

街区尺度下的通勤出行方式挖掘及其影响因子——以北京市为例

阚长城¹, 马琦伟², 党安荣², 万涛³, 高唱⁴

(1. 百度时代网络技术(北京)有限公司, 北京 100085; 2. 清华大学建筑学院, 北京 100084; 3. 天津市城市规划设计研究院, 天津 300201; 4. 中国城市规划设计研究院, 北京 100037)

摘要: 通勤数据的传统获取手段存在成本高、覆盖面小、更新慢等问题, 难以满足实时、高效监测和管理的需求。基于百度地图时空大数据, 综合应用多种机器学习方法, 构建一套识别城市街区尺度下通勤出行方式的技术框架, 具有准确率高、覆盖面广、空间分辨率高等优势。挖掘北京市六环高速公路以内各街区的通勤出行方式构成特征, 结果显示各类交通方式的通勤出行比例相对均衡。进而考察通勤出行方式的具体空间布局, 探索其与路网密度、用地功能混合密度、公共交通设施服务水平3项建成环境因子之间的内在联系。对比分析结果验证了路网密度、用地功能混合密度与小汽车通勤出行呈负相关关系, 轨道交通服务水平对通勤出行的影响具有空间异质性。

关键词: 通勤交通方式; 时空大数据; 机器学习; 绿色交通

Commuting Travel Mode and Its Influencing Factors at Street Block Scale: A Case Study in Beijing
Kan Changcheng¹, Ma Qiwei², Dang Anrong², Wan Tao³, Gao Chang⁴

(1. Baidu.com Times Technology (Beijing) Co., Ltd., Beijing 100085, China; 2. School of Architecture, Tsinghua University, Beijing 100084, China; 3. Tianjin Urban Planning & Design Institute, Tianjin 300201, China; 4. China Academy of Urban Planning & Design, Beijing 100037, China)

Abstract: The disadvantages of the traditional commuting data collection methods are high cost, insufficient special coverage, and slow update, which makes it hard for meeting the real-time effective traffic monitoring and managing needs. Based on the Baidu Map's big spatial-temporal data, this paper develops a technical framework that uses the multiple machine learning methods to identify commuting travel mode at street block scale, which has high accuracy, wide coverage, and high spatial resolution. By analyzing the characteristics of commuting travel mode in different street blocks within the Sixth Ring freeway in Beijing, the results show that all types of commuting travel modes are relatively balanced. Based on the spatial distribution of commuting travel modes, the paper discusses the relationship among three built environment factors, such as roadway network density, mix land use density, and level of service of public transit facilities. The comparative analysis results show that roadway network density and mix land use density are negatively correlated with car commuting travel, and the impact of rail transit level of service on commuting travel is spatially heterogeneous.

Keywords: commuting travel modes; big spatial-temporal data; machine learning; green transportation

收稿日期: 2020-05-09

基金项目: 国家政府间国际科技创新合作重点专项“数字城市规划新技术研发”(2017YFE0118600)课题“数字城市规划技术集成方法研究”资助

作者简介: 阚长城(1985—), 男, 山东临沂人, 硕士, 百度慧眼技术架构师, 主要研究方向: 时空大数据、AI在城市规划治理和智慧城市中的应用。E-mail: 448211363@qq.com

通信作者: 马琦伟(1983—), 男, 浙江绍兴人, 硕士, 高级城市规划师, 主要研究方向: 城市规划、地理信息系统、时空大数据分析。E-mail: maqw15@mails.tsinghua.edu.cn

0 引言

通勤活动指就业者在居住地与工作地之间采用一定交通方式开展的有节律移动, 它是城市居民最主要的日常活动之一。对多数

人而言, 由于居住地点和工作地点一般不会频繁更换, 全体居民的通勤活动所构成的城市通勤格局具有相对稳定性, 能够反映城市空间结构的特点, 因而在空间规划与研究具有重要的地位^[1]。通勤出行在时间上具有

爆发性，因此成为早晚时段交通高峰的最重要原因，公共交通、小汽车等不同的通勤出行方式会对公共交通系统、城市道路系统日常运营造成巨大压力，小汽车通勤出行是造成交通拥堵、城市大气污染等城市病的重要来源^[2]。因此，在国家大力提倡高质量发展，推动绿色出行、低碳城市^[3]的背景下，科学优化城市通勤交通格局十分重要，而准确、高效地获取城市通勤出行方式则是开展城市通勤格局优化的基础性工作，对于城市交通系统的整体优化具有重大现实意义。

