

区域出行需求模型构建关键技术 ——以南京都市圈为例

耿天霜¹, 孟丹青¹, 戴骏晨², 凌小静²

(1.南京地铁建设有限责任公司, 江苏 南京 210008; 2.中咨城建设计有限公司江苏分公司, 江苏 南京 210012)

摘要: 区域出行在频率和规律性方面显著区别于城市出行, 需构建针对性的出行需求模型, 而该类模型相较于传统方法的具体提升效果及低成本实现路径尚缺乏充分实证研究。以南京都市圈为例, 基于手机信令和小样本调查数据, 重点针对出行生成与出行分布两个阶段开展建模方法研究, 并通过划分训练集与测试集进行模型验证与对比分析。提出一种融合小样本调查数据与出行大数据的建模流程: 首先通过调查数据标定离散选择模型, 再借助手机信令数据校正泊松回归模型, 在保证精度的同时保留对个人特征的解释力。研究表明: 未引入个人特征及起终点交互变量时, 目的地选择模型性能甚至弱于重力模型; 而加入交互变量后, 其表现显著优于重力模型, 嵌套Logit模型中Nest结构的合理构建可进一步优化效果; 所提出的方法在出行生成和分布阶段精度较传统模型分别提升24%和18%。

关键词: 区域出行需求; 离散选择模型; 出行生成; 出行分布; 出行频率; 目的地选择; 交互效应; 南京都市圈

Key Technologies for Constructing Regional Travel Demand Models: A Case Study of the Nanjing Metropolitan Area

GENG Tianshuang¹, MENG Danqing¹, DAI Junchen², LING Xiaojing²

(1. Nanjing Metro Construction Co., Ltd., Nanjing Jiangsu 210008, China; 2. Jiangsu Branch, CIECC Urban Construction Design Co., Ltd., Nanjing Jiangsu 210012, China)

Abstract: Regional travel differs substantially from urban travel in terms of frequency and regularity, thus requiring a specialized travel demand model. However, current empirical research is insufficient regarding the specific improvements and cost-effective implementation pathways of such models compared to conventional modeling approaches. Using the Nanjing Metropolitan Area as a case study, this paper focuses on modeling methodology research on the trip generation and trip distribution stages based on mobile signaling data and small-sample survey data. Model validation and comparative analysis are conducted by dividing the data into training and testing datasets. The paper proposes a modeling framework that integrates small-sample survey data with large-scale travel data: a discrete choice model is first calibrated using survey data, followed by a Poisson regression model corrected through mobile signaling data, thereby ensuring the accuracy while preserving explanatory power with respect to individual attributes. Findings indicate that, when individual attributes and origin-destination interaction variables are excluded, the performance of the destination choice model is weaker than that of gravity models. After incorporating interaction variables, however, the model significantly outperforms gravity models, and a reasonable construction of nesting structures within nested Logit models can further enhance the results. The proposed approach improves the modeling accuracy in trip generation and distribution stages by 24% and 18%, respectively, compared with traditional models.

Keywords: regional travel demand; discrete choice model; trip generation; trip distribution; travel frequency; destination choice; interaction effects; Nanjing Metropolitan Area

收稿日期: 2025-03-14

基金项目: 南京地铁集团有限公司科研项目“都市圈城际轨道交通客流预测关键技术研究”(JS-D.S05.X-XY01-00-2211-0332)

作者简介: 耿天霜(1980—), 男, 江苏南通人, 学士, 高级工程师, 总工办副主任, 研究方向为地铁建设与管理, 电子邮箱1219801024@qq.com。

通信作者: 戴骏晨(1990—), 男, 江西新余人, 硕士, 工程师, 研究方向为交通模型、数据科学, 电子邮箱djesse@foxmail.com。

0 引言

都市圈是城市群内部以超(特)大城市或辐射带动功能强的大城市为中心、以1小时通勤圈为基本范围的城镇化空间形态。建设现代化都市圈是推进新型城镇化的重要手段,需着力打造轨道上的都市圈^[1]。实现这一目标,需科学把握都市圈范围内的出行需求,即构建相应的区域出行需求模型。

欧美国家通常由官方机构建立并维护州(省)级乃至全国范围的出行需求模型。例如,美国由大都会规划组织(Metropolitan Planning Organization, MPO)负责建立各州的出行需求模型^[2],瑞典则由交通局管理 Sampers 模型并长期用于项目评估^[3]。而中国除个别发达地区外,出行需求模型多由规划设计单位按城市分别建立,缺乏更高层次的统筹。这一现状导致都市圈层面的客流预测建模通常仍沿用基于单一城市的建模框架,仅将模型范围扩展至研究对象区域。然而,区域出行具有明显的跨城特征,其出行频率远低于城市内部出行,行为规律也更难把握。因此,有必要建立适应性更强的区域出行需求模型,并对其优势开展深入研究。此外,模型构建还需考虑现实成本与可行性。本文以南京都市圈为例,开展区域出行需求模型构建关键技术研究,提出一种经证实有效的低成本建模方法,并明确区域出行需求模型相较于传统模型的关键优势,弥补了该领域现有研究的不足。

1 区域出行需求模型研究现状

尽管基于活动的模型是当前最先进的模型,但美国MPO出于实用性因素考虑,仍普遍使用传统的四阶段模型。K. Park等^[4]调查的25个MPO几乎全部采用传统的四阶段模型,其中20个仍使用重力模型进行交通分布预测。重力模型并不符合第一性原理,即它不是从最基本假设或经验观察出发并通过逻辑推理或数学推导得出,而是通过形式类比万有引力定律构建的^[5]。在城市交通中,居住与就业对应的出行发生和吸引相互关联,高频的通勤出行占据主导地位,且出行强度随距离和成本增加而衰减,重力模型反映了该现象;加之重力模型应用便捷,使其在城市内部交通建模中具有广泛的实用价值。然而,对于城市间的低频出行,重力模

