

区域交通需求预测OD细分技术研究

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: 区域交通需求预测获得的通常是区县级以上的OD, 而进行交通分配需要更细颗粒度的数据, 因此提出OD细分模型以解决已知大区OD求细分OD的问题。以南京都市圈区域出行为例, 基于手机信令等数据构建普通回归模型、混合模型、一体化标定模型、一体化标定混合模型, 并对模型效果进行对比。结果表明: 已知大区OD求细分小区OD可采用指数函数表征小区间出行量, 在进行大区间出行量限制时转换为基于随机效用理论的离散选择模型形式; 进行一体化标定可最大限度提升模型解释能力, 一体化标定模型的效果优于其他模型, 且考虑混合效应能进一步提升效果。此外, 区域出行影响因素按影响程度从大到小分别为: 出行成本、就业岗位规模、居住人口规模、旅游类POI数量和用地混合度, 但中心地区吸引作用可能无法单独通过社会经济指标完全体现, 考虑混合效应更能反映现实。

关键词: 区域出行; 出行需求; 细分OD; 出行分布; 随机效用; 混合模型

OD Disaggregation Technologies for Regional Travel Demand Forecasting

MENG Danqing¹, DAI Junchen², LING Xiaojing²

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

Abstract: Regional travel demand forecasting typically yields origin-destination (OD) data at a coarse spatial resolution, such as the district or county level. However, finer-grained OD matrices are required for detailed traffic assignment. To address this gap, this paper proposes an OD disaggregation modeling approach to derive disaggregated OD matrices from known large-zone OD data. Taking regional travel within the Nanjing Metropolitan Area as a case study, the paper develops four models based on mobile signaling data, including a conventional regression model, a mixed model, an integrated calibration model, and an integrated calibration mixed model, and then compares the model outcomes. The results indicate that disaggregating large-zone OD flows into small-zone OD flows can be effectively represented using an exponential function to characterize the volumes of travel volumes between small zones, which can be transformed into a discrete choice model grounded in random utility theory when imposing constraints on travel volumes between large zones. Integrated calibration substantially enhances the explanatory power of the models, with the integrated calibration model outperforming the other alternatives, and the incorporation of mixed effects further improving model performance. In addition, the key factors influencing regional travel demand, ranked in descending order of impact, are travel cost, employment scale, residential population size, the number of tourism-related points of interest (POIs), and land-use mix. Notably, the attractiveness of central areas may not be fully captured by socioeconomic indicators alone; incorporating mixed effects provides a more realistic representation.

Keywords: regional travel; travel demand; disaggregated OD; trip distribution; random utility; mixed model

收稿日期: 2024-12-10

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: 孟丹青(1986—), 女, 江苏南京人, 硕士, 高级工程师, 研究方向为轨道交通规划设计、交通一体化, 电子邮箱357941105@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。

引用格式: 孟丹青, 戴骏晨, 凌小静. 区域交通需求预测OD细分技术研究[J]. 城市交通, 2025, 23(6): 71-79.

MENG D Q, DAI J C, LING X J. OD disaggregation technologies for regional travel demand forecasting [J]. Urban transport of China, 2025, 23(6): 71-79.

大尺度分区之间的出行量相比于小尺度通常更易获得, 如百度迁徙数据等开源数据公布了部分城市级出行的相对指数。而在模型方面, 模拟中长距离出行的区域出行需求模型(以下简称“区域模型”)与模拟城市内部出行的交通需求模型(以下简称“城市模型”)不同, 其获得的是区县及以上颗粒度的OD, 若要与城市模型衔接进行交通分配, 还需要对数据做进一步细分。因此, 已知大区OD、求细分OD是一个较普遍的问题, 同时在小颗粒度主要数据缺失时, 也需要通过将大颗粒度数据以及小颗粒度次要数据输入模型以获得细粒度数据。解决这些问题对客流预测有明显益处, 也使得对跨区域规划研究项目的数据支撑从粗放式的量级判断提升至可指导方案优化的精细层面。此外, 现实还面临着基于居民出行调查的OD对数据量不足、基于大数据的OD数据小尺度精度不足等挑战, 而实现良好的OD细分可在一定程度上改善这些问题。为此, 本文以已知区县级区域出行OD求细分至街道级OD为例进行模型OD细分技术研究。

1 细分OD研究现状

细分OD的需求通常出现在区域模型与城市模型衔接的阶段。因为区域模型首先得到的是大尺度OD(如区县级或以上), 通过OD细分后才能达到城市模型的交通小区级, 进而完成OD叠加并实现交通分配。关于区域模型, 西方国家对长距离出行通常采用区别于传统城市模型的专门建模方法, 最典型的是将出行分布阶段的重力模型改为目的地选择模型, 这也正是区域模型获得的OD尺度较大的直接原因——可选目的地无法精确至具体地点。西方国家交通模型由官方机构管理并有法定分析的地位, 应用较为普遍^[1], 而中国的区域模型仅在个别城市有研究应用, 更多地区仍然采用扩大地域范围的传统城市模型。