另一方面，城市空间作为通勤交通的物质载体，对通勤出行方式的影响十分复杂多样^[4]。从时间维度上看，通勤出行方式受到轨道交通建设、汽车尾号限行、停车供给与政策、道路拥堵水平等多方面的影响而变动^[5]；从空间维度上看，城市范围内不同的区域由于交通可达性、公共交通供给水平和建成环境的不同，通勤出行方式构成也会存在较大差异性^[6]。在科学获取通勤出行方式数据的基础上，系统分析多种城市空间要素对通勤出行方式的影响，对分析发展趋势、制订合理的交通政策、编制交通规划具有很强的支撑作用。

传统上通过问卷调查的方式可以获知居民的通勤出行方式，但由于流程烦琐、获取成本高等问题，一般仅在小范围开展^[7]。国内外许多城市一般以5~10年为周期，开展抽样率为0.5%~2.0%的居民出行调查，获取居民的家庭、个人信息以及一个典型工作日的包含通勤出行在内的全部出行信息。但抽样无法覆盖全部社区，一般而言即便是一次大规模的居民出行调查，也仅能涉及约1/2~2/3的社区^[8]。而且居民出行调查的成本较高，两次调查之间一般间隔5~10年甚至更长时间，且无法对城市通勤产生较大影响的事件进行及时追踪。

基于位置信息的时空大数据是一类研究城市空间关系和出行的新型数据，通过此类数据可以实现用户居住地和就业地的判别，进而识别城市整体的通勤空间流动模式^[9]，近年来已有较多的应用^[10, 11]。本文使用百度地图时空大数据，该数据源具有精度高、覆盖人群广、数据采集速度快等优势，在挖掘城市通勤出行方式方面具备独特优势。基于此数据，本文提出一种挖掘用户通勤出行方式的新技术框架。该框架融合互联网定位数据、地图基础信息数据及路径规划信息，使

用高准确率的机器学习方法进行通勤出行方式的识别，最终可将所有个体的通勤出行方式判别为小汽车、公共汽车、轨道交通、自行车、步行五种类型。以北京市六环以内地区为研究范围，应用该框架挖掘街区层面的通勤出行方式，进行通勤出行方式与建成环境因子之间的关联性分析。

1 研究范围与数据

1.1 研究范围

本文研究范围为北京市六环高速公路以内地区，该地区总面积为2 267 km²，包括2 153个基于路网划分的基础研究单元。研究范围内覆盖北京市大部分常住人口、就业岗位和城市建设用地，构成了高度复杂的城市系统。

1.2 研究数据

本文主要使用互联网位置数据、兴趣点(Point of Interest, POI)数据、用户行为数据、常驻点数据、用户画像数据、路网数据共6类时空大数据，每类数据经过匿名化处理，确保本研究中不涉及个体隐私的相关信息。

1) 互联网位置服务数据。

百度地图日均位置服务请求次数超过1 200亿次，每月活跃智能设备数达到12亿台，涵盖全国各级行政区划。位置服务数据一方面用于计算小汽车、轨道交通、公共汽车、自行车、步行5类样本，另一方面用于挖掘通勤出行方式中定位点数量、速度中位数、最大和最小速度等特征。

2) POI数据。

POI数据一方面用于通勤特征的计算，包括公共交通可达性，例如与居住地、工作地最近的公共汽车站和轨道交通车站的距离，居住地、工作地附近的公共汽车站和地铁车站数量，通勤途中地铁和公共汽车站附近的定位数量等特征。另一方面，POI数据也用于计算研究单元的用地功能混合密度和轨道交通设施的服务水平。

3) 用户行为数据。

用户行为数据包括地图中“驾车、公交、轨道交通、骑行、步行”等各交通方式的使用频率，导航、路况、路径规划等功能的使用情况，用于构建各类通勤出行方式的使用特征。

4) 常驻点数据。

本文使用的常驻点数据由研究基准期向前回溯6个月的互联网位置数据挖掘得到,在处理过程中整合了去隐私化的位置、POI等多源数据类型,具有精度高、覆盖广的特点。基于这一数据,进一步计算用户通勤OD的空间分布、通勤距离等特征。

其中,根据常驻点数据提取得到的居住地和工作地信息是通勤出行研究的基础,在以往的多项通勤出行研究^[8,12-13]中,经过与统计数据、实地调查数据等其他各类数据源的对比表明,常驻点数据具有较高的精确度和有效性。通过计算用户居住地和工作地之间的空间距离可得到通勤距离。根据天津市居民出行调查数据与百度地图数据的比对验证,两者的吻合度较高,表明基于互联网位置数据计算得到的通勤距离分布具有较高的可信度。