型对城际出行复杂成因的解释力可能不足。

因此,对于长距离出行,各国通常单独建模。不同国家对长距离出行的界定阈值存在差异:以单程出行距离计算,美国和英国约为80.5 km,瑞典约为100 km,荷兰约为50 km^[6]。根据美国交通部2002年全国家庭出行调查,长距离出行中个人出行占59%,商业出行占28%,通勤出行占13%^[7]。2015年英国出行调查数据显示:每人每周平均出行17.2次,其中长距离出行仅0.4次,且收入越高,长距离出行频率越高^[8]。长距离出行模型通常采用基于随机效用理论的离散选择模型。例如M. Bradley等^[9]建立了覆盖全美的实用型长距离出行模型,包括机动车拥车情况模型、目的地-出行方式选择模型、出行生成模型等,并通过人口合成技术生成离散数据以支持模型应用。在长距离出行模型的标定中,通过居民出行调查获得足够样本非常困难且成本极高,因此实践中很少采用——在过去20年中,美国各地仅成功开展过5次具备全国代表性样本的有效调查。目前常需借手机信令等大数据进行补充与校正^[10]。Zhang D.等^[11]则通过SP问卷调查,建立了针对城际交通方式的诱增客流模型。

在国际实践中,轨道交通的客流需求预测常常被高估。B. Flyvbjerg等^[12]指出,尽管建模技术在进步,但轨道交通需求高估的问题比公路更为突出,其原因并非模型不可靠,而是受到政治压力影响。因此,开发可靠且透明(模型结构及参数公开)的出行需求模型,不仅有利于交通学术界,也符合公众利益^[11]。在中国,雷磊^[13]较早对区域出行需求模型进行了系统研究,丘建栋等^[14]则开发了应用于广东省规划设计实践的区域交通模型。

2 研究内容及方法

2.1 研究内容

区域出行需求模型与城市交通模型的主要差异体现在建模方法上,具体表现为四阶段模型中出行生成与出行分布两个阶段所采用方法的不同。区域出行需求模型不再局限于城市交通模型中常用的回归分析或重力模型,而是转向基于随机效用理论的离散选择模型。因此,本文重点研究区域出行需求模型在出行生成与出行分布阶段的建模方法。

2.2 区域出行界定

本文所界定的区域出行，主要指区别于城市内部出行的跨城出行。在都市圈范围内，该界定不包含进出都市圈的出入境出行及跨都市圈的过境出行。此外，虽然通勤圈周边乡镇街道的通勤出行以及相邻区域的短距离出行也跨越了行政区边界，但由于其属于日常出行范畴，性质上仍等同于城市内部出行，因此也不在本文的界定范围。都市圈

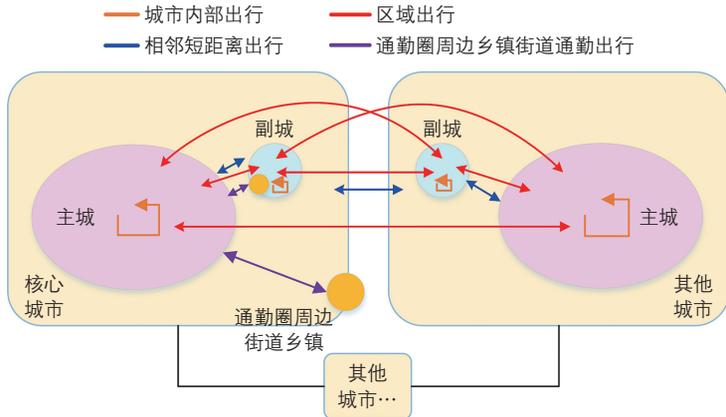


图1 都市圈区域出行界定示意
Fig.1 Regional travel definition in the metropolitan area



图2 南京都市圈区域单元划分示意
Fig.2 Regional unit division in the Nanjing Metropolitan Area

区域出行界定示意图1。

尽管图1中主城和副城的范围可能与实际行政区划有所不同，但为便于实践应用和数据收集，仍将主城界定为地级市的市区，副城界定为所辖县或县级市(通常县或县级市相对独立，其与市区之间的出行规律已不同于日常通勤出行)，并统称为区域单元。若某市仅由市辖区组成，则将其相对独立的区划定为副城。以南京都市圈为例：核心城市南京的副城包括溧水和高淳；都市圈内其他地级市的主城为市区，副城为所辖县或县级市。南京都市圈区域单元划分示意图2(同颜色表示同一区域单元)。

手机信令数据显示，相邻区域单元之间通常存在大量跨边界的短距离出行。如前所述，此类出行属于日常出行，不应计入区域出行。理想情况下，应基于出行频率对区域出行进行划分，但由于数据资源有限，且出行频率通常与距离呈负相关，为便于操作，本文采用距离阈值进行划定。同时，考虑到区域出行需求模型的重要应用之一是跨区域轨道交通客流预测——其中市域(郊)铁路平均站间距约为3 km，城际铁路站间距更大，结合主要乘客的出行距离特征及区县级模型精度等因素，最终将阈值确定为5 km。基于手机信令数据统计，跨区域单元的相邻区县间出行中，距离小于5 km的出行比例约为40%，将此类出行与区域出行加以区分，有利于开展更具针对性的建模工作。因此，本文将区域出行具体界定为都市圈内部出行距离大于5 km并跨区域单元的非通勤出行。通勤圈周边乡镇街道的通勤出行以及相邻区域间的短距离出行可单独建模，但本文不对其展开论述。

2.3 研究方法 with 数据

分别采用基于随机效用理论的离散选择模型和传统回归分析方法，对区域出行生成与出行分布进行建模，以获取分区县的人均区域出行次数和出行分布比例。利用训练数据集进行参数标定，并通过测试数据集检验模型效果，以此进行模型对比，并探索优化模型性能的途径。