研究区域模型的文献一般对细分OD未作详细说明, 或处理过程较为简单。如R. Moeckel等^[2]将细分OD的过程描述为: 考虑

起终点的居住人口、就业岗位, 并且用重力模型确保细分后的出行距离分布与实际情况一致; C. Llorca等^[3]对大区模型OD细分方法的描述则是: 基于大区内居住、就业、学校位置分布, 将端点类型为居住、就业、学校的大区O/D分别分散至具体的地址, 端点为其他类型则随机细分至小区。

细分OD问题本质上是对大区间交换量加以限制的出行分布问题, 出行分布预测是重点。基于OD间回归关系建立的直接需求预测模型使用更方便, 但信息利用并不充分, 而基于个人行为、更离散化的方式可产生更准确的出行分布预测结果, 提供更具政策敏感性的模型^[4]。使用Logit目的地选择模型进行出行分布可最充分地挖掘和利用数据。然而, 按细粒度划分目的地并不现实, 如南京都市圈内有超过700个街道级行政区——这作为Logit模型选择枝来说数量过多, 也使得细分小区内各类统计数据的收集变得困难。针对该问题, 一个解决办法是使用条件概率, 在起终点大区OD的概率 P_0 确定后, 再分别计算属于起点大区O的小区o作为起点的条件概率 P_1 和属于终点大区D的小区d作为终点的条件概率 P_2 , 三者相乘即为小区od的概率^[5]。然而, P_1 和 P_2 往往不是相互独立的, 因为在终点大区D确定的情况下, 选择小区d的概率与起点大区O中小区o的确定有关——时空距离更近的选择概率更大。因此对 P_1 和 P_2 单独计算意味着忽略了(大区之间的)小区相对分布与时空距离相关的事实。虽然在长距离出行中该问题影响不大, 但在相邻大区的出行建模中应对此予以关注。

因此, 直接建立出行起终点间的回归模型是一个符合实际需求的解决办法。S. BAIN等^[6]基于大区OD数据进行了分方式的回归分析, 通过假设回归参数不变、应用于小区, 并根据大区间OD量进行控制, 实现了OD细分。H. Yamaguchi等^[7]以基于手机信令数据的OD出行量作为因变量, 以两类指标——吸引力指标(设施水平)和出行成本指标(服务水平)作为自变量进行回归分析, 结果显示吸引力指标的影响随季节变化较大,

而出行成本指标的影响随季节变化较小。

在区域出行回归分析具体使用的自变量(影响因素)方面, S. F. A. Baqueri 等^[8]将外部出行模型整合至城市模型, 将涉及外部的大尺度 OD 在城市端细分, 所用变量包含小区的城市内出行量、与 CBD 的距离、小区中心度、兴趣点(Point of interest, POI)密度、基于土地利用的吸引力指标等。S. F. A. Baqueri 等^[9]还对巴基斯坦和比利时的 10 个不同区域或城市建立了以“外部出行占小区出行的比例”为因变量的回归模型, 使用的自变量包含与外部站点(入城点)的距离、用地混合度、人口密度、交叉口密度等。M. Outwater 等^[10]则考虑了出发地、目的地是城市地区、郊区还是乡村的哑元变量, 以提高模型解释能力。

综上, OD 细分在区域模型的处理中通常较为简单, 存在遗漏重要影响因素的风险, 研究中也缺乏对大区间出行量限制的考虑。因此, 对 OD 细分问题有待进一步开展针对性研究。

2 研究方法及数据描述

本文通过建立小区级出行回归分析模型、并进行大区间出行量控制的方式完成大区 OD 细分, 同时通过建立多种模型并对拟合数据的好坏、泛化性能进行评价, 给出建议模型。

使用的因变量数据集为南京都市圈范围内、基于手机信令数据的街道级区域出行 OD 表, 包含 8.7 万个有出行的街道级出行 OD 对。本文将南京都市圈内的区域出行定义为跨越市辖区(同一城市内所有市辖区合并)或相对独立的区县级行政边界的出行, 其中南京溧水区、高淳区作为 2 个独立的区域单元, 都市圈内其他地级市均以市辖区、县市为区域单元。将以上 OD 按区县集计。研究使用的其他数据主要包含街道级的居住人口、就业岗位数据等。

为验证模型的泛化性, 随机抽取 70% 的数据作为训练数据集用于标定参数, 剩余的 30% 作为测试数据集用于检验模型效果。模型不涉及调整超参数实现最佳效果, 故无须设置验证数据集来指导调整, 直接以测试数据集反映模型效果。其中, 训练数据集、测试数据集中的已知大区量, 为各自数据集所涵盖街道 OD 的总量。

