



武汉大学学报(信息科学版)

*Geomatics and Information Science of Wuhan University*

ISSN 1671-8860,CN 42-1676/TN

## 《武汉大学学报(信息科学版)》网络首发论文

题目: 多源空间大数据场景下的家装品牌线下广告选址布局研究  
作者: 张嘉琪, 杜开虎, 任书良, 王瑞凡, 关庆锋, 陈文辉, 姚尧  
DOI: 10.13203/j.whugis20190468  
收稿日期: 2020-06-04  
网络首发日期: 2020-12-08  
引用格式: 张嘉琪, 杜开虎, 任书良, 王瑞凡, 关庆锋, 陈文辉, 姚尧. 多源空间大数据场景下的家装品牌线下广告选址布局研究. 武汉大学学报(信息科学版).  
<https://doi.org/10.13203/j.whugis20190468>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:10.13203/j.whugis20190468

## 引用格式:

张嘉琪, 杜开虎, 任书良, 等. 多源空间大数据场景下的家装品牌线下广告选址布局研究[J]. 武汉大学学报·信息科学版, 2020,

DOI:10.13203/j.whugis20190468 (ZHANG Jiaqi, DU Kaihu, REN Shuliang, et al. Research on Location Selection of Outdoor

Advertisement of Home Decoration Brand Based on Multi-source Spatial Big Data[J]. Geomatics and Information Science of

Wuhan University, 2020, DOI:10.13203/j.whugis20190468)

## 多源空间大数据场景下的家装品牌线下广告选址布局研究

张嘉琪<sup>1</sup> 杜开虎<sup>1</sup> 任书良<sup>1</sup> 王瑞凡<sup>1</sup> 关庆锋<sup>1</sup> 陈文辉<sup>2</sup> 姚尧<sup>1,3</sup>

1 中国地质大学(武汉)地理与信息工程学院, 湖北 武汉, 430074

2 浙江农林大学 信息工程学院, 浙江 杭州, 311300

3 阿里巴巴集团, 浙江 杭州, 311121

**摘要:** 合理进行线下广告牌投放位置的选择, 对商家宣传品牌以及扩大营销市场具有十分积极的作用。由于商业数据较难获取, 以往研究多停留在宏观理论层面, 未能对线下广告选址的实际布局进行细尺度分析。本研究以北京为研究区, 通过耦合某大型家装品牌户外广告到店转化率和路网、感兴趣点数据 (point of interest, POI) 等表征地理特征和商业经济特征的多源空间数据, 构建了基于随机森林的广告到店转化率预测模型 ( $Standard R^2=0.758$ ), 得到该品牌连锁家装商店在北京市广告适宜性空间分布结果, 并对各影响特征进行分析。研究表明: 该家装品牌线下广告到店转化率在北京整体呈现“中心高外围低”的格局, 且存在较强的空间自相关和高值聚集现象; 同时线下广告到店转化率与社会经济、商业政治和人群活动等具有较强相关性, 且对同一群体持续进行广告曝光的位置对到店转化率的影响较大。本研究所构建模型的精细制图结果, 可为线下广告牌布局和商业选址等相关研究提供参考依据和理论基础。

**关键词:** 多源数据; 户外广告; 随机森林; 地块特征集; 到店转化率

**中图分类号:** P28

**文献标志码:** A

**收稿日期:** 2020-06-04

**项目资助:** 国家重点研发计划(2019YFB2102903); 国家自然科学基金 (41801306, 41671408, 41901332); 武汉大学测绘遥感信息工程国家重点实验室开放基金(18S01); 湖北省自然科学基金(2017CFA041); 浙江省自然科学基金(LY18D010001)。

**第一作者:** 张嘉琪, 硕士生, 主要研究方向为城市计算。ningzetaolover@cug.edu.cn

**通讯作者:** 姚尧, 博士, 副教授。yaoy@cug.edu.cn

户外广告包括可以提供户外曝光的多媒体形式,如图片、文字和口头对话等<sup>[1]</sup>,它们的重要特性是能够吸引目标用户群体的注意<sup>[2]</sup>。在竞争日益激烈的市场,企业为了增加产品的曝光率和品牌影响力,将户外媒体这种重要的媒体形式作为优先选择的对象<sup>[3]</sup>。为了最大限度地提高户外广告的效率,广告应该有针对性,即“在适当的时间以适当的形式向适当的人提供适当的内容”<sup>[4]</sup>。户外广告的成功与许多因素有关,如位置选择、内容设计、可见度,其中位置选择是最重要的因素<sup>[5]</sup>。