5) 用户画像数据。

用户画像数据包括性别、年龄、资产状况、教育水平、消费水平、收入水平等,用于辅助判断用户的通勤出行方式。

6) 路网数据。

本文使用百度地图16级城市路网数据,一方面用于提取研究单元边界,另一方面用于计算各研究单元周边的路网密度指数。

2 研究方法

2.1 技术路线

本文首先通过部分标注数据,提取小汽

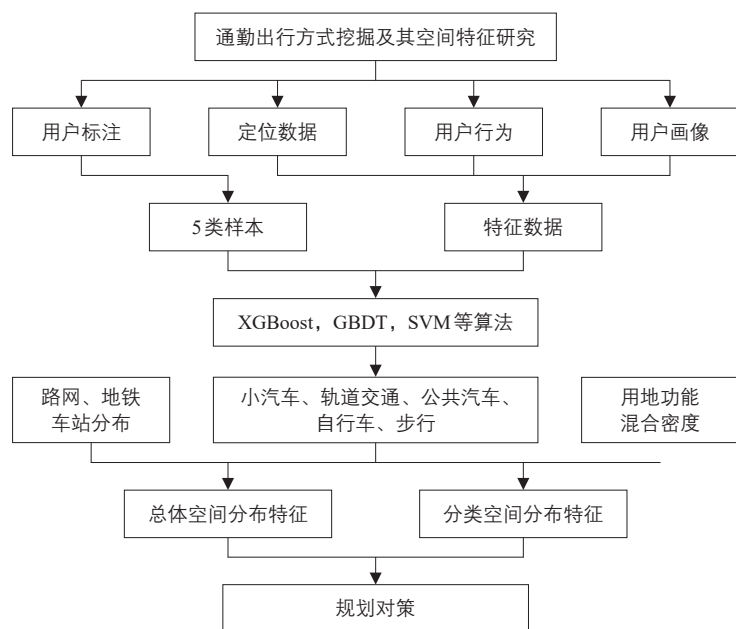


图1 通勤出行方式挖掘技术路线

Fig.1 Techniques of commuting travel modes analysis

车、轨道交通、公共汽车、自行车、步行5类样本。而后基于位置数据、用户行为和用户画像数据,提取通勤距离、定位速度、公共交通便利性等71个特征,使用极端梯度提升(eXtreme Gradient Boosting, XGBoost)、梯度提升决策树(Gradient Boosting Decision Tree, GBDT)、支持向量机(Support Vector Machine, SVM)等机器学习算法模型进行训练,得到每个用户的通勤出行方式,经验证计算准确率超过89%。结合地块空间单元、路网、轨道交通车站分布、用地功能混合等数据,分析不同通勤出行方式的空间分布格局,并探讨其与路网密度、用地功能混合密度、轨道交通设施服务水平等建成环境因子之间的内在关联。最终面向低碳交通、城市精细化治理等方向,提出具有操作性的规划、建设和管理建议(见图1)。

2.2 基于机器学习的通勤出行方式挖掘

基于互联网位置、POI、用户行为、常驻点、用户画像5类基础数据,构建通勤出行方式挖掘样本集,提取小汽车、轨道交通、公共汽车、自行车、步行5类样本集,总样本数量为42万个。

通过对样本集及其特征的分析,结合交通调查数据可以发现,采用不同通勤出行方式的用户在时空大数据中呈现出迥然不同的特点,这些差异构成了提取特征、构建分类数据集的基础。以通勤距离分布为例(见图2),不同出行方式之间存在明显差异,自行车和步行的通勤距离偏短,公共汽车和轨道交通的通勤距离偏长,小汽车通勤距离分布则较为均衡。据此,可以将通勤距离作为分类的一项特征。

采取这一思路,本文构建了4大类71项特征,深入刻画用户的通勤行为。4大类特征包括:7个定位特征、6个用户画像特征、45个用户行为特征、13个公交可达性特征。

在通勤出行方式挖掘算法的构建中,本文综合比较了贝叶斯、SVM、决策树、随机森林、GBDT和XGBoost等机器学习算法。评估结果表明,XGBoost算法的准确率和召回率均最高,整体精度超过87%(见图3),因此,本文最终选择此算法作为主要的通勤出行方式分类模型。

确定分类模型后,对用户通勤特征数据进行分类计算,挖掘各用户的通勤出行方式。将个体通勤出行方式数据按照其所在的

街区进行聚合，最终得到城市中各街区的居民通勤出行特征信息。

根据上述技术框架，本文对研究范围内全部街区的通勤出行方式进行计算。

2.3 通勤出行方式的空间影响因子分析

为探索通勤出行方式与城市建成环境之间的内在关联，本文计算了路网密度、用地功能混合密度、轨道交通设施服务水平3项因子(见表1)，并分析其与通勤出行方式之间的关系。

3 结果讨论

3.1 总体特征

基于上述方法，本文对研究范围内各街区的通勤出行方式进行计算，结果表现为街区各类通勤出行方式的比例(见图4)。从研究区域总体汇总数据来看，目前北京市各类通勤出行方式比例相对均衡，其中公共汽车通勤出行比例最高，约为26%，其次是小汽车通勤和轨道交通通勤比例，分别约为24%和23%。公共交通通勤出行比例接近一半，表明北京市公共交通系统发展日趋完善。与2014年居民出行调查数据^[17]相比，小汽车通勤比例下降约8个百分点，体现了城市在低碳交通方面的成就。在非机动车交通方面，自行车通勤出行比例为15%，与2014年相比略有上升，体现了城市在非机动交通环境改善、共享单车发展、城市职住结构优化等方面的综合提升。