3 基于回归分析的模型构建

3.1 自变量选择

自变量选取需兼顾有效性与实用性。结合既有研究及实际经验, 初选影响区域出行的自变量包括: 公共交通出行广义时耗、居住人口规模、就业岗位规模、旅游类 POI 数量、用地混合度、起终点街道至所在区县中心的空间距离、交叉口密度、起终点属性(街道、乡、镇、各类园区等)。

其中, 旅游类 POI 数量直接采用高德开放平台 POI 分类中的“风景名胜类”大类, 用于反映景点的客流吸引能力; 公共交通出行广义时耗通过出行导航时间与费用综合校正获得; 居住人口、就业岗位规模通过手机信令数据采集; 用地混合度通过兴趣面中的面积与建筑轮廓中的层高数据综合计算获得(面向规划时, 根据规划的用地布局以及片区的指导容积率进行计算); 街道至区县中心的空间距离通过 TransCAD 模型网络计算; 交叉口密度通过开放街道地图(Open Street Map, OSM)进行计算; 起终点属性则根据实际确定。

使用 SPSS 的逐步选择法筛选普通回归模型的变量。该方法通过逐步选择最能解释因变量变异性的自变量来简化模型, 结合了前向选择、后向消除两种策略的优点。其选择过程主要包括 4 部分: 1) 从空模型开始; 2) 每一步逐一考虑尚未进入模型的所有候选变量, 选择一个能使模型改进最多的变量加入, 改进的标准为统计显著性; 3) 对已存在的变量逐个评估是否应该被移除; 4) 当没有新变量可以加入也没有现有变量可移除时过程结束。按 5% 统计显著性标准筛选后, 自变量为: 公共交通出行广义时耗, 终点街道的就业岗位规模, 起点街道的居住人口规模, 起、终点街道的旅游类 POI 数量, 起、终点街道的用地混合度。

3.2 模型构建

3.2.1 普通回归模型

参照重力模型构建出行量模型, 分子为出行发生、吸引函数, 分母为出行阻抗函数。函数形式一般为幂函数、指数函数。本文采用指数函数, 小区间出行量

$$q_{od0} = \exp(\beta^T X + \alpha), \quad (1)$$

式中: q_{od0} 为小区级区域间未经总量控制的出行量/(人次·d⁻¹), 可作为细分离散权重;

β^T 为待估计系数列向量的转置; α 为模型的常数项; X 为自变量的列向量, 包括起点小区居住人口规模、旅游类 POI 数量、用地混合度, 终点小区就业岗位规模、旅游类 POI 数量、用地混合度, 起终点小区间公共交通出行广义时耗。

随后进行大区间出行量的限制, 小区间出行量

$$q_{od} = \frac{q_{od0}^k}{\sum_{x \in O, y \in D} q_{xy0}^k} Q_{OD}, \quad (2)$$

式中: q_{od} 为小区间(不含小区内)出行量/(人次·d⁻¹); Q_{OD} 为出发点小区所属的区县、终点小区所属的区县间的跨区域出行量/(人次·d⁻¹); k 为参数, 对于式(1)单独参数标定的情况, $k=1$, 因为标定时已经以小区间出行量观测值为拟合对象。

只有当式(1)的函数形式是指数函数时, 将式(1)和式(2)组合后得到的式(3)才能转换为基于随机效用理论的离散选择模型。小区间出行量是对应大区出行量乘以选择该小区

OD 基于效用的概率的结果, 小区间出行量

$$q_{od} = \frac{\exp(U_{od}k)}{\sum_{x \in O, y \in D} \exp(U_{xy}k)} Q_{OD}, \quad (3)$$

式中: U_{od} 为小区间的出行效用值; k 此时转化为效用的缩放系数。

式(1)两边取对数后转换为线性模型, 以小区间出行量的对数为因变量, 进行回归分析, 参数估计值见表 1 相应列, 其中普通回归模型、混合模型因变量均为 $\ln(q_{od})$, 而一体化标定模型、条件概率模型因变量为 q_{od} ; 仅混合模型的自变量和因变量进行了 Z 标准化, 其他模型自变量、因变量均未标准化。按式(2)进行大区间出行量限制, 并计算模型评价指标(见表 2)。其中, 通勤者共同部分指数(Common Part of Commuters, CPC)广泛用于出行预测问题, 其衡量了预测值与目标值之间的一致性^[11]。当没有任何一致时 CPC 为 0, 当两者相同时 CPC 为 1, CPC 计算方法见式(4); 均方根误差(Root Mean Square Error, RMSE)用于衡量预测值与真实值的差异; “(25%, 75%)分位误差”表示误差的 25%, 75%分位值; “相对误差 10%内占比”表示相对误差在 -10%~10%内的比例。此外, 在应用式(2)时, Q_{OD} 为对应数据集(训练数据集或测试数据集)而非总体数据集按区县集计的出行量。

$$CPC = \frac{2 \sum \min(\hat{y}, y)}{\sum \hat{y} + \sum y} Q_{OD}, \quad (4)$$