由于区域出行需求模型所需数据覆盖空间范围大，在此范围内开展全面居民出行调查难度大、成本高，因此通常以大数据为主要数据来源。本文采用符合区域出行定义在南京都市圈手机信令数据，具体为区县级工

作日日均出行量,仅包含去程数据(不含回程),且无与出行关联的个人特征数据。

离散选择模型的标定需要个人特征数据,尤其在出行生成阶段。为此,本研究在南京都市圈范围内开展了小样本问卷调查,共获取1163份有效问卷,内容包括个人社会经济特征(如性别、年龄、职业、收入、拥车情况及常住地址)及出行特征(如都市圈内区域出行频率、出行目的、目的地、出行方式及成本等),用于支持离散选择模型的标定。具体调查方式为:以网络问卷形式在公立机构官方微博、微信公众号等渠道征集受访者,并对出行方式、时间、费用等信息逐项校验,严控数据质量。

3 区域出行生成

各区县基于手机信令数据的年人均区域出行次数(去程)见图3。该数据将作为真实值,分别用于模型的标定(训练集)与效果检验(测试集)。

3.1 线性回归模型

1) 变量选取。

线性回归模型是城市交通模型在出行生成阶段的常用方法,本文采用该方法对人均区域出行次数进行建模。模型的因变量为各区县的年人均区域出行次数(去程)。候选自变量包括各区县的第三产业GDP、第三产业GDP、总GDP、常住人口、年旅游接待人数、人均GDP、所在市相邻市的数量、相邻区域单元数量、所在城市GDP、所在区域单元GDP、所在市主城区区域单元GDP(非主城区则为0,以下简称 GDP_c)、副城与主城GDP乘积、目的地选择模型Root层的Logsum(以下简称 L_d),以及副城至主城方式划分Root层的Logsum(以下简称 L_s)。

2) 关键变量定义与建模逻辑说明。

副城与主城GDP乘积用于反映同一地级行政区内副城与主城之间的向心作用,若该区县本身为主城,则取值为0。Logsum为Logit模型中所有子选择肢效用期望值的对数,可用于表征综合效用。 L_d 表示所有目的地对出行者或出发地的综合吸引力,其值来源于区域出行分布模型的标定结果。因此,在模型标定过程中,实际顺序为先进行分布模型标定,再进行生成模型标定;但为

符合应用中的逻辑顺序,本文仍按“先生成、后分布”的顺序进行论述。 L_s 用于表征副城至本市主城的综合出行便捷程度,体现同一市域范围内副城与主城的联系强度(若区县属于主城区,则 L_s 取最小值)。 L_s 的计算基于出行方式划分模型,由于出行方式划分不属于本文研究内容,故直接引用其结果。

3) 模型估计结果与变量影响分析。

将南京都市圈60个区县按80%与20%的比例随机划分训练集(用于参数标定)和测试集(用于效果检验)。由于 L_d 是出行分布阶段采用目的地选择模型(非重力模型)所特有的变量,为便于比较,分别对包含与不包含 L_d 的情况进行参数估计。候选变量中存在明显线性相关的变量,将其输入SPSS软件,采用“步进”方法进行线性回归自动迭代,筛选后的变量及参数估计值见表1。采用平均绝对误差(Mean Absolute Error, MAE)和均方根误差(Root Mean Square Error, RMSE)衡量预测值与真实值的差异,模型验证结果见表2。

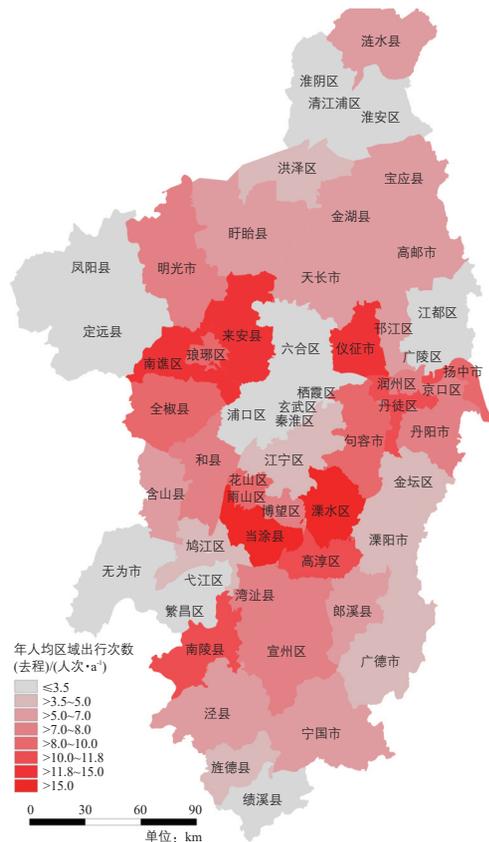


图3 基于手机信令数据的年人均区域出行次数(去程)
Fig.3 Annual per-capita regional travel frequency (outbound) based on mobile signaling data

由表1和表2可知：①不含 L_d 的模型所筛选出的变量(除 L_d 外)与含 L_d 的模型一致，说明没有其他变量可以替代 L_d 的作用； L_d 系数显著为正，表明外界对该区县的综合吸引力直接影响区域出行频率，且不包含 L_d 的模型误差显著增大；② L_s 系数为正，说明副城至主城的出行越便捷，越容易诱发额外出行；③所在区域单元GDP系数为负，说明当地经济越发达，居民越倾向于本地活动，减少跨区域出行；④相邻区域单元数量系数为正，表示相邻区域单元越多，区域出行发生的可能性越大。

3.2 出行生成频率选择模型

区域出行生成模型通常采用出行频率选择Logit模型，其选择肢为不同的出行频率等级(见图4)。

1) 问卷调查设计与变量处理。

以问卷调查获取的个人区域出行频率作为因变量。由于区域出行频率普遍较低，为充分捕捉其分布特征，调查中将非通勤出行频率划分为6个等级：一年内无外出、每年

1~2次、每年3~10次、每月1~2次、每月3~5次、每月6~10次。为应用图4所示模型结构，可根据调查年期期望出行次数以及出行频率分布，将其转换为12次月出行频率(每次月出行频率对应图4中的4项)，同时保证年期期望出行次数不变。