目前户外广告的位置选择无法满足消费者的地理集中度及购物需求,并未广泛应用于企业及市场<sup>[1]</sup>;同时,公共压力和政府监管也使得户外广告的投放量大幅减少,导致商家对户外广告所带来的经济效益提出质疑<sup>[6]</sup>。因此,选择合理的户外广告位置具有极其重要的意义,不仅可以避免盲目投放的浪费,而且对商家产品销售和品牌宣传都会产生积极的影响。

已有许多学者对户外广告的选址布局进行了相关研究。传统的户外广告选址多基于调查数据和经典条件作用进行户外广告和变量间的相关性研究。如 Jimenez 等通过酒精广告牌的调查数据研究广告与居民年龄、收入等的相关性<sup>[7]</sup>; Jennifer 等则通过烟草广告的调查数据研究广告与社区人口、烟草价格等的相关性<sup>[8]</sup>;此外随着空间统计学的发展,许多学者将广告位置和区域与购买人群相联系,如 Rosewarne 等研究探讨户外广告内容与区域的空间特征间的相关性<sup>[9]</sup>; Wilson 等设计对照实验对广告位置对人们购买意愿的影响进行研究分析<sup>[6]</sup>。

早期的研究方式往往依赖于主观记忆,而主观记忆可能会因时间而被扭曲,并且有些研究数据需要专业机构进行采集,费时费力,灵活性较低<sup>[10,11]</sup>。随着信息技

术的发展和多源空间数据的出现,广告选址布局的相关研究有了新的思路。如 Liu 等基于大规模出租车全球定位系统(Global Positioning System, GPS)轨迹通过可视化分析进行广告牌位置选择<sup>[5]</sup>; Anagnostopoulos 等利用社交媒体 Twitter 数据挖掘用户兴趣来进行广告牌选址研究<sup>[12]</sup>; Huang 等耦合用户手机上网数据和位置数据进行了广告选址布局相关研究<sup>[10]</sup>。

但这些研究仍然存在一些问题,数据的有效性低导致研究的尺度较大,无法进行精细尺度的分析。由于缺乏数据和模型,大多研究只考虑影响广告效果的因素和用户兴趣,未能对线下广告实际布局和产生的效果进行更进一步的探讨。针对以往研究的不足,本研究利用某移动应用程序,跟踪了人口位置信息,获取了某家装品牌广告到店转化率数据,结合影响选址地块属性的多源空间数据,基于随机森林模型,构建一个有效的数据挖掘框架。并以此进行研究区内细尺度的广告到店转化率预测,同时对户外广告选址的影响因子进行分析。

## 1 研究区与数据

### 1.1 研究区

北京市地处中国华北地区,是我国的首都以及政治、文化中心,总面积 16,410.540 平方千米,常住人口 215,420,000 人。如图 1 所示,北京市共辖 16 个市辖区,包括中心主城区:东城区、西城区、海淀区、朝阳区、丰台区;近郊区:大兴区、通州区、顺义区、昌平区、门头沟区、石景山区、房山区;远郊区:怀柔区、平谷区、密云区、延庆区<sup>[13]</sup>。