3.2 通勤出行方式的空间区位特征

从空间分布来看，研究范围内不同通勤出行方式的分布具有明显的区位偏好。就小汽车通勤而言，距离城市中心越远，则小汽车通勤出行比例越高，且城市东部和北部地区的小汽车通勤出行比例略高于城市南部和西部地区(见图5)。非机动车通勤出行的空间分布特征则与之相反，距离城市中心越远比例越低，且城市二环快速路以内的中心地带及城市南部、西部地区的通勤出行比例显著高于二环快速路以外的东部和北部地区(见图6)。

可见，通勤出行方式的空间分布兼具同心圆和扇形两种空间分布模式，体现出距离衰减性和空间异质性的特征。从城市经济学理论来看，同心圆模式体现通勤成本对通勤

出行方式的影响。由于研究范围内的就业岗位主要集中在城市中心周边的若干地区^[18]，使得城市通勤出行结构大体上呈现中心-外围格局。随着居住地与城市就业中心的空间距离增加，就业人口的通勤距离和通勤时间随之上升，公共交通在便捷性、经济性上的优势有所下降，而小汽车的舒适性、灵活性优势则逐渐体现出来，导致其比例相对升高。扇形模式则体现了不同社会群体差异化的空间布局^[19]对通勤出行方式的影响，不同群体在职业、支付能力等方面存在差异，进

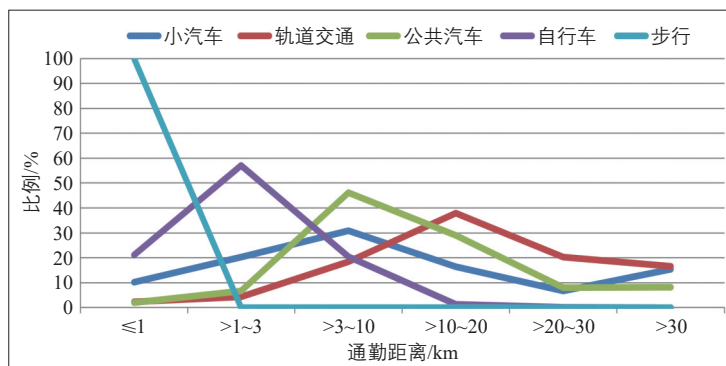


图2 样本通勤距离分布

Fig.2 Distribution of commuting distance

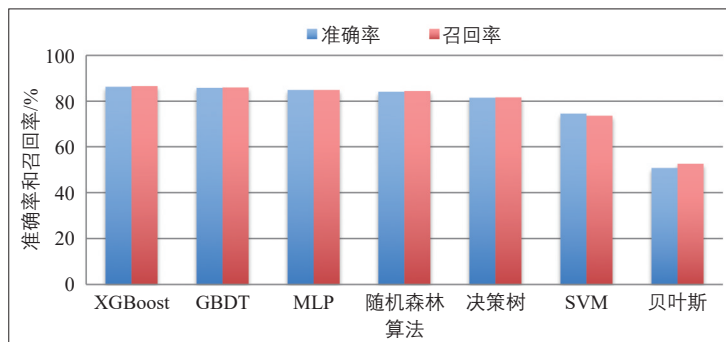


图3 通勤出行方式准确率和召回率评估

Fig.3 Accuracy and recall rates of commuting travel modes

表1 城市建成环境因子一览

Tab.1 Urban built environmental factors

因子名	因子含义
路网密度	本文将街区的城市路网密度界定为街区周边地区的路网线段核密度。线段核密度的计算方法及搜索半径的确定详见文献[14]。路网密度越低，则街区周边的路网越稀疏，一般而言街区尺度也较大
用地功能混合密度	本文以街区的加权功能混合密度来测度街区的用地功能多样性，具体计算方式：将POI的空间分布及其访问频次与地块及其主导功能数据结合，借助词频-逆文档频率(Term Frequency-Inverse Document Frequency, TF-IDF)算法，计算每个街区中各类POI功能点的贡献度；借助信息熵算法，将功能点贡献度聚合为街区的功能混合度，并使用地块面积对此指标进行标准化，最终得到各街区用地功能混合密度 ^[15]
轨道交通设施服务水平	本文用街区的轨道交通车站核密度值 ^[16] 来测度街区的轨道交通设施服务水平。该值越高，则街区周边轨道交通车站的分布密度越大，街区获取轨道交通设施服务的成本越低