需要说明的是，区域出行调查难度大，问卷完成率和有效率普遍较低。为了充分获取有效信息，本次调查以年度为周期进行询问，未区分工作日与非工作日出行，因此与基于工作日的手机信令数据存在统计口径差异。鉴于本文重点在于验证参数逻辑的合理性，且后续将对参数进行校正，该偏差尚在可接受范围内。未来调查中建议增设工作日/非工作日选项，并可结合比例估计等方法简化操作。

自变量包括问卷调查获得的个人特征：性别(男=1，女=0)，年龄(20~49岁=1，其他=0)，职业(单位负责人、专业技术人员、办事人员和有关人员、商业服务业人员=1，其他=0)，拥车情况(有=1，无=0)，收入水平($\leq 2\,499$ 元=1，2 500~3 999元=2，4 000~7 999元=3，8 000~19 999元=4， $\geq 20\,000$ 元=5)。同时将 L_d 和所在市相邻市数量作为自变量。区域出行生成频率选择模型参数估计值见表3。由参数符号可知：男性、中青年、常差旅职业人员、高收入人员及有车群体的区域出行频率更高； L_d 对出行频率具有显著正向影响；而出发地所在都市圈的中心程度则呈负向影响，即外围地区因受中心城区吸引更易产生区域出行。

2) 人口合成方法与数据代表性分析。

为将出行频率选择模型应用于区县层面，需先获取反映个人特征组合的人口合成数据。该方法以调查数据为种子，以第七次全国人口普查长表数据为边界约束，通过迭代调整使合成人口在各特征维度上与区县统计资料一致，其过程类似于人口扩样，具体

表1 区域出行生成线性回归模型参数估计值

Tab.1 Estimates of the linear regression model parameters for regional trip generation

变量	参数估计值	显著性水平	
常数项	2.389	0.123	
L_s	1.279	0.022	
含 L_d 的筛选变量	所在区域单元GDP/万亿元	-5.689	0.000
	相邻区域单元数量/个	0.701	0.029
L_d	2.134	0.002	
常数项	4.237	0.010	
L_s	1.934	0.001	
不含 L_d 的筛选变量	所在区域单元GDP/万亿元	-6.481	0.000
	相邻区域单元数量/个	1.211	0.000

表2 出行生成阶段各模型评价指标

Tab.2 Evaluation indicators of models in the trip generation stage

数据集	评价指标	含 L_d 的线性回归模型	不含 L_d 的线性回归模型	小样本出行频率选择模型	小样本出行频率选择模型总量校正	校正泊松回归模型
训练集	MAE	1.461	1.654	2.839	2.567	1.808
	RMSE	2.014	2.269	3.420	3.387	2.539
测试集	MAE	1.618	2.484	1.945	2.735	1.793
	RMSE	2.424	3.163	3.438	4.115	2.401

步骤可参考文献[15]。

网络问卷尽管难以完全保证样本代表性，但通过在都市圈空间层面尽可能均匀收集问卷，基本覆盖了各年龄段人群。针对部分外围区县人口属性组合缺失的问题，通过在种子数据中补充一组低权重的各类人口属性组合，确保了人口合成过程的收敛。图5展示了合成后各区县部分人口属性(性别、年龄、职业)与“七普”数据的对比情况。

需要说明的是，对于样本较少的地区，该方法通过均匀调整属性组合比例以满足区县级的各项统计数据，可能与实际存在偏差。因此，在有条件时仍建议使用更全面的种子数据。

3) 模型校正的局限性与区域出行建模的现实挑战。

受限于问卷样本量，出行频率选择模型直接输出的区县人均区域出行次数与手机信令数据存在规模差异。常规做法是通过调整常数项使模型输出结果与观测值一致，但此类强行拟合数据可能损害模型的泛化能力。本文对比以下两种情形的模型效果：直接应用模型，以及通过调整“不出行”选择肢的常数项使训练集出行总量与真实值一致。结果如表2所示，可见不仅模型效果(对应“小样本出行频率选择模型”列)明显差于线性回归模型，常数项校正后效果(对应“小样本出行频率选择模型总量校正”列)进一步恶化。然而，这并不表明出行频率选择模型效果一定不如线性回归模型，而是反映出小样本调查数据难以准确标定复杂模型参数。这也说明常数项校正虽可勉强拟合当前数据，却可能削弱模型的预测能力。

这一问题凸显了区域出行建模的现实困境：大数据通常缺乏个人特征，无法直接用于标定出行频率选择模型；而区域层面大规模出行调查成本高昂、难以实施。因此，如何有效融合大数据与小样本调查数据，成为提升区域出行需求分析能力的关键挑战。

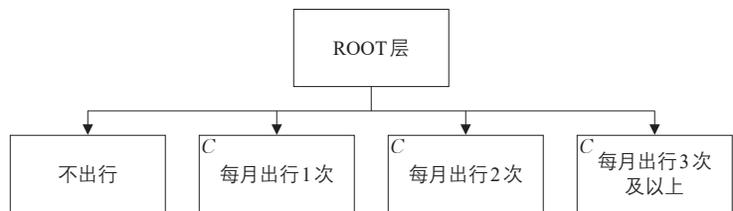
3.3 校正泊松回归模型

基于调查数据标定的模型虽能提供个人特征对出行影响的有效信息，且变量参数取值合理，但因样本量有限，难以充分反映全局规律。因此，需引入区县级特征进行重新标定，并利用训练集区县的出行量信息校正出行频率。具体步骤如下：

