

DOI:10.16799/j.cnki.csdqyfh.2022.11.055

基于大数据的出行方式分配预测研究

潘健英¹, 周侃², 周琳²

(1.深圳市地铁集团有限公司, 广东 深圳 518038; 2.深圳市市政设计研究院有限公司, 广东 深圳 518029)

摘要: 随着城镇化及城市交通机动化的进程加快,在城镇化和机动化的共同作用下,交通拥挤问题已经成为制约城市发展和影响居民生活质量的主要问题之一。居民出行行为研究在交通管理和控制中有着非常重要的作用。结合多元大数据,构建出行方式选择的概率模型,有助于不同小区的出行方式分担率的预测,对缓解城市交通压力、提升运营管理能力有直接作用。基于小区出行分配、公共交通和非公共交通以及公交和轨道出行三个层级,分别得出出行分配预测模型,并形成计算机程序。利于实际应用,对识别城市出行特征、缓解城市交通压力提供有效措施。

关键词: 出行方式;大数据;预测

中图分类号: U491.1

文献标志码: A

文章编号: 1009-7716(2022)11-0205-04

0 引言

经济社会的不断发展在改善城市生活面貌的同时也给城市交通带来了越来越大的压力。随着城镇化及城市交通机动化的进程加快,在城镇化和机动化的共同作用下,交通拥挤问题已经成为制约城市发展和影响居民生活质量的主要问题之一。

出行者作为交通出行的主体,其出行方式选择行为对交通组织、运营及管理有显著影响。居民出行行为研究在交通管理和控制中有着非常重要的作用。对城市居民个体出行行为的研究,是城市规划、建设、管理中一项不可缺少的基础性工作。居民的出行方式选择通过改变城市交通客运结构来影响城市交通系统的运行情况^[1]。

居民出行方式选择的基本过程是:居民由于自身的需要产生出行需求,然后结合自己的社会经济条件、历史经验、交通运输的历史供给情况等,确定适合自己的运输方式集合以及预期各种运输方式会产生的效用,在此基础上结合出行目的建立选择标准,最终确定最适合自己的运输方式。居民个体出行方式的选择实质是出行者一个不断学习的过程所表现出来的结果。在交通出行方式的选择中,出行者根据以往的出行经验,再结合当下的实际情况,在充分考虑出行时间、路线和费用的情况下,对出行方式做出选择。城市居民个体出行方式的选择是一个

反馈的过程,先前的经验对后面的选择产生影响,后面的选择又会形成新的经验。

因此,通过结合多元大数据,如手机信令数据和刷卡数据、互联网数据和 GIS 数据的分析,获得小区之间不同出行方式、小区的基础属性数据以及小区之间不同出行方式的分担率,进而构建出行方式选择的概率模型,有助于不同小区的出行方式分担率的预测,对提升出行方式分担率预测的准确性有着积极的意义。

1 出行方式分担率的预测方法

通过对手机信令数据和深圳通刷卡数据^[2]的分析,得到现状小区之间不同出行方式的分担率数据;通过对手机信令数据、深圳通刷卡数据、互联网数据和 GIS 数据的分析,得到小区之间不同出行方式的出行数据以及小区的基础属性数据;构建出行方式选择的概率模型,并利用现状数据标定模型参数;利用参数已经标定好的概率模型和规划年的相关基础数据,对规划年小区之间不同出行方式的分担率进行预测。其技术路线如图 1 所示。

1.1 第一阶段

第一阶段构建非机动化分担率与小区间出行距离、小区间出行时间、平均房价、住房租金、无固定工作居住人口占总居住人口比例之间的第一回归模型。其特征在于所述小区之间不同出行方式的出行数据,包括:小区间的平均出行时间、小区间的平均出行距离、小区间不同出行方式的平均出行费用和小区间不