而影响其通勤出行方式的选择。

3.3 高用地功能混合密度压缩小汽车通勤出行比例

对比图5和图7可以看出, 用地功能混合密度与小汽车通勤出行比例呈明显的反比

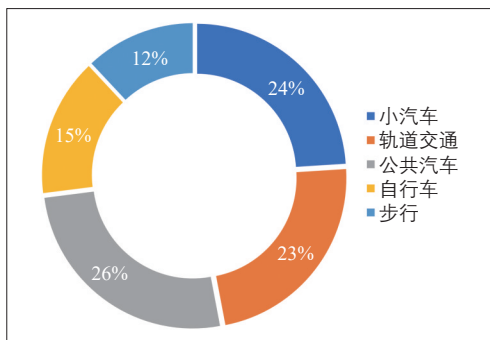


图4 北京市通勤出行方式构成

Fig.4 Commuting travel modes in Beijing

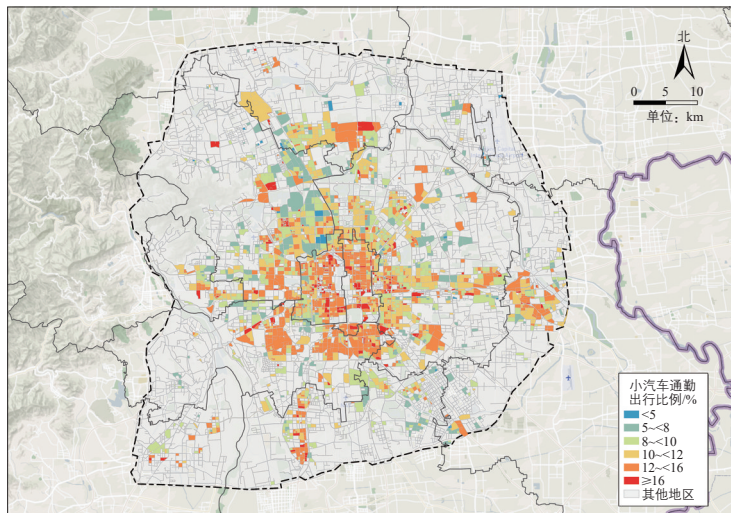


图5 小汽车通勤出行空间分布

Fig.5 Spatial distribution of car commuting trips

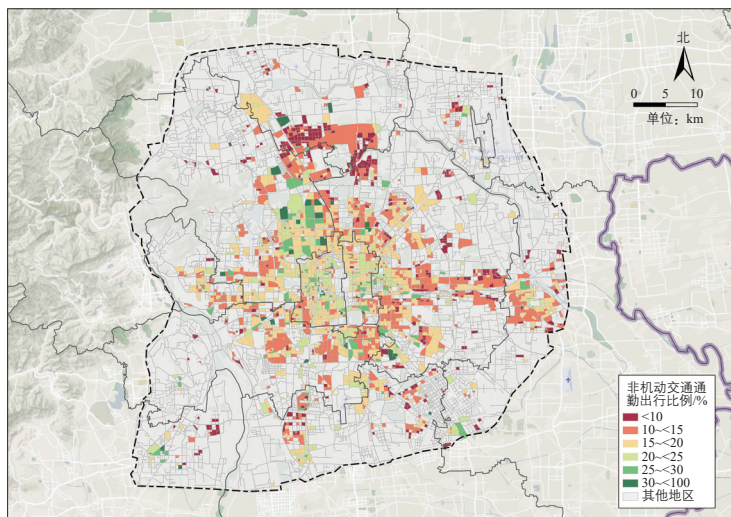


图6 非机动车通勤出行空间分布

Fig.6 Spatial distribution of non-motorized commuting trips

关系, 用地功能混合密度越高的地区, 小汽车通勤出行比例越低, 反之亦然, 这表明提高土地的混合利用率对压缩小汽车出行比例卓有成效。其原因可归结为两个方面: 一方面, 土地混合利用将一系列相互关联的功能紧凑地安排在同一区域内, 从而大大缩减出行成本, 降低市民的小汽车出行需求, 减少小汽车通勤的可能性; 另一方面, 多元化的城市功能可以有效促进居民就近就业, 在一定程度上减少通勤中使用小汽车的概率^[15]。