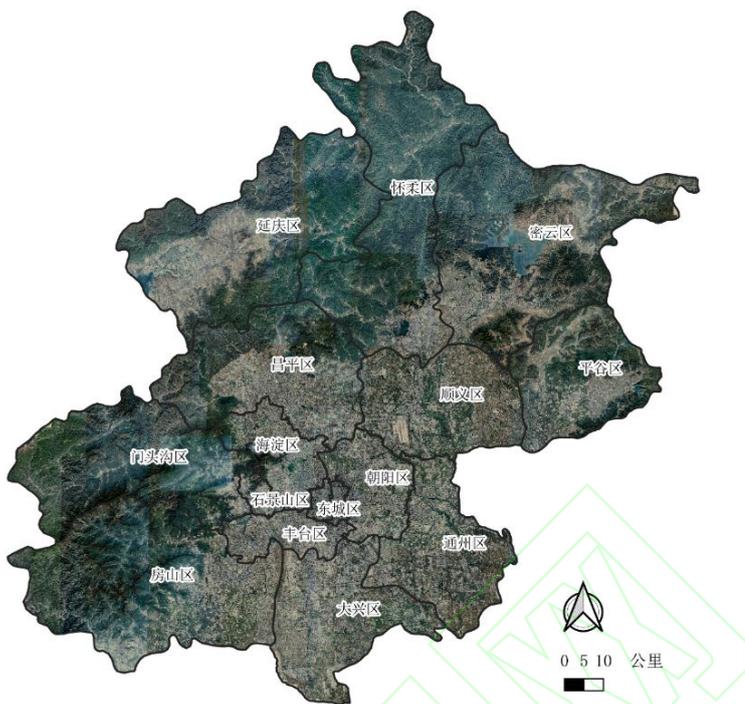


图 1 北京市行政区划图

Fig. 1 Administrative Map of Beijing

由于政府支持、国家政策等因素，北京市下辖各区经济发展水平存在差异，较为发达的辖区主要分布在中心主城区。以 2017 年为例，五个中心主城区的地区生产总值之和约占当年全市生产总值的 68%<sup>[14]</sup>。

### 1.2 研究数据

到店转化率是本文研究最重要的数据。该数据集来自一款移动应用程序，在北京市用户渗透率达 80% 以上，保证了数据的可靠性<sup>[15]</sup>。该应用程序根据用户隐私协议，采集用户的位置信息。若用户见到广告牌并于 3 个月内至广告牌所属店内消费，则该用户被判定为广告转化成功用户。转化成功用户和看到广告牌的用户数量比例，

即为本研究采用的到店转化率数据。

由于原始到店转化率数据与研究区域网格间不完全重合，对数据进行重采样处理，得到网格化后的转化率数据集。由于采集的到店转化率是连续数值且存在冗余过高、部分区域异常的问题，本文通过对它们进行标准差等级划分来进行离散化处理<sup>[16]</sup>。设研究区内每个网格内到店转化率的平均值和标准差分别为 $\mu$ 和 $\sigma$ ，在数据预处理中，为了保证数据的可靠性，删除低于 $\mu-3\sigma$ 或者高于 $\mu+3\sigma$ 的样本数据<sup>[17]</sup>，并将到店转化率数据进行归一化。图 2 显示了北京市该家装品牌到店转化率的空间分布情况。

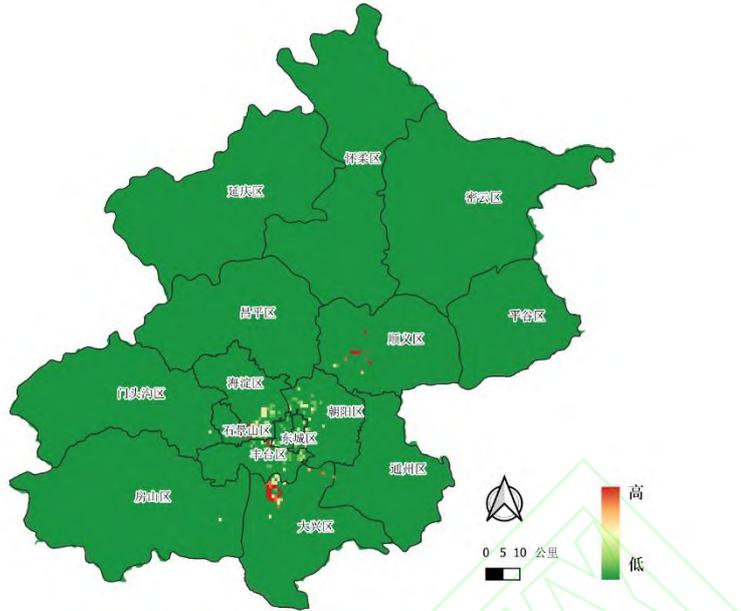


图 2 北京市某家装品牌到店转化率分布

Fig. 2 Distribution of the in-store conversion rate of a home decoration brand in Beijing

除到店转化率外，本研究用到的多源空间大数据还包括道路网数据、感兴趣点 (point of interest, POI) 数据、房价数据及其他基础地理数据。北京市道路网数据来源于 OpenStreetMap(OSM)网站 (<https://www.openstreetmap.org/>)。OSM 数据库存储了一个巨大的环境空间数

据集，其中包含了非常丰富的元数据，覆盖了世界上的主要城市<sup>[18]</sup>。本研究在对原始 OSM 数据集进行清洗后，提取出一级公路、二级公路和三级公路三个类别。图 3 为北京市 OSM 路网分布。