收稿日期: 2022-06-17

作者简介: 潘健英(1971—),女,工学学士,高级工程师,从事轨道交通科技创新研究工作。

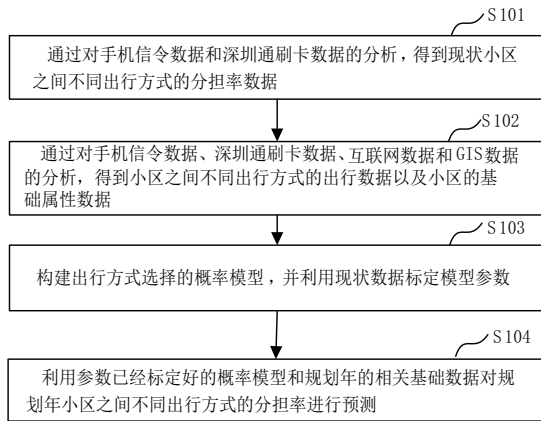


图1 出行方式分配预测技术路线

同出行方式的平均出行时间。所述不同出行方式包括公交、非公交、轨道和常规公交。

其中,小区间的平均出行时间、小区间的平均出行距离、小区间的非公交平均出行时间和小区间的非公交平均出行费用,是根据网络爬取的驾车出行数据计算得到的;小区间的公交平均出行时间、小区间的公交平均出行费用,是根据网络爬取的公交出行数据计算得到的;小区间的轨道平均出行时间、小区间的轨道平均出行费用、常规公交的平均出行时间和常规公交的平均出行费用,是根据网络爬取的公交出行数据筛选出的轨道和常规公交数据计算得到的。

$$P_{ijT} = a_0 + a \times (s_{ij}^{0.2}) + b \times (t_{ij}^{0.4}) + c \times (\text{price}_i^{0.25}) + d \times (\text{rent}_i^{0.2}) + e \times (\text{ratio}_i^2)$$

式中: P_{ijT} 表示*i*小区到*j*小区的出行中选择非机动化工具出行的概率; s_{ij} 表示*i*小区到*j*小区的平均出行距离; t_{ij} 表示*i*小区到*j*小区的平均出行时间; price_i 表示*i*小区的平均房价; rent_i 表示*i*小区的租金; ratio_i 表示*i*小区的无固定工作居住人口占总居住人口的比例;其他为模型待标定参数。

1.2 第二阶段

第二阶段构建公交和非公交的二元 Logit 模型,具体是指公交分担率与小区之间的公交平均出行时间、公交平均出行费用、非公交平均出行时间、非公交平均出行费用、起点小区的公交站点总数、起点小区的轨道站点数量以及终点小区的轨道站点数量之间的第二回归模型。其中,所述非机动化分担率和轨道分担率数据通过对手机信令数据分析得到,所述公交分担率数据通过对手机信令数据和深圳通刷卡数据综合分析得到。

所述小区的基础属性数据包括平均房价、住房租金、无固定工作居住人口占总居住人口比例、常规

公交站点数和轨道站点数^[3]。其中,所述无固定工作居住人口占总居住人口比例,是依据手机信令数据识别出的居住人口数和岗位数计算得到的;所述平均房价和住房租金数据,是利用对网络爬取得到的房价和租金的点位置信息数据以及 GIS 数据的综合分析得到的;所述轨道站点数和常规公交站点数,是利用 GIS 数据得到的。

第二回归模型的具体形式为:

$$P_{ijT} = \frac{e^{vT_{ij}}}{e^{vT_{ij}} + e^{vA_{ij}}}$$

$$vT_{ij} = a \times t_{ij}^{bus} + b \times \text{cost}_{ij}^{bus} + c \times \text{stop}_{ij}^{bus} + d \times \text{stop}_{i}^{rail} + e \times \text{stop}_{j}^{rail}$$

$$vA_{ij} = a_0 + a \times t_{ij}^{not-bus} + b \times \text{cost}_{ij}^{not-bus}$$