3.4 轨道交通对通勤出行的影响具有空间异质性

轨道交通服务供应充足时可以压缩小汽车通勤比例, 不足时效果不显著, 轨道交通服务水平对出行方式的影响更为复杂。从研究范围内的情况来看, 大致以地铁10号线为分界, 内外呈现不同的相关性。在10号线以内, 轨道交通车站覆盖度较高, 此时轨道交通通勤比例与车站的密度正相关, 小汽车通勤比例则与车站密度呈负相关关系; 而在10号线以外, 情况正好相反, 车站建设对压缩小汽车通勤作用不显著, 甚至车站周边地区小汽车通勤出行比例更高(见图8和图9)。

考虑通勤成本主要由时间成本、经济成本、舒适度成本等组成。地铁10号线以内地区可以获取相对快捷的交通服务, 且乘车环境较好, 区位较居中, 同时交通比较拥堵, 因此轨道交通通勤成本相对较小; 在10号线以外地区, 轨道交通成为大型住区建设的驱动力, 车站周边集聚了大量通勤人口, 而轨道交通的服务能力相对滞后, 使乘坐舒适度大大下降, 候车、换乘时间变长, 外部效应的存在使得轨道交通通勤成本反而高于小汽车通勤成本, 造成车站周边地区小汽车通勤出行比例“逆增长”。

3.5 高密度路网鼓励非机动车出行

与功能多样性类似, 高密度路网会压缩小汽车的出行比例, 并对非机动车产生正向激励(见图10和图11)。高密度路网具有更高的通达性, 为出行提供更多选择可能, 且在小尺度的街区非机动车环境相对较好, 非机动车通勤出行比例相对较高。高密度路网的交叉口、临街出入口均比较多, 周边用地功能较复杂, 人流集聚度较高, 客观上削弱了小汽车通勤的优势。这一发现与城市空间和城市交通领域的相关研究一致^[20]。

4 结语

本文针对传统通勤出行方式调查中存在的不足,提出一种基于时空大数据的通勤出行方式识别方法。该方法通过多源数据融合,使用XGBoost算法对用户的通勤出行方式进行挖掘。经验证,该方法具有较高准确率,同时具有覆盖面广、更新速度快、空间分辨率高等特征,是对传统方法的有效补充和强化。以北京市六环高速公路以内地区为研究范围,在街区尺度上对各类通勤出行方式的空间分布特征进行分析,进而探讨其与城市路网密度、用地功能混合密度、公共交通设施服务能力3项因子之间的相关性。分析结果表明,高用地混合、充足的轨道交通服务和高密度路网可以有效压缩小汽车通勤出行比例、激励非机动车通勤出行。根据这一结果,在城市建成环境的优化中应有针对性地提升街区的功能多样性,在通勤需求密集地区加强轨道交通设施的建设和提升,同时改善城市路网结构,营造更加良好的非机动车出行环境。通过上述举措,不断改善城市绿色交通水平,提升城市空间发展品质。

本文所使用的数据和模型主要针对居民通勤需求,在未来研究中将进一步拓展研究方法,涵盖城市其他出行需求。为此,首先要充实调查数据集,使之支撑出行方式计算中模型训练和验证的全过程;其次,整体出行需求较通勤出行需求更加复杂多样,本文将深入比较机器学习、深度学习的多种模型,提高挖掘精度;第三,将对出行方式与城市建成环境诸要素的关系开展更为深入、系统的研究,涵盖更加全面的要素类型,量化分析各类要素的影响力,以便为建设低碳城市提供更加坚实的决策支持。

参考文献:

References:

- [1] White M J. Urban Commuting Journeys Are Not “Wasteful” [J]. Journal of Political Economy, 1988, 96(5): 1097-1110.
- [2] Karanasiou A, Viana M, Querol X. et al. Assessment of Personal Exposure to Particulate Air Pollution During Commuting in European Cities: Recommendations and Policy Implications[J]. Science of the Total Environment, 2014, 490: 785-797.
- [3] 刘志林, 戴亦欣, 董长贵, 等. 低碳城市理念与国际经验[J]. 城市发展研究, 2009, 16