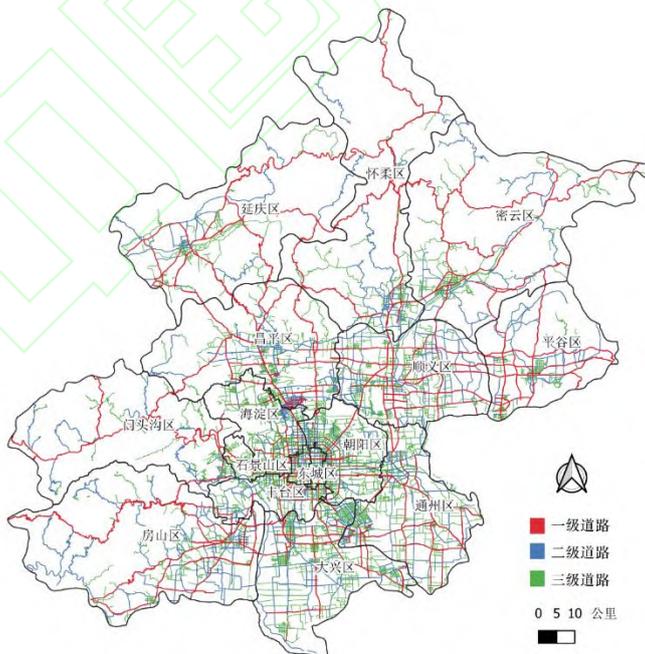


图 3 北京市 OSM 路网

Fig. 3 OSM road network distribution in Beijing

POI 是指具有地理标识的空间特征物，包含名称、类别、经纬度等信息，是空间大数据分析的基础性数据。POI 信息是数据供应商通过实地调查采集的，信息丰富

而准确，更新及时<sup>[19]</sup>。本研究所使用的 POI 数据来自中国最大的地图服务商之一高德地图服务 (<https://lbs.amap.com/>)。共获取到 701,494 条高德 POI

记录，包括公司企业、餐饮服务、购物服务、交通设施、金融服务、科教文化、汽车服务、汽车维修、汽车销售、商务住宅、生活服务、体育休闲、停车场、政府机构、商场专卖店 15 种类别。采用核密度估计法，计算离散

点在周边领域内的密度<sup>[20]</sup>，使得研究对象密度呈现空间连续的变化<sup>[21]</sup>，生成相应密度数据集，能够表达 POI 的分布模式和分布密度等特征<sup>[22]</sup>。图 4 显示了这 15 类 POI 数据的核密度分布情况。