式中: \hat{y} 和 y 分别为预测值和真实值。

表 1 中的各参数均呈现统计显著且正负号符合逻辑的特征。从系数值看, 就业岗位规模相比居住人口规模更能影响出行量。使用标准化变量进行分析能消除量纲影响, 但使用未标准化变量在应用时更为方便。表 2

表 1 各模型参数估计值

Tab.1 Estimated parameter values of each model

参数	普通回归模型	混合模型 固定效应	一体化标定模型	条件概率模型
终点就业岗位规模/万人	0.013	0.170	0.020	0.054
起点居住人口规模/万人	0.008	0.143	0.014	0.034
起点旅游类 POI 数量/个	0.005	0.063	0.007	0.005
终点旅游类 POI 数量/个	0.006	0.058	0.007	0.003
起点用地混合度	0.698	0.038	1.557	1.743
终点用地混合度	0.875	0.055	1.413	1.776
公共交通出行广义时耗/h	-0.160	-0.292	-0.501	
α	1.276	-0.186	0.000	
k	1.000	1.000	1.266	1.000

表 2 各模型评价指标

Tab.2 Evaluation metrics of each model

数据集	评价指标	普通回归模型	混合模型	一体化标定模型	一体化标定混合模型	条件概率模型
训练	CPC	0.493	0.636	0.584	0.648	0.472
	RMSE	81.99	63.64	66.31	45.10	84.60
	(25%, 75%)分位误差	(-0.33, 6.24)	(-0.33, 3.65)	(-1.60, 4.09)	(-1.60, 3.57)	(-0.51, 6.15)
	相对误差 10%内占比/%	7.82	9.35	6.79	7.32	7.61
测试	CPC	0.519	0.643	0.613	0.644	0.497
	RMSE	56.13	44.73	44.71	40.65	58.70
	(25%, 75%)分位误差	(-0.43, 5.29)	(-0.58, 3.45)	(-1.58, 3.55)	(-1.63, 3.59)	(-0.61, 5.33)
	相对误差 10%内占比/%	9.41	10.98	8.14	6.7	9.15

结果显示，测试数据集误差比训练数据集误差小，这是由于测试数据集在每个大区内抽样的小区数更少，在按小区集计的大区间出行量限制的情况下，模型更易拟合。

3.2.2 混合模型

大区间限制总量情况下的小区回归模型呈现层次结构，每个大区OD是一个层次，同一层次内的出行特征更相似。传统的统计分析模型均假设不同样本间数据完全独立，因此使用混合模型能更准确地解释大区间限制总量情况下的小区回归问题。混合模型除了考虑常规固定效应外，还考虑随机效应，部分变量和常数项有一个随层次变化的系数^[12]，见式(5)。

ln(q_{od0}) = (β^TX + α) + (μ_j^TX + v_j), (5)

式中：j 为小区 OD 所属的第 j 个层次；μ_j^T 指在第 j 个层次中随机效应变化系数期望值列向量的转置；v_j 指在第 j 个层次中随机效应变化常数项的期望值。

由于混合模型较为复杂，数据的扰动对结果影响较大，故对自变量和因变量都进行 Z 标准化后再按式(5)进行拟合。固定效应参数估计值见表 1 对应列。影响区域出行程度从大到小的变量依次是出行成本、就业岗位规模、居住人口规模、旅游类 POI 数量和用地混合度。混合模型中随机效应的检验结果见表 3，其中 Z(ln(q_{od})) 为因变量，自变量均进行了 Z 标准化处理。结果显示，各随机项均拒绝了方差为 0 的假设，即各随机项确实存在。

标准化预测值返回原量纲值后经式(2)进行大区间出行量限制，按限制后的结果计算模型评价指标，见表 2 “混合模型” 列。结果显示，测试数据集 CPC 相比于普通回归模型提升明显，为 0.643。

3.2.3 一体化标定模型

以上模型均为在控制大区间总量前就直接以小区间出行量的观测值为拟合对象，此时式(2)的 k 值为 1，否则为有偏估计。将式(1)和式(2)进行一体化标定，即按式(3)统一估计包含 k 在内的各参数值，可以更充分地提升模型解释能力。因此，使用最大似然估计法估计参数，估计值见表 1 相应列，模型评价指标见表 2。结果显示，参数估计值正负号及大小相对关系与同为未标准化的回归模型基本一致；k 值 1.266 说明存在效用放大作用；常数项在模型中不会带来小区间效用差异、不产生作用，标定结果为 0。

3.2.4 一体化标定混合模型

进一步将混合模型纳入，考虑不同层参数的变化。若按大区 OD 分层则参数过多，故按大区本身而非大区间组合分层：每个大区有作为起点、终点的混合效应参数，其中公共交通出行广义时耗的混合效用由起点、终点混合效应参数加和得到。以表 1 一体化标定模型参数为固定效应参数(实际应重新标定，在此为简化处理)，小区间的出行效用值

U_{od} = U_f + (μ_o^TX_o + μ_d^TX_d), (6)