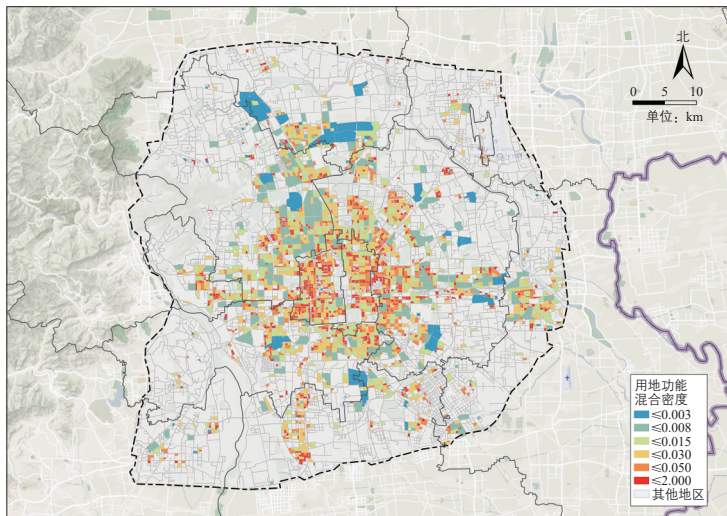


图7 用地功能混合密度空间分布

Fig.7 Spatial distribution of mix land use density

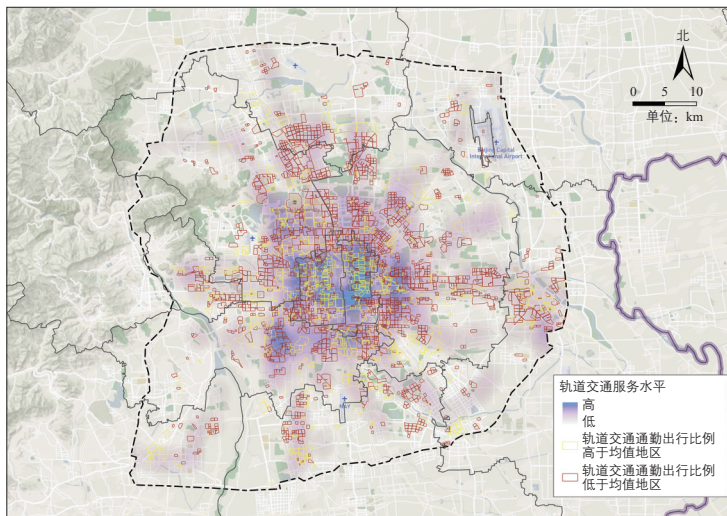


图8 轨道交通通勤出行比例与轨道交通服务水平的关系

Fig.8 Relationship between the proportion of rail transit commuting travel and level of service of rail transit

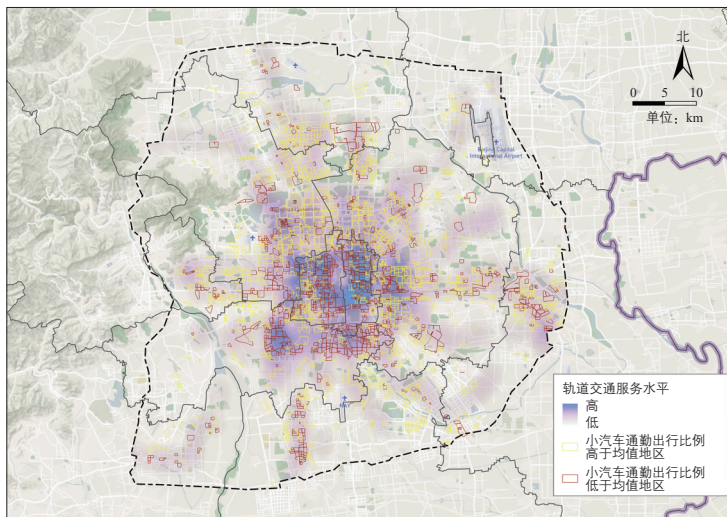


图9 小汽车通勤出行比例与轨道交通服务水平的关系

Fig.9 Relationship between the proportion of car commuting travel and level of service of rail transit

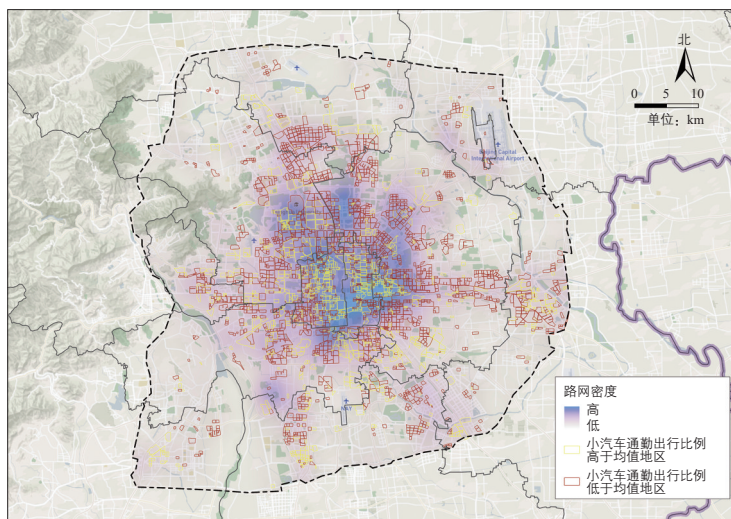


图10 小汽车通勤出行比例与路网密度的关系

Fig.10 Relationship between the proportion of car commuting travel and roadway network density