式中: P_{ijT} 表示*i*小区到*j*小区的出行中选择公交出行的概率; vT_{ij} 表示公交系统的效用函数; vA_{ij} 表示非公交系统的效用函数; t_{ij}^{bus} 表示*i*小区到*j*小区的公交平均出行时间; cost_{ij}^{bus} 表示*i*小区到*j*小区的公交平均出行费用; stop_{ij}^{bus} 表示*i*小区和*j*小区的常规公交站点总数; stop_{i}^{rail} 表示*i*小区的轨道站点总数; stop_{j}^{rail} 表示*j*小区的轨道站点总数; $t_{ij}^{not-bus}$ 表示*i*小区到*j*小区的非公交平均出行时间; $\text{cost}_{ij}^{not-bus}$ 表示*i*小区到*j*小区的非公交平均出行费用;其他为模型待标定参数。

1.3 第三阶段

第三阶段构建轨道和常规公交的二元 Logit 模型,具体是指轨道分担率与小区之间的轨道平均出行时间、轨道平均出行费用、常规公交平均出行时间、常规公交平均出行费用、起点小区的轨道站点数量、终点小区的轨道站点数量、起点小区的常规公交站点数量、终点小区的常规公交站点数量之间的第三回归模型。其中,所述非机动化分担率和轨道分担率数据是通过对手机信令数据分析得到。

第三回归模型的具体形式如下:

$$P_{ijT} = \frac{e^{vT_{ij}}}{e^{vT_{ij}} + e^{vA_{ij}}}$$

$$vT_{ij} = a \times t_{ij}^{rail} + b \times \text{cost}_{ij}^{rail} + c \times \text{stop}_{i}^{rail} + d \times \text{stop}_{j}^{rail}$$

$$vA_{ij} = a \times t_{ij}^{bus} + b \times \text{cost}_{ij}^{bus} + c \times \text{stop}_{i}^{bus} + d \times \text{stop}_{j}^{bus}$$

式中: P_{ijT} 表示*i*小区到*j*小区的出行中选择轨道出行的概率; vT_{ij} 表示常规公交系统的效用函数; vA_{ij} 表示轨道系统的效用函数; t_{ij}^{rail} 表示*i*小区到*j*小区的轨道平均出行时间; cost_{ij}^{rail} 表示*i*小区到*j*小区的轨道平均出行费用; stop_{i}^{rail} 表示*i*小区的轨道站点总数; stop_{j}^{rail} 表示*j*小区的轨道站点总数; t_{ij}^{bus} 表示*i*小区到

j 小区的常规公交平均出行时间; $cost_{ij}^{bus}$ 表示*i*小区到*j*小区的常规公交平均出行费用; $stop_{i}^{bus}$ 表示*i*小区的常规公交站点总数; $stop_{j}^{bus}$ 表示*j*小区的常规公交站点总数。

通过对手机信令数据和深圳通刷卡数据的分析,得到现状小区之间不同出行方式的分担率数据。利用对手机信令数据、深圳通刷卡数据、互联网数据和GIS数据的分析,得到小区之间不同出行方式的出行数据以及小区的基础属性数据;构建模型单元,基于上述数据,构建出行方式选择的概率模型,并利用现状数据标定模型参数。进而,根据参数已经标定好的概率模型和规划年的相关基础数据对规划年小区之间不同出行方式的分担率进行预测^[4]。

表1 标定第一回归模型的部分现状数据

i 小区编号	j 小区编号	非机动车化分担率	出行时间 /min	出行距离 /m	无固定工作人口占比	i 小区平均房价 /(元·m ²)	i 小区每月平均租金 /(元·m ⁻²)
1	2	0.714 3	5.683 33	1 293	0.644 209 00	75 703	115
1	4	0.857 1	4.883 33	1 687	0.644 209 00	75 703	115
1	5	0.833 3	7.166 66	2 955	0.644 209 00	75 703	115
1	6	0.754 5	5.316 66	1 196	0.644 209 00	75 703	115
1	7	0.800 0	6.850 00	2 170	0.644 209 00	75 703	115
1	8	1.000 0	6.283 33	2 104	0.644 209 00	75 703	115
1	9	0.625 0	8.866 66	3 401	0.644 209 00	75 703	115
1	11	0.666 7	9.233 33	3 529	0.644 209 00	75 703	115

