_{ij} \ln \frac{X_{ij}}{X_j} - X_{ij}) \\ \text{s.t. } V^k = \sum_i \sum_j X_{ij} p_{ij}^k, k=1, 2, \dots, 2(n-1) \end{cases}。 \quad (7)$$

式中： V^k 为方程组(4)中矩阵 V 的第 k 项； p_{ij}^k 为方程组(4)中 V^k 对应的方程组系数。

由于目标函数的形式类似于热力学系统熵的表达式，故上述模型被称为极大熵模型。

构造拉格朗日函数：

$$L(X, \mu) = - \sum_i \sum_j (X_{ij} \ln \frac{X_{ij}}{X_j} - X_{ij}) + \sum_k \mu^k (V^k - \sum_i \sum_j p_{ij}^k X_{ij}), \quad (8)$$

式中 μ^k 为拉格朗日算子，根据函数取极值的一阶条件

$\frac{\partial L(X, \mu)}{\partial X_{ij}} = 0$ 可求得：

$$X_{ij} = x_{ij} \exp[- \sum_k \mu^k p_{ij}^k]。 \quad (9)$$

令 $\lambda^k = \exp(-\mu^k)$ ，则式(9)表示为：

$$X_{ij} = x_{ij} \prod_k (\lambda^k)^{p_{ij}^k}。 \quad (10)$$

如果极大熵模型中不包含先验OD矩阵信息，那么可构造一个极大熵模型：

$$\begin{cases} \max S' = - \sum_i \sum_j (X_{ij} \ln X_{ij} - X_{ij}) \\ \text{s.t. } V^k = \sum_i \sum_j X_{ij} p_{ij}^k, k=1, 2, \dots, 2(n-1) \end{cases}。 \quad (11)$$

这时模型的解为：

$$X_{ij} = \prod_k (\lambda^k)^{p_{ij}^k}。 \quad (12)$$

3 模型求解

将式(10)代入式(7)可得

$$V^k = \sum_i \sum_j [x_{ij} \prod_k (\lambda^k)^{p_{ij}^k}] p_{ij}^k。 \quad (13)$$

式(13)中只有 $\{\lambda^k\}$ 的值未知。实际上，式(13)表示一个包含 k 个未知数 $\{\lambda^k\}$ 和 k 个等式的非线性方程组。考虑到这是一个典型的“multi-proportional”问题，可以用以下的迭代算法求解。算法中需要引入辅助变量 y^m 。具体的求解步骤为：

- ① 置迭代次数 $n=0$ ，令 $\{X_{ij}^{(n)}\} = \{x_{ij}\}$ ， $\lambda^{(n)k} = 1, k=1, 2, \dots, K$ ；
- ② 令 $m=0$ ；
- ③ 置 $m=m+1$ ；
- ④ 用二分法求解下列方程中的 y^m ：

$$V^m = \sum_i \sum_j [x_{ij} \prod_k (\lambda^k)^{p_{ij}^k}] p_{ij}^m (y^m)^{p_{ij}^m}； \quad (14)$$
- ⑤ 用 $\lambda^{(n+1)m} = \lambda^{(n)m} y^m$ 更新 $\lambda^{(n)m}$ ；
- ⑥ 计算 $X_{ij}^{(n+1)} = x_{ij} \prod_k (\lambda^{(n+1)k})^{p_{ij}^k}$ ；
- ⑦ 若 $m < K$ ，返回③；否则，转向⑧；

$$\textcircled{8} \text{ 计算 } V_n^k = \sum_i \sum_j X_{ij}^{(n)} p_{ij}^k,$$

若对于全部的 k 均满足 $\frac{|V_{(n)}^k - V^k|}{V^k} \leq \sigma$,

转向 $\textcircled{10}$; 否则, 转向 $\textcircled{9}$;

$\textcircled{9}$ 若 $n < N$, 令 $n = n + 1$, 转向 $\textcircled{2}$; 否则, 转向 $\textcircled{10}$;