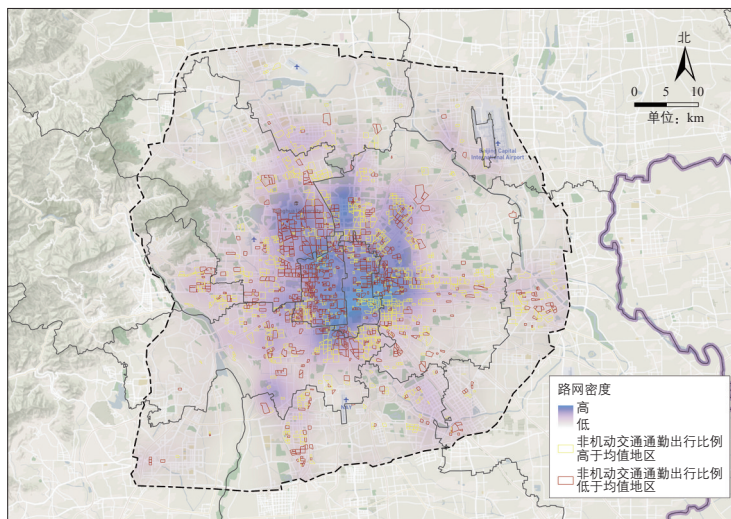


图11 非机动车通勤出行比例与路网密度的关系

Fig.11 Relationship between the proportion of non-motorized commuting travel and roadway network density

(6): 1-7+12.

Liu Zhilin, Dai Yixin, Dong Changgui, et al. Low-Carbon City: Concepts, International Practice and Implications for China[J]. Urban Development Studies, 2009, 16(6): 1-7+12.

[4] Susilo Y O, Maat K. The Influence of Built Environment to the Trends in Commuting Journeys in the Netherlands[J]. Transportation, 2007, 34: 589-609.

[5] Tammaru T. Suburbanisation, Employment Change, and Commuting in the Tallinn Metropolitan Area[J]. Environment and Planning A, 2005, 37(9): 1669-1687.

[6] Long Y, Zhang Y, Cui C. Identifying Commut-

ing Pattern of Beijing Using Bus Smart Card Data[J]. Acta Geographica Sinica, 2012, 67 (10): 1339-1352.

[7] 黄晓燕, 刘夏琼, 曹小曙. 广州市三个圈层社区居民通勤碳排放特征: 以都府小区、南雅苑小区和丽江花园为例[J]. 地理研究, 2015, 34(4): 751-761.

Huang Xiaoyan, Liu Xiaqiong, Cao Xiaoshu. Commuting Carbon Emission Characteristics of Community Residents of Three Spheres: A Case Study of Three Communities in Guangzhou City[J]. Geographical Research, 2015, 34 (4): 751-761.

[8] 天津市城市规划设计研究院. 2017年天津市居民出行调查报告[R]. 天津: 天津市城市规划设计研究院, 2017.

[9] 张天然. 基于手机信令数据的上海市域职住空间分析[J]. 城市交通, 2016, 14(1): 15-23.

Zhang Tianran. Job-Housing Spatial Distribution Analysis in Shanghai Metropolitan Area Based on Cellular Signaling Data[J]. Urban Transport of China. 2016, 14(1): 15-23.

[10] 蒋寅, 郑海星, 于士元, 等. 天津市职住空间分布与轨道交通网络耦合关系: 基于手机信令数据分析[J]. 城市交通, 2018, 16(6): 26-35.

Jiang Yin, Zheng Haixing, Yu Shiyuan, et al. Relationship Between Job-Housing Spatial Distribution and Rail Transit Network in Tianjin: An Analysis Based on Cellular Data [J]. Urban Transport of China, 2018, 16(6): 26-35.

[11] 钮心毅, 丁亮. 利用手机数据分析上海市域的职住空间关系: 若干结论和讨论[J]. 上海城市规划, 2015(2): 39-43.

Niu Xinyin, Ding Liang. Analyzing Job-Housing Spatial Relationship in Shanghai Using Mobile Phone Data: Some Conclusions and Discussions[J]. Shanghai Urban Planning Review, 2015(2): 39-43.

[12] 田轲, 李宇. 城市人口时空分布研究新视角新工具[R]. 深圳: 城市数据派, 2018.

[13] 青岛市城市规划设计研究院大数据中心. 百度大数据与多源数据的人口校核分析[R]. 深圳: 城市数据派, 2018.

[14] ESRI. How Kernel Density Works[EB/OL]. 2018[2020-05-03]. <https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst>.

(下转第67页)