式中：U_f 为小区 OD 间固定效应的出行效用值，等于按一体化标定模型的参数计算 ln(q_{od0})；μ_o^T，μ_d^T 分别指在起点、终点所在区县中涉及起点、终点变量的随机效应变化系数期望值列向量的转置；X_o，X_d 分别为自变量中涉及起点、终点的变量的列向量，两者均包含起终点小区间公共交通出行广义时耗。

将式(6)代入式(3)进行参数标定，k 值同样取一体化标定模型所得参数。该模型评价指标为几个模型中最优，测试数据集 RMSE 仅为 40.65(见表 2)。以公共交通出行广义时耗的随机效应参数为例分析其在各区县的分布(见图 1)。

该参数的意义在于：在其他指标相同但局部目标 OD 较大或较小时，通过局部阻抗的减弱或增强实现进一步拟合。图 1 中，南京市中心城区随机效用系数为正，相当于减弱了阻抗本身的负效用；而外围地市如镇江、淮安等的随机效用系数多为负，相当于进一步增强了阻抗的作用。结果说明，真实区域出行分布中的中心城区吸引作用比单通过社会经济指标体现出来的中心作用更大。

表 3 混合模型随机效应协方差参数估计值
Tab.3 Estimated covariance parameter values of random effects in the mixed model

对应变量	随机项参数的方差	显著性
终点就业岗位规模/万人	0.063	0.000
起点居住人口规模/万人	0.053	0.000
起点旅游类 POI 数量/个	0.023	0.000
终点旅游类 POI 数量/个	0.021	0.000
起点用地混合度	0.007	0.000
终点用地混合度	0.007	0.000
公共交通出行广义时耗/h	0.144	0.000
α	0.111	0.000

3.3 参数估计方法

普通回归模型及混合模型参数估计直接通过SPSS软件实现,而一体化标定模型及其混合模型需要自定义编程。采用最大似然估计方法进行参数估计时,通过找到使一组观测数据出现概率最大的参数值来确定模型的最佳参数。计算一组数据的概率即概率的连乘函数 \prod 时,为了计算简便取对数,变成求和函数 \sum 时,求优化问题的最小值,因此再乘以负号。此时转化为求负对数似然函数的最小值。

标定用负对数似然函数为 $-\sum(\text{真实值} \times \ln(\text{预测值}) - \text{预测值})$ 。采用Python编程并使用“L-BFGS-B”方法优化,迭代得出参数估计值。

4 模型对比分析

4.1 模型关系讨论

指数函数表征小区间出行量时,进行大区之间出行量限制可转换为离散选择模型,其与应用于OD细分时的条件概率模型本质

相同,但形式存在区别:后者从出发或到达大区角度,考虑各细分出发地或到达地的概率;前者则是以每个起点—终点大区为一组进行概率选择,细分出发点和细分到达点同时进行选择。因此,条件概率模型相当于分别在起点、终点大区独立进行两次选择,再将概率结果相乘;本文模型则是起点选择和终点选择同时进行,模型形式涵盖了条件概率模型,当使用的模型变量不含起终点交互时转化为条件概率模型。

4.2 评价与验证

进一步计算条件概率模型结果。条件概率模型等同于不考虑公共交通出行广义时耗的一体化标定模型,使用的其他变量相同,其参数估计值、评价指标分别见表1、表2。

模型对比结果显示,其他模型均优于条件概率模型(测试数据集CPC为0.497,为各模型中最低);普通回归模型使用简便,但模型解释能力较低(测试数据集CPC为0.519);混合模型操作较为复杂,数据集拟合效果较好(测试数据集CPC为0.643),解释能力较普通回归模型提升显著,但泛化性部分指标(测试数据集RMSE)并不优于一体化标定模型;一体化标定模型使用方便,且模型解释能力良好(测试数据集CPC为0.613);一体化标定混合模型进一步提升了准确性(测试数据集CPC为0.644),操作相对方便,且泛化性强。虽然一体化标定模型及其混合模型的高精度预测占比表现不及其他模型,但反映整体表现的CPC与RMSE均优于其他模型,因此推荐使用一体化标定模型或其混合模型作为OD细分建模形式。以南京市内部区县级区域出行OD细分为例,模型结果(使用基于训练数据集标定的一体化标定混合模型)与手机信令数据的空间分布对比见图2,整体分布态势基本一致。

值得注意的是,最优模型测试数据集CPC为0.644,与1仍有一定差距,这是因为出行的影响因素复杂,如H. Yamaguchi等^[7]基于实证研究指出出行量是随时间波动的,不存在恒定的值,预测也就更为困难。而宏观规划设计注重把握量级,以及基于空间交互的相对关系做出方案判断,因此追求不现实的高准确率不是唯一目标,应更注重模型的可解释性,因此本文验证的模型更符合实际需求。针对更高精度需求的场景,可考虑更多、更精细的变量,但形式上建议使用本