表2 标定第二回归模型的部分现状数据

i 小区编号	j 小区编号	公交分担率	公交平均出行时间 /min	公交平均出行费用 /元	公交站点数	i 小区轨道站点数	j 小区轨道站点数	非公交平均出行时间 /min	非公交平均出行费用 /元
5	56	0.106 6	59.146 6	1.8	5	0	0	15.450 0	20
32	3 734	0.699 5	96.663 3	7.4	3	0	0	44.250 0	68
37	1 814	0.048 4	143.636 0	1.4	6	0	0	52.666 6	83
13	103	0.029 5	45.340 0	1.6	6	1	0	22.166 6	29
26	3 978	0.334 0	85.253 3	4.6	9	0	0	34.533 3	65
4	379	0.181 2	83.593 3	2.6	11	0	0	34.933 3	76

根据构建的第二回归模型的具体公式和现状数据,使用R语言标定模型的相关参数,可以得到第二回归模型的公式为:

$$P_{ijT} = \frac{e^{vT_{ij}}}{e^{vT_{ij}} + e^{vA_{ij}}}$$

$$vT_{ij} = -0.006 29 \times t_{ij}^{bus} - 0.000 000 62 \times cost_{ij}^{bus} + 0.005 91 \times stop_{ij}^{bus} + 0.223 \times stop_{i}^{rail} + 0.319 \times stop_{j}^{rail}$$

$$vA_{ij} = -3.52 - 0.020 5 \times t_{ij}^{not_bus} - 0.000 167 \times cost_{ij}^{not_bus}$$

2.3 第三阶段

标定第三回归模型的部分现状数据见表3。

根据构建的第三回归模型的具体公式和现状数据,使用R语言标定模型的相关参数,可以得到第三

2 预测方法应用

2.1 第一阶段

标定第一回归模型的部分现状数据见表1。

根据构建的第一回归模型的具体公式和现状数据,使用R语言标定模型的相关参数,可得第一回归模型的公式为:

$$P_{ijT} = 1.686 629 - 0.209 793 2 \times (s_{ij}^{0.2}) - 0.019 769 9 \times (t_{ij}^{0.4}) - 0.005 963 016 \times (price_i^{0.25}) + 0.373 041 9 \times (rent_i^{0.2}) + 0.018 508 4 \times (ratio_i^2)$$

第一回归模型的R²为0.81。

2.2 第二阶段

标定第二回归模型的部分现状数据见表2。

回归模型的具体公式为:

$$P_{ijT} = \frac{e^{vT_{ij}}}{e^{vT_{ij}} + e^{vA_{ij}}}$$

$$vT_{ij} = -1.38 - 0.000 326 \times t_{ij}^{rail} - 0.000 003 03 \times cost_{ij}^{rail} + 0.037 9 \times stop_{i}^{rail} + 0.075 3 \times stop_{j}^{rail}$$

$$vA_{ij} = -0.000 529 \times t_{ij}^{bus} - 0.011 4 \times cost_{ij}^{bus} + 0.026 2 \times stop_{i}^{bus} + 0.022 \times stop_{j}^{bus}$$