1) 将基于调查数据标定的模型应用于训练集区县的人口合成数据，通过各选择肢对应的出行频率与其概率加权求和，得到每个离散个体(组合)的年区域出行次数，再按区县内对应个体(组合)的人数加权求和，得到区县级区域出行次数的初始预测值；

2) 将训练集区县的真实区域出行次数除以步骤1)得到的预测值，得到训练集各区县的调整系数，用以校正每个离散个体(组合)的年区域出行次数；

3) 以校正后的出行次数作为因变量，在原有自变量基础上，补充区县级特征变量(包括所在城市GDP、GDP_c、相邻区域单元数、副城与主城GDP乘积、 L_s)，进行泊松



注：C表示对应选择肢带常数项。

图4 出行频率选择Logit模型结构

Fig.4 Structure of the Logit model for travel frequency choice

资料来源：文献[13]和[14]。

表3 区域出行生成频率选择模型参数估计值

Tab.3 Estimates of the frequency choice model parameters for regional trip generation

参数对应变量_所属选择肢	估计值	显著性水平
性别_0	-0.216	0.000
年龄_0	-0.863	0.000
职业_0	-0.343	0.000
收入水平_1	0.145	0.000
收入水平_2	0.145	0.000
收入水平_≥3	0.311	0.000
拥车情况_1	0.331	0.000
拥车情况_2	0.331	0.000
拥车情况_≥3	0.199	0.001
所在市相邻市数量_所有有出行	-0.169	0.000
L_d _所有有出行	0.407	0.000
常数项_1	-3.223	0.000
常数项_2	-3.916	0.000
常数项_≥3	-3.449	0.000

1)表中“所属选择肢”非该变量取值水平，同一变量可能出现在不同选择肢中；
2)若以调整常数项方式按训练集的区县区域出行总量进行校正，则增加0次选择肢的常数项取值0.233。

回归建模。

泊松回归假设个体在一定周期内的区域出行次数服从泊松分布，并对其期望值进行回归建模，公式如下：

$$\ln \mu_i = \alpha + \beta^T X, \quad (1)$$

式中： μ_i 为第 i 个分类变量组合(除收入分为 5 个水平外，其余 4 个人特征均为 2 个水平，共 80 种特征组合)对应的年期望出行

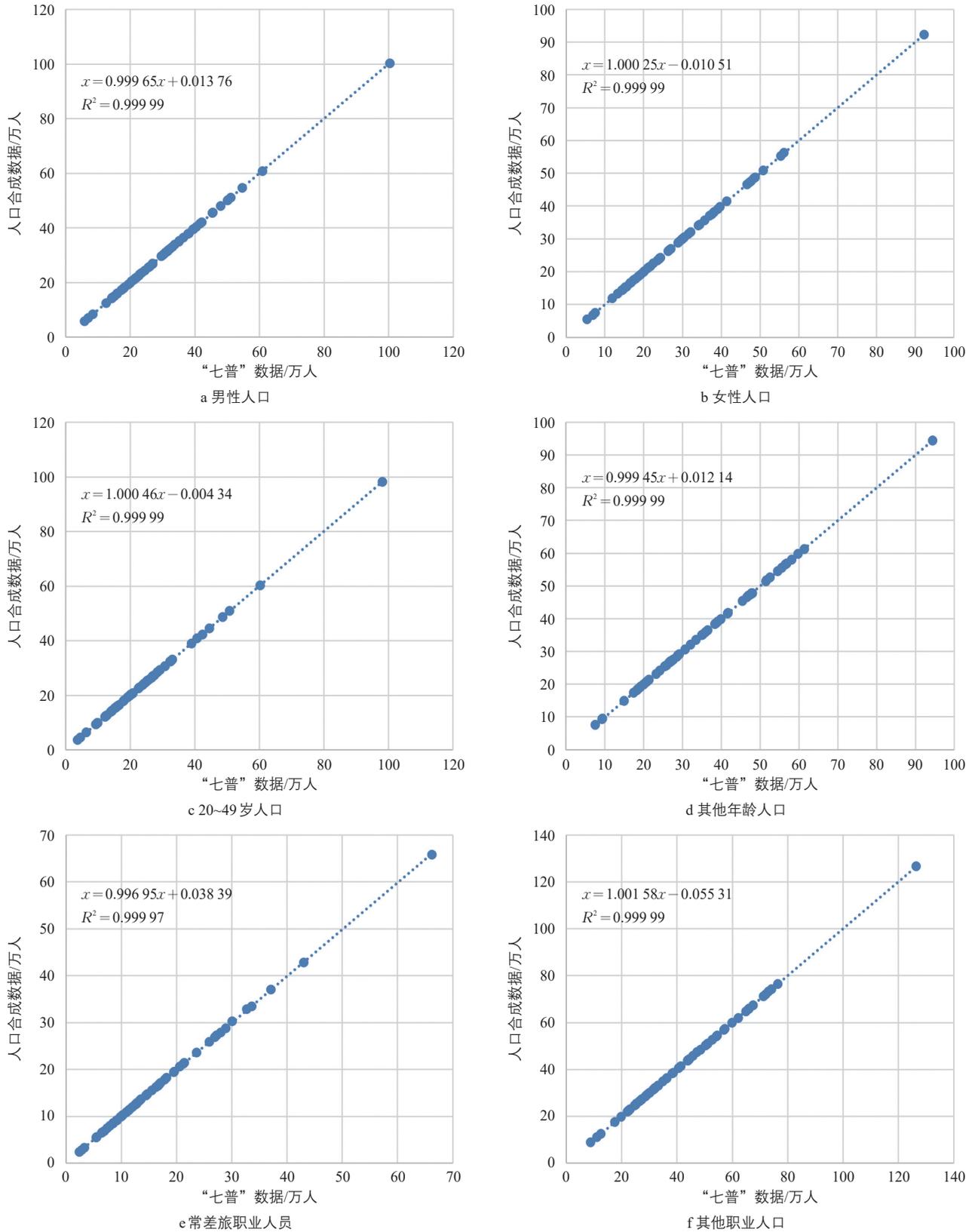


图5 南京都市圈区县级人口合成部分人员属性边界约束拟合情况

Fig.5 Boundary constraint fitting of selected individual attributes in district/county-level population synthesis in the Nanjing Metropolitan area