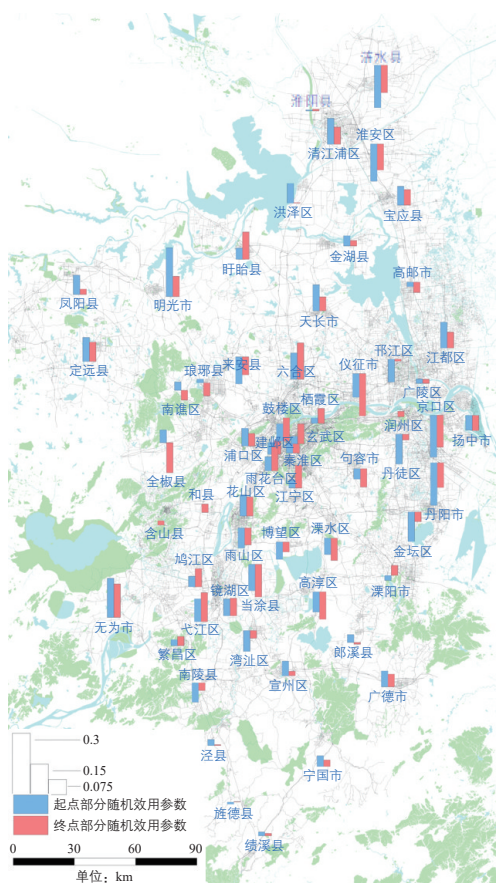


图1 公共交通出行广义时耗的随机效应参数取值分布
Fig.1 Distribution of random effect parameter values for public transportation generalized travel time

文提出的模型，因为该模型并非随意类比构建，而是在形式上等价于基于随机效用理论的离散选择模型。

此外，研究提出的OD细分方法可一定程度上缓解手机信令数据在较细颗粒度空间层次中误差较大的问题，以南京市范围内手机信令数据为例重新建模进行简要分析。将东山片区至主城区、麒麟片区至主城区的小区间全日出行数据分别作为训练数据集、测试数据集，给出麒麟片区至主城区的总量、预测其小区间出行分布(见图3a)：手机信令数据在此颗粒度显示了不合逻辑的出行分布，即在两居住小区间异常集聚(见图3b)；细分模型消除了不合理的预测结果，呈现的出行分布分散于主城区内部(见图3c)。

5 结束语

本文主要面向区域交通需求预测中的OD细分问题，以南京都市圈为例，基于手机信令等数据构建了普通回归模型、混合模型、一体化标定模型、一体化标定混合模型等多种模型，主要结论如下：

1) 已知大区OD求细分小区OD可采用指数函数表征小区间出行量，在进行大区间出行量限制时可转换为基于随机效用理论的离散选择模型形式；进行一体化标定可最大限度地提升模型解释能力，一体化标定模型的效果优于普通回归模型及混合模型，且同时考虑混合效应能进一步提升模型效果。

2) 区域出行影响因素的影响程度从大到小分别为：出行成本、就业岗位规模、居住人口规模、旅游类POI数量、用地混合度。但中心地区吸引作用现象可能无法单独通过社会经济指标完全体现，考虑混合效应更能反映实际情况。

需要说明的是，本文提出的细分模型有其适用范围：应用时大区、小区的划分尺度应与标定时一致，例如标定时使用区县细分街道数据，则应用也只适用于区县细分为街道。

关于时空条件变化下模型的应对能力：对于空间变化，本文模型应用于标定时未使用的、空间不同的测试数据集中仍然表现良好，显示出其具有一定的空间迁移能力；对于时间变化，大区OD相比于小区OD更恒