$\textcircled{10}$ 输出OD矩阵 $\{X_{ij}^{(n)}\}$, 过程结束。

上述过程中, σ 为一小值, 可设为 $10e^{-6}$ 。N为预设的最大迭代次数, 可设为200次。

4 先验OD矩阵的确定

居民公交出行具有一定的稳定性, 在选择公交出行时, 大部分的居民会选择固定时间、固定线路出行, 尤其在早晚高峰时。同时, 居民公交出行也具有相当大的随机性, 很难用一定的规律来完美地表达, 因此有原始资料时应尽量以原始资料为先验OD, 在缺乏原始资料时, 根据上述对居民公交出行规律的分析, 先验OD矩阵可采用重力模型的形式: $x_{ij} = U_j \frac{T_i W_i P_{ij}}{\sum_i T_i W_i P_{ij}}$ 。

5 实例分析

以哈尔滨市八路公交线路20个运营车次的后4个车次为研究对象, 利用公交乘客上下车人数, 分别计算以前6个运营车次为先验OD和采用重力模型先验OD的公交线路客流OD矩阵。

引入预测误差指数指标

$$EC = 1 - \frac{\sqrt{\sum_i \sum_j (R_{ij} - X_{ij})^2}}{\sqrt{\sum_i \sum_j R_{ij}^2} + \sqrt{\sum_i \sum_j X_{ij}^2}}, \quad (20)$$

式中: R_{ij} 表示 i 站到 j 站的实际OD分布量; X_{ij} 表示 i 站到 j 站的推算OD分布量。该预测误差指数指标表示预测值与实测值之间的拟合度, 数值在(0,1), 数值越大表示拟合的效果越好。EC的大小充分反映了预测效果的好坏, EC值大于0.85的预测视为较好的预测, EC值大于0.9的预测可视为满意的预测。EC值越高, 则整体预测效果越与实际测量值接近, 效果也越理想。

将推算得到的公交线路客流OD矩阵与实测的公交线路客流OD矩阵进行比较并检验其误差, 结果见表2。

表2 推算OD矩阵与实测OD矩阵的误差比较表

Tab.2 Errors comparison between estimation OD matrix and actual OD matrix

	上行		下行	
	以原始数据为先验OD	采用重力模型形式的先验OD	以原始数据为先验OD	采用重力模型形式的先验OD
EC	0.900	0.868	0.887	0.845

上述结果说明两种方法计算精度都较好, 方法可靠, 但以原始数据为先验OD的推算结果误差要小于采用重力模型形式的先验OD的推算结果。因此, 有原始数据时应尽量使用原始数据作为先验OD。

6 结语

本文在公交线路客流OD调查的基础, 分析了居民公交出行规律, 发现居民公交出行规律主要受公交出行距离和站点用地性质影响, 在引入吸引权系数表示公交站点用地性质影响后, 发现公交乘客乘行站数概率符合泊松分布。建立了根据公交乘客上下车人数推算公交线路客流OD矩阵的极大熵模型, 并根据居民公交出行规律确定了缺乏原始资料情况下的先验OD的重力模型形式, 分别采用原始资料作为先验OD和重力模型形式的先验OD, 用哈尔滨市八路公交线路的客流OD调查数据验证了模型, 结果证明该模型精度较高。

参考文献

- 1 陈森发, 周振国, 于栋华.一种动态 OD 矩阵估计算法的理论及应用 [J].东南大学学报(自然科学版), 2003, 33 (1): 106~110
- 2 H.P.Lo, N.Zhang, W.H.K.Lam.Decomposition Algorithm for Statistical Estimation of OD Matrix with Random Link Choice Proportions form Traffic Counts [J].Transportation Research (PartB), 1999, (33): 369~385
- 3 S.C.Wong, C.O.Tong.Estimation of Time-Dependent Origin-Destination Matrices for Transit Networks [J]. Transportation Research (PartB), 1998, 32 (1): 35~48
- 4 尹娟, 郭国会.神经网络在交通网络 OD 矩阵推算中的应用 [J].浙江大学学报(工学版), 2000, 34 (4): 423~427