次数； α 为常数项； β^T 为待估计系数列向量的转置； X 为自变量的列向量，其中连续变量取该组内的平均值。

将训练集校正后的数据整理为 80 条记录(对应 80 种特征组合)，字段包括 5 个人特征变量的取值、各连续变量在该组合中的平均值，以及组合人数和年区域出行次数总和。在 SPSS 中进行分析时，将年区域出行次数设为权重变量，组合人数设为单元格结构，分类变量作为因子，连续变量平均值作为协变量，运行泊松回归模型。

校正泊松回归模型参数估计值如表 4 所示，对于共有变量，泊松回归模型中各变量的符号含义与出行频率选择模型一致。由表 4 可知：所在城市 GDP 系数为正，表明全市经济越发达，区域出行倾向越高；GDP_c 系数为负，反映主城区居民的区域出行倾向低于副城；副城与主城 GDP 乘积系数为负，原因在于控制城市 GDP 后，该乘积越大意味着主城区与副城经济规模越接近，相互吸引作用反而减弱。

模型验证结果见表 2。测试集的 RMSE 为所有模型中最优，虽然 MAE 不及含 L_d 的线性回归模型，但 RMSE 对较大误差更为敏感，有利于在客流预测中避免局部出现较大偏差。因此，可认为校正泊松回归模型整体表现最优。尽管优势不显著，但该模型具备个人特征解释能力，在未来数据条件改善时，能够提供更合理的预测和更精确的决策支持。同时，由于传统模型方法无法获得目的地选择模型的 L_d ，与不含 L_d 的线性回归模型(RMSE 为 3.163)相比，本模型测试集的 RMSE 为 2.401，误差降低了 24%。

4 区域出行分布

为便于评估出行分布模型所得结果的准确性，所有分布模型的对比都基于真实的出行生成数据展开。基于手机信令数据获取的日均区域出行分布(去程)见图 6，以此作为出行分布真实值用于模型标定(训练集部分)与效果检验(测试集部分)。

4.1 重力模型

重力模型普遍应用在城市交通模型的出行分布阶段，通常采用双约束形式(即各交通小区的发生量和吸引量均预先给定)。由于区域出行生成模型仅提供出行发生量，为

保持模型间的可比性，本文采用仅约束出行发生量的单约束形式，并以特征变量的线性组合表征出行吸引量。重力模型计算公式为：

$$y_{ij} = P_i \frac{(\gamma^T X_j + \alpha) e^{k L_{sij}}}{\sum_j (\gamma^T X_j + \alpha) e^{k L_{sij}}}, \quad (2)$$

式中： y_{ij} 为区县 i 至区县 j 的日均区域出行去程出行量/(人次·d⁻¹)； P_i 为区县 i 的日均区域出行去程总量/(人次·d⁻¹)； γ^T 为待估计系数列向量的转置； X_j 为表征区县 j 出行吸引力的自变量列向量； α 为常数项； k 为待估计参数； L_{sij} 为区县 i 至区县 j 的 L_s 值。

在以上重力模型中，阻抗函数采用指数函数形式，其核心变量——综合出行便捷程度 L_s 本质上反映的是出行效用，更适合用指数函数表达；出发量约束则采用线性权重形式，与重力模型的迭代方法一致。该模型相当于在给定各区县出行发生量的基础上，求解各自对外区域出行量的分布。由于不同出发区县的出行分布相对独立，本文继续沿用出行生成阶段所划分的 80% 训练集和 20% 测试集区县，以便系统分析模型效果。

X_j 的候选变量包括：区县第二产业 GDP、第三产业 GDP、人均 GDP、常住人

表 4 校正泊松回归模型参数估计值

Tab.4 Estimates of the corrected Poisson regression model parameters

参数	估计值	显著性水平
常数项	2.496	0.000
性别 0	-0.156	0.000
年龄 0	-0.638	0.000
职业 0	-0.215	0.000
收入 1	-0.550	0.000
收入 2	-0.275	0.000
收入 3	-0.223	0.000
收入 4	-0.053	0.000
拥车情况 0	-0.151	0.000
所在市相邻市数量/个	0.006	0.000
L_d	0.115	0.000
所在城市 GDP/万亿元	0.921	0.000
GDP _c /万亿元	-1.797	0.000
相邻区域单元数量/个	0.141	0.000
副城与主城 GDP 乘积/万亿元 ²	-9.681	0.000
L_s	0.445	0.000

注：收入为 5 及其他个人特征取值为 1 的情况为参照水平，估计值为 0。

表6 出行分布阶段各模型评价指标

Tab.6 Evaluation indicators of models in the trip distribution stage

数据集	评价指标	重力模型	目的地选择模型MNL	目的地选择模型城市NL	含交互MNL	含交互城市NL	含交互区域单元NL
训练集	CPC	0.653	0.652	0.673	0.699	0.713	0.721
	RMSE	301.0	306.2	283.4	259.9	246.3	241.7
测试集	CPC	0.694	0.689	0.688	0.728	0.740	0.750
	RMSE	349.2	359.6	347.3	299.4	290.5	285.7

性越高。

4.2.1 常规模型

MNL模型在形式上与重力模型类似，但所有计算项均采用指数函数形式；NL模型则将相互关联的区县组成Nest，例如可根据“先选城市、再选区县”的出行决策过程，按地级市划分Nest。本文分别对这两种模型采用与重力模型相同的候选变量进行估计，结果见表5和表6。结果显示，各选择肢共有的常数项因不影响最终选择而不产生作用；与重力模型相比，模型中仅多出一个影响较小的常住人口变量。然而，MNL模型测试集的CPC低于重力模型，而NL模型测试集的CPC进一步降低。这说明在使用相同变量的情况下，目的地选择模型并未表现出明显优势，预测性能反而略有下降。