2.4 预测结果

利用参数已经标定好的概率模型和规划年的相关基础数据,对规划年小区之间不同出行方式的分担率进行预测。

表4为预测得到的小区间分担率的部分数据。

表3 标定第三回归模型的部分现状数据

i 小区编号	j 小区编号	轨道分担率	轨道平均出行时间/min	轨道平均出行费用/元	i 小区轨道站点数	j 小区轨道站点数	常规公交出行时间/min	常规公交出行费用/元	i 小区公交站点数	j 小区公交站点数
6	707	0.500 0	74.316 6	8	0	1	110.700 0	0	23	8
16	22	0.027 3	22.166 6	2	0	1	31.433 3	2	2	1
16	532	0.333 3	56.816 6	5	0	0	96.516 6	0	2	0
17	1 214	0.666 6	96.300 0	8	0	0	120.266 0	0	4	7
27	49	0.076 9	24.350 0	2	0	0	49.250 0	3	8	4
27	339	0.230 7	54.083 3	6	0	0	72.366 6	3	8	11
31	387	0.090 9	42.250 0	4	0	0	50.683 3	2	2	0
39	121	0.032 2	28.533 3	2	0	0	32.300 0	2	5	4

表4 预测得到的小区间分担率的部分数据

i 小区编号	j 小区编号	非机动化分担率	公交分担率	轨道分担率
6	705	0.09	0.52	0.32
16	22	0.42	0.18	0.05
16	532	0.06	0.47	0.28
17	1 213	0.13	0.32	0.17
27	49	0.11	0.36	0.12
27	339	0.07	0.51	0.23

2.5 计算程序形式

图2为一种出行方式分担率预测装置的示意性框图。该预测装置200用于执行前述的基于大数据的出行方式分担率的预测方法。其中,该预测装置可以配置于服务器或终端。如图2所示,该预测装置200包括:第一分析单元201、第二分析单元202、构建模型单元203和预测单元204。该装置可以实现为一种计算机程序的形式。该计算机程序可以在计算设备上运行。

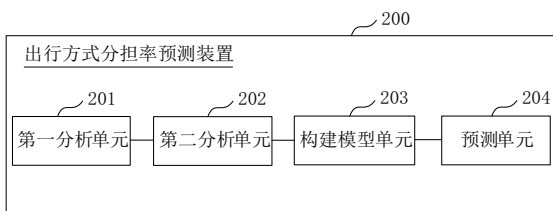


图2 计算机程序应用过程

第一分析单元,用于通过对手机信令数据和深圳通刷卡数据的分析,得到现状小区之间不同出行方式的分担率数据;第二分析单元,用于通过对手机信令数据、深圳通刷卡数据、互联网数据和GIS数据的分析,得到小区之间不同出行方式的出行数据以

及小区的基础属性数据;构建模型单元,用于构建出行方式选择的概率模型,并利用现状数据标定模型参数;预测单元,用于利用参数已经标定好的概率模型和规划年的相关基础数据,对规划年小区之间不同出行方式的分担率进行预测。

3 结论

本文基于城市大数据,对出行方式分担率进行预测。构建出行方式选择的概率模型,并利用现状数据标定模型参数。最后利用参数已经标定好的概率模型,预测小区之间不同出行方式的分担率,并预测出公共交通方式以及非公共交通方式的分配。进一步预测出在公共交通中,公交和轨道交通的分配,并按照计算机程序形式,构建预测过程。对判断城市交通出行特征,缓解城市交通压力具有实际意义。该方法已实际应用并申请相关专利^[5]。

参考文献:

[1] 刘小艳,城市居民出行方式选择行为及影响因素分析[J].交通企业管理,2012(6):52-53.
 [2] Tu, W., et al. Spatial variations in urban public ridership derived from GPS trajectories and smart card data [J]. Journal of Transport Geography, 2018(69):45-57.
 [3] An, D., et al. Understanding the impact of built environment on metro ridership using open source in Shanghai [J]. Cities, 2019(93): 177-187.
 [4] 关桢,社区公交发展模式研究[C]//城乡治理与规划改革——2014年中国城市规划年会论文集.北京:中国建筑工业出版社,2014: 221-228.
 [5] 深圳市市政设计研究院有限公司.一种基于大数据的出行方式分担率的预测方法和预测装置:中国,CN201911252177.X [P]. 2020-05-12.