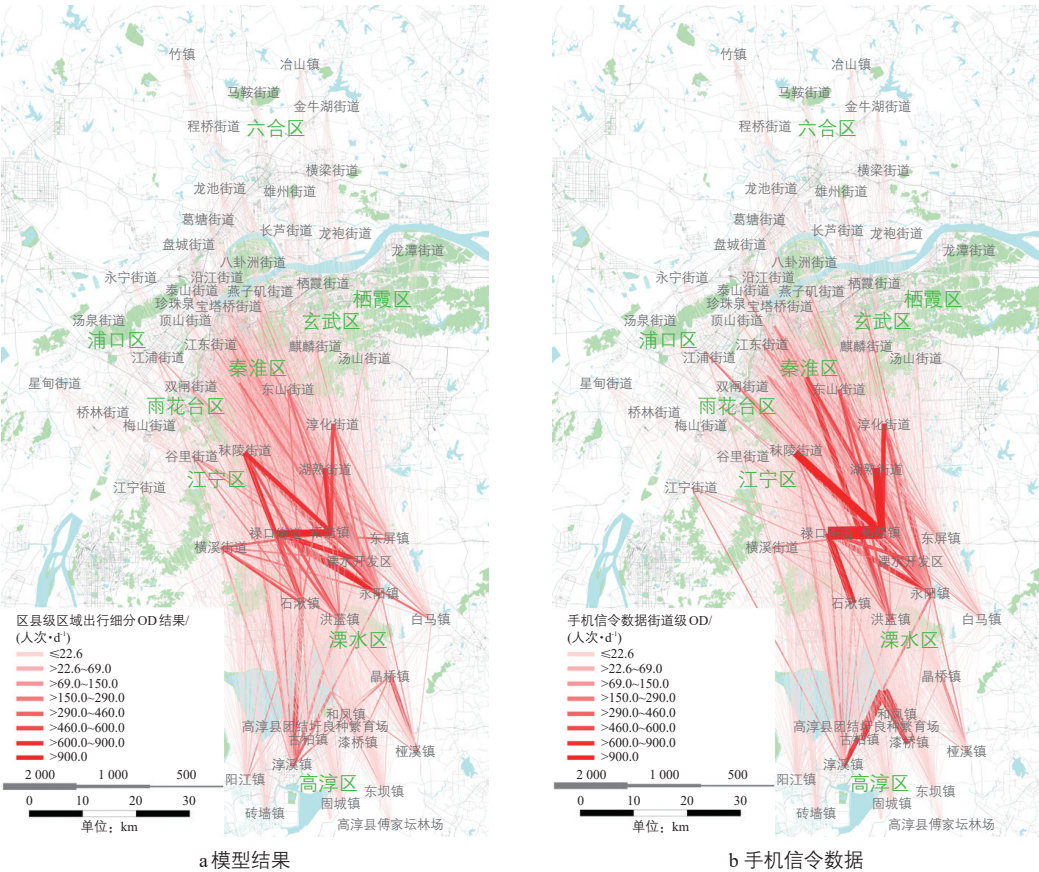
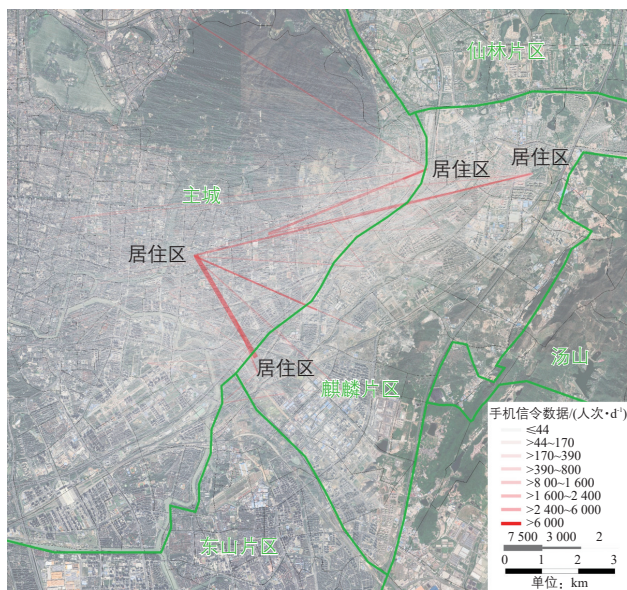


图2 南京内部细分OD模型结果与手机信令数据对比

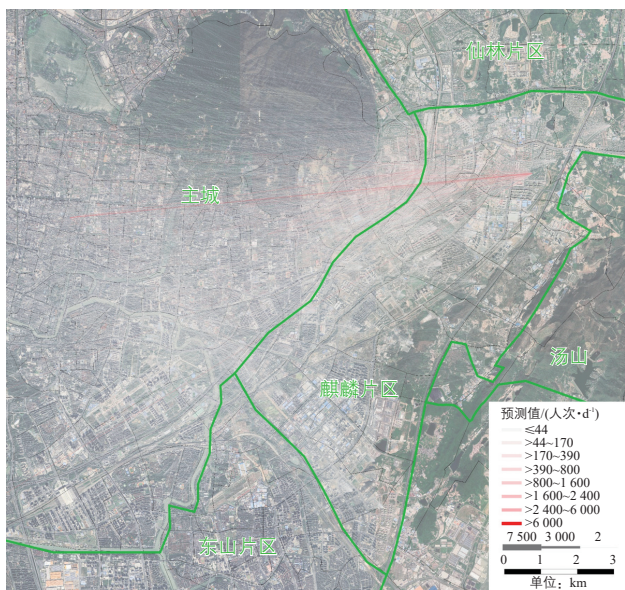
Fig.2 Comparison between OD disaggregation model results and mobile signaling data within Nanjing



a 分析区域与数据划分



b 手机信令数据全日出行分布



c 细分模型预测出行分布

图3 细分模型缓解细颗粒度手机信令数据误差示意图

Fig.3 Illustration of the disaggregation model alleviating errors in fine-grained mobile signaling data

定，因为小区中的随机性在集聚为大区时可一定程度相互抵消，所以即使同一量级的大区 OD 在不同时间可能存在不同的小区 OD 分布，这也正是模型的准确度有所限制的原因——模型解释的是规律性特征，而无法解释随机性特征。在不同的时间，模型表现可能下降，但若相关行为习惯未发生根本变化(如由于新科技出现、行为习惯发生巨大改变)，仍能解释出行的基本规律部分。未来建议进一步研究时空变化条件下的模型迁移性，并寻找达到混合模型效果的所有地区统一的特征变量。

参考文献:

References:

[1] 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.