造成上述结果的主要原因是Logit模型的潜能未被充分利用。重力模型仅能考虑出行发生量、吸引力以及起讫点间的阻抗/效用，若Logit模型也仅采用相同变量，自然无法发挥其在解释个体离散特征方面的优势。然而，如前所述，包含离散特征的区域出行数据较难获取，现有建模只能基于区县级集计数据。在此条件下，Logit模型中的目的地选择仅与目的地属性有关，而与出发地特征无关(因为在所有目的地选择肢中出发地特征相同，其影响会被抵消)。但这显然与真实情况不符：不同出发地的居民对目的地的偏好必然存在差异。

综合以上情况，在建模中引入起点与终点之间的交互效应，是解决上述问题的关键。

4.2.2 含交互模型

不同出发地的居民在选择目的地时存在偏好差异，这意味着起点与终点变量之间必然存在多种交互效应。为分析其中最主要的交互影响，本文考虑市域内主城与副城之间因行政隶属关系而产生的额外联系——出行生成阶段相关变量的有效性已初步印证这一

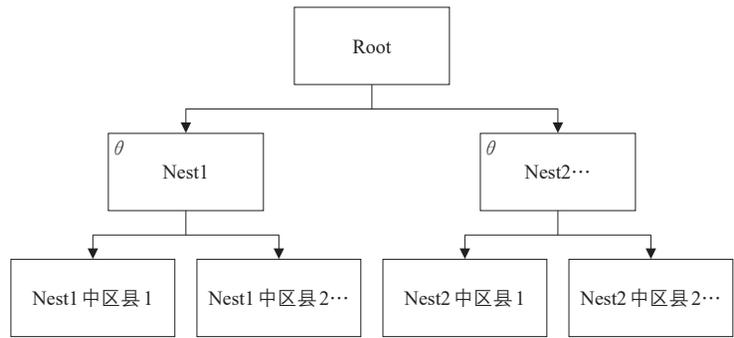


图7 目的地选择Logit模型结构

Fig.7 Structure of the Logit model for destination choice

点。据此，构建交互变量 GDP_{od} ，其取值为起点区县与终点区域单元GDP乘积的平方根；若起点与终点跨越地级市行政区，则该变量取值为0。

将此交互变量纳入候选变量集进行模型标定与验证，结果见表5和表6。可以看出， GDP_{od} 系数较大且具有显著正向影响。MNL模型测试集的CPC从重力模型的0.694提升至0.728，NL模型进一步提升至0.740；如果将Nest按区域单元划定，CPC可达0.750；测试集RMSE较重力模型提升18%。这说明，即使在缺乏离散个体数据的情况下，引入有效的交互项仍能显著提升目的地选择模型在区域出行分布预测中的性能，使其优于重力模型。同时，Nest划分越符合实际决策结构，越能准确刻画出行选择行为。然而，若模型中缺失关键交互变量，NL模型的表现可能反而不如MNL模型。这是因为模型无法捕捉不同出发地居民的出行习惯差异，导致选择行为描述出现偏差，此时即便优化Nest结构也难以弥补这一缺陷。

5 结论

本文研究了区域出行需求模型与城市交通模型在四阶段模型中的主要差异——出行

生成与出行分布中的建模方法及其效果。基于手机信令数据, 本文通过划分训练集与测试集的验证方式, 为所构建的区域出行需求模型的优势提供了实证依据。研究提出了一种结合通用手机信令数据与小样本问卷调查数据的低成本建模流程, 能够在有限资源下构建具有一定精度的区域出行需求模型。主要结论如下:

1) 在符合相关数据管理规定的前提下, 可对特征变量进行适当合并与简化以获取与出行行为关联的个人特征数据, 这样能够最大限度地发挥出行频率选择模型和目的地选择模型的优势;

2) 在无法获得大量含个人特征的出行数据时, 可先通过小样本调查数据标定离散选择模型, 再通过出行大数据校正、标定泊松回归模型, 从而在预测精度上超越传统回归模型, 同时保留对个人特征的解释能力;

3) 若模型中既不引入个人特征, 也不包含起终点交互变量, 则目的地选择模型在区域出行分布中的表现甚至弱于重力模型; 而引入起终点交互变量后, 其预测效果显著优于重力模型, 若进一步结合合理的Nest结构划分, 可进一步提升模型性能;

4) 综合来看, 即使在缺乏大量个人特征数据的情况下, 本文所提出的区域出行需求建模方法, 在出行生成和出行分布阶段的预测精度相比传统城市交通模型方法仍分别提升约24%和18%。

综上, 区域出行需求模型的构建应区别于传统城市交通模型的建模思路。若仅沿用城市交通模型框架, 并简单将模型范围扩展至区域尺度, 则跨区域部分的预测往往需依赖额外简化建模或主观调控手段, 稳定性较差; 而区域出行需求模型基于大范围出行特征进行整体建模, 能够提供更为客观的出行规律。与既有区域出行需求模型构建方法相比, 本文明确了应充分利用的信息类型(如Logsum、起终点交互变量等), 并提出了相应的校正技术建议。

参考文献:

References:

[1] 中华人民共和国国家发展和改革委员会. 国家发展改革委关于培育发展现代化都市圈的指导意见(发改规划(2019)328号)[A/OL]. (2019-02-19)[2025-03-05]. <https://www.gov.cn/zhengce/zhengceku/2019-09/29/>

[content_5434981.htm](https://www.gov.cn/zhengce/zhengceku/2019-09/29/content_5434981.htm).

- [2] MARSHALL N, GRADY B. Travel demand modeling for regional visioning and scenario analysis[J]. *Transportation research record: journal of the Transportation Research Board*, 2005, 1921(1): 44-52.
- [3] BÖRJESSON M. Forecasting demand for high speed rail[J]. *Transportation research part A: policy and practice*, 2014, 70: 81-92.
- [4] PARK K, SABOURI S, LYONS T, et al. Intrazonal or interzonal? improving intrazonal travel forecast in a four-step travel demand model [J]. *Transportation*, 2020, 47(5): 2087-2108.
- [5] SIMINI F, GONZÁLEZ M C, MARITAN A, et al. A universal model for mobility and migration patterns[J]. *Nature*, 2012, 484(7392): 96-100.
- [6] LIMTANAKOOL N, DIJST M, SCHWANEN T. The influence of socioeconomic characteristics, land use and travel time considerations on mode choice for medium- and longer-distance trips[J]. *Journal of transport geography*, 2006, 14(5): 327-341.
- [7] MOECKEL R, FUSSELL R, DONNELLY R. Mode choice modeling for long-distance travel [J]. *Transportation letters*, 2015, 7(1): 35-46.
- [8] ROHR C, FOX J, DALY A, et al. Modeling long-distance travel in Great Britain[J]. *Transportation research record: journal of the Transportation Research Board*, 2013, 2344(1): 144-151.
- [9] BRADLEY M, OUTWATER M L, FERDOUS N. Implementation of a practical model system to predict long-distance travel for the entire US population[J]. *Transportation research record: journal of the Transportation Research Board*, 2016, 2563(1): 9-17.
- [10] BERNARDIN V L, FERDOUS N, SADRSDAT H, et al. Integration of national long-distance passenger travel demand model with Tennessee statewide model and calibration to big data[J]. *Transportation research record: journal of the Transportation Research Board*, 2017, 2653(1): 75-81.
- [11] ZHANG D, LUCHIAN S, RAYCROFT J, et al. Induced travel demand modeling for high-speed intercity transportation[J]. *Transportation research record: journal of the Transportation Research Board*, 2019, 2673(3): 189-

- 198.
- [12] FLYVBJERG B, HOLM M K S. Demand forecasts in the case of transportation[J]. Journal of the American Planning Association, 2005, 71(2): 131-146.
- [13] 雷磊. 城市与区域一体化的出行需求分析理论与方法研究[D]. 成都: 西南交通大学, 2010.
- LEI L. Research on integrated urban and regional travel demand analysis theory and method[D]. Chengdu: Southwest Jiaotong University, 2010.
- [14] 丘建栋, 刘恒, 金双泉, 等. 区域交通模型建设思考: 以广东省综合交通模型为例[J]. 城市交通, 2016, 14(2): 59-66.
- QIU J D, LIU H, JIN S Q, et al. Regional transportation model development: a case study of Guangdong comprehensive transportation models[J]. Urban transport of China, 2016, 14(2): 59-66.
- [15] 张天然, 朱春节, 陈先龙, 等. 高级出行需求模型的增强人口仿真研究[J]. 城市交通, 2023, 21(5): 113-124.
- ZHANG T R, ZHU C J, CHEN X L, et al. Enhanced population simulation for advanced travel demand modeling[J]. Urban transport of China, 2023, 21(5): 113-124.
- [16] LIU Z, MIRANDA F, XIONG W, et al. Learning geo-contextual embeddings for commuting flow prediction[C]//Proceedings of the AAAI conference on artificial intelligence. New York: Association for the Advancement of Artificial Intelligence (AAAI), 2020: 808-816.

(上接第43页)

- [5] 物流评论. 物流近期趋势: 劳动力短缺、物流共享、技术趋势等[EB/OL]. (2017-06-20)[2024-06-10]. <https://www.sakata.co.jp/logistics-366/>.
- [6] City Logistic. City logistic reduces congestion and emissions by rethinking city logistics with a customer-centric focus and electrified deliveries[EB/OL]. (2023-12-20)[2024-06-10]. <https://www.citylogistic.dk/>.
- [7] Port of Hamburg. Europe's largest rail port[EB/OL]. (2022-05-25)[2024-06-10]. <https://www.hafen-hamburg.de/cn/transport/rail/>.
- [8] New York City. Trucks & commercial vehicles[EB/OL]. (2021-11-15)[2024-06-10]. <https://www.nyc.gov/html/dot/html/motorist/trucks.shtml>.
- [9] Transport for London. Ultra Low Emission Zone[EB/OL]. (2023-10-18)[2024-06-10]. <https://tfl.gov.uk/modes/driving/ultra-low-emission-zone>.
- [10] 美国中文网. 2035年一半以上卡车需电动加利福尼亚州新规扫清障碍[EB/OL]. (2020-09-29)[2024-06-10]. <https://www.sinovision.net/society/202303/00528841.htm>.
- [11] Port of Antwerp Bruges. HyTrucks consortium aims to have 300 hydrogen-powered trucks on the road in Belgium by 2025[EB/OL]. (2022-04-21)[2024-06-10]. <https://newsroom.portofantwerpbruges.com/hytrucks-consortium-aims-to-have-300-hydrogen-powered-trucks-on-the-road-in-belgium-by-2025>.
- [12] 邱诗永. 基于“避免—转移—提升”的绿色出行经验与启示[J]. 世界环境, 2019(4): 33-36.
- QIU S Y. The green travel experience based on “avoid-shift-improve” and its inspiration [J]. World environment, 2019(4): 33-36.
- [13] 上海市交通委员会交通指挥中心. 2022年上海市交通运行年报[R]. 上海: 上海市交通委员会交通指挥中心, 2023.
- [14] 金昱, 刘皓冰. 基于GPS数据的重型货车碳排放计算及时空分布研究[J]. 交通运输研究, 2022, 8(6): 90-97.
- JIN Y, LIU H B. Carbon emissions calculation and spatio-temporal distribution of heavy duty trucks based on GPS data[J]. Transport research, 2022, 8(6): 90-97.