[2] MOECKEL R, FUSSELL R, DONNELLY R. Mode choice modeling for long-distance travel [J]. Transportation letters, 2015, 7(1): 35-46.

[3] LLORCA C, WINKLER C, MOCANU T, et al. Long-distance and daily travel demand: integration of various travel markets and modeling approaches[J]. Procedia computer science, 2019, 151: 788-793.

[4] ZHANG L, SOUTHWORTH F, XIONG C F, et al. Methodological options and data sources for the development of long-distance passenger travel demand models: a comprehensive review[J]. Transport reviews, 2012, 32(4): 399-433.

- [5] 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.
- [6] BAIN S, HENSHER D A, LI Z. R-Tresis: developing a transport model system for regional New South Wales[J]. Journal of transport geography, 2011, 19(4): 615-622.
- [7] YAMAGUCHI H, SHIBATA M, NAKAYAMA S. Travel matrix decomposition for understanding spatial long-distance travel structure[J]. Complexity, 2023, 2023(1): 1090277.
- [8] BAQUERI S F A, ADNAN M, BELLEMANS T, et al. Comprehensive modeling framework to integrate external trips in a travel demand model[J]. Journal of computing in civil engineering, 2019, 33(3): 04019011.
- [9] BAQUERI S F A, ADNAN M, KNAPEN L, et al. Modelling distribution of external - internal trips and its intra-region and inter-region transferability[J]. Arabian journal for science and engineering, 2019, 44(5): 4517-4532.
- [10] OUTWATER M, TIERNEY K, BRADLEY M, et al. California statewide model for high-speed rail[J]. Journal of choice modelling, 2010, 3(1): 58-83.
- [11] LIU Z C, MIRANDA F, XIONG W T, et al. Learning geo-contextual embeddings for commuting flow prediction[J]. Proceedings of the AAAI conference on artificial intelligence, 2020, 34(1): 808-816.
- [12] 张文彤, 董伟. SPSS 统计分析高级教程 (第3版)[M]. 北京: 高等教育出版社, 2022: 63-75.

(上接第25页)

- [19] 香港规划署. 北往南来2017年跨界旅运统计调查[R]. 香港: 香港规划署, 2017.
- [20] 沈立, 倪鹏飞. 全球产业链演变趋势下中国城市发展格局的未来走向及政策建议[J]. 经济纵横, 2022(2): 60-68.
- SHEN L, NI P F. Future trend and policy suggestions of China's urban development structure under the evolving trend of global industrial chain[J]. Economic review journal, 2022(2): 60-68.
- [21] 王庆云, 毛保华. 科技进步对交通运输系统发展的影响[J]. 交通运输系统工程与信息, 2020, 20(6): 1-8.
- WANG Q Y, MAO B H. Impacts of science and technology on transportation[J]. Journal of transportation systems engineering and information technology, 2020, 20(6): 1-8.
- [22] 孙永海, 马亮, 邓琪, 等. 人性化交通出行角度的都市圈空间尺度研究[J]. 规划师, 2014, 30(7): 32-39.
- SUN Y H, MA L, DENG Q, et al. Metropolitan area scale from humane transportation viewpoint[J]. Planners, 2014, 30(7): 32-39.
- [23] 周军, 孙永海, 孙夕雄. 新时期城市物流空间分类研究[J]. 城市交通, 2021, 19(2): 14-22.
- ZHOU J, SUN Y H, SUN X X. Urban logistics space classification in the new era[J]. Urban transport of China, 2021, 19(2): 14-22.
- [24] MOAVENZADEH J, LANG N S. Reshaping urban mobility with autonomous vehicles: lessons from the city of Boston[C]//The World Economic Forum. System initiative on shaping the future of mobility. Geneva, Switzerland: The World Economic Forum, 2018, 27: 3-33.
- [25] 胡箭, 罗书源, 赖金涛, 等. 自动驾驶对交通运输系统规划的影响综述[J]. 交通运输系统工程与信息, 2021, 21(5): 52-65.
- HU J, LUO S Y, LAI J T, et al. A review of the impact of autonomous driving on transportation planning[J]. Journal of transportation systems engineering and information technology, 2021, 21(5): 52-65.
- [26] 孙永海, 邓琪, 陆锡明. 从综合交通规划角度看小汽车限制政策的实施: 以深圳市为例[J]. 城市交通, 2016, 14(4): 51-58.
- SUN Y H, DENG Q, LU X M. Discussion on car restriction policy from a perspective of comprehensive transportation planning: a case study of Shenzhen[J]. Urban transport of China, 2016, 14(4): 51-58.