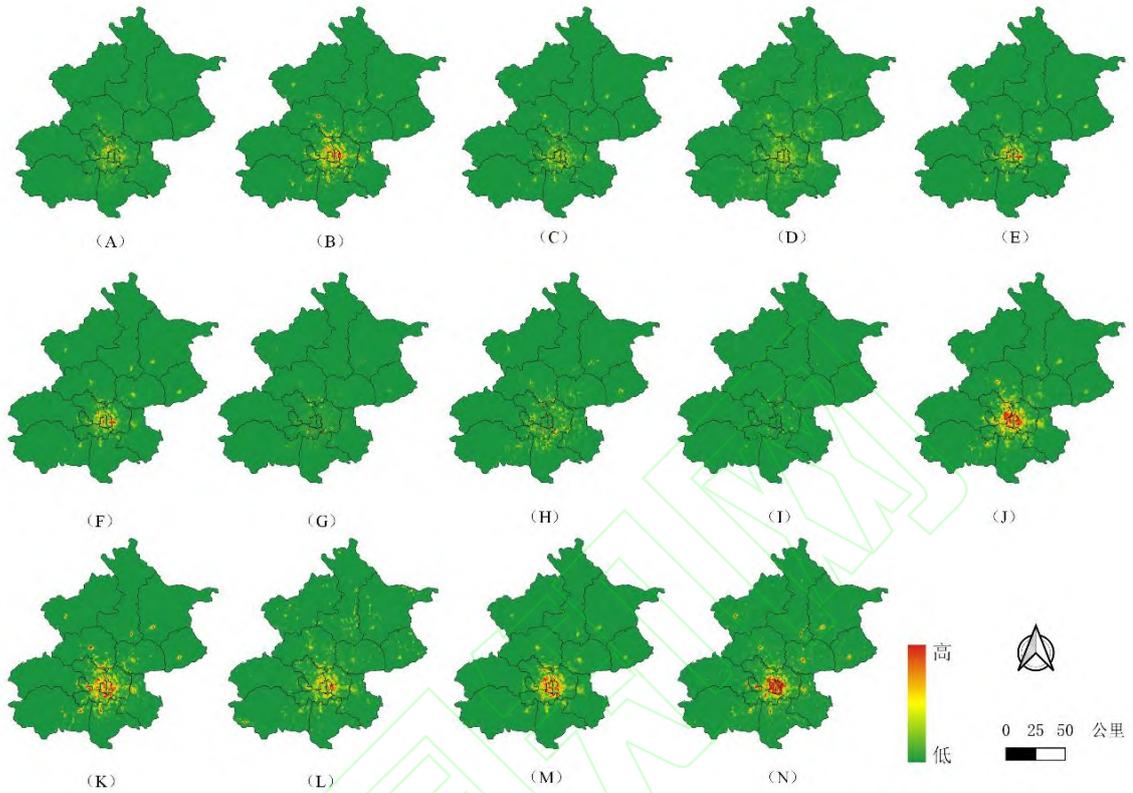


图 4 北京市各类 POI 核密度分布

(A) 公司企业；(B) 餐饮服务；(C) 购物服务；(D) 交通设施；(E) 金融服务；(F) 科教文化；(G) 汽车服务；(H) 汽车维修；(I) 汽车销售；(J) 商务住宅；(K) 生活服务；(L) 体育休闲；(M) 停车场；(N) 政府机构

Fig. 4 Kernel Density Distribution of All Kinds of POI in Beijing

(A) incorporated business; (B) catering service; (C) shopping service; (D) traffic facilities; (E) financial service; (F) education and culture; (G) automobile service; (H) auto repair; (I) auto sales; (J) commercial housing; (K) life services; (L) sports leisure; (M) parking lot; (N) governmental agencies

北京市房价数据来源于中国最大的在线住宅市场网站房天下 (<https://www.fang.com/>)。房价和收入被认为是由稳定的长期关系联系在一起的，在我国城市住宅市场中，城市房价与城市平均收入水平正相关，城市居

民的劳动收入水平越高，或者城市生活质量越好，城市的房价就会越高<sup>[23, 24]</sup>。在对原始数据进行清理后，采用反距离加权插值法将原始房价插值<sup>[25]</sup>。图 5 为北京市平均房价分布。

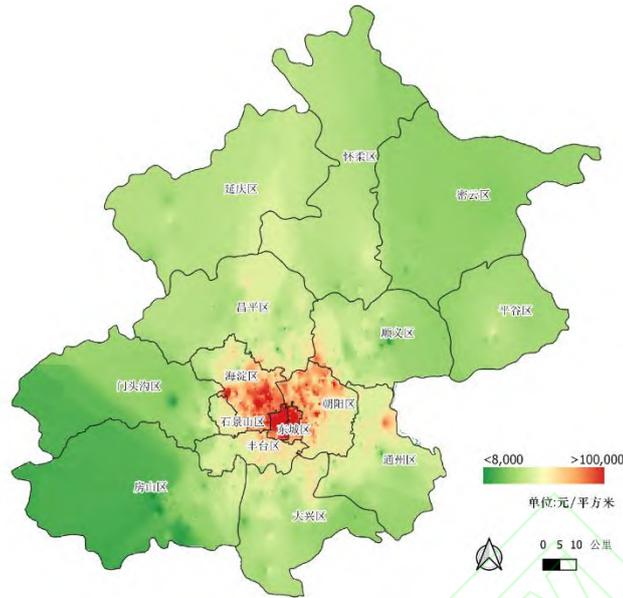


图 5 北京市平均房价分布

Fig. 5 Average Housing Price Distribution in Beijing

## 2 研究方法

本文基于随机森林模型建立多源空间数据与户外广告到店转化率间的相关关系。在多源空间数据影响特征集的选择上，基于前人研究的经验，参考了 Yao 等以及 Guo 等在商业店铺选址中所使用到的特征数据集，包括对应位置的居民属性和商业经济属性两个方面<sup>[26, 27]</sup>，共选取了包括路网数据、POI 数据和基础地理数据共 19 个空间变量作为影响特征集。前人研究表明，将研究区

进行网格级别的划分能够轻松地聚合或分解多源数据，促进了数据集成<sup>[28]</sup>，同时能够对研究区不同位置的比较提供统一的尺度<sup>[29]</sup>。在网格分辨率上，采用了 Zhou 等和 Zheng 等在研究中使用的 1km×1km 大小<sup>[30, 31]</sup>，这样的网格划分能够同时避免因网格太小而没有足够特征，以及网格太大而导致研究过于粗糙的问题。拟合研究区内的户外广告到店转化率得到广告适宜性制图结果，给出各影响特征的重要程度并对结果进行分析。本文相关技术路线如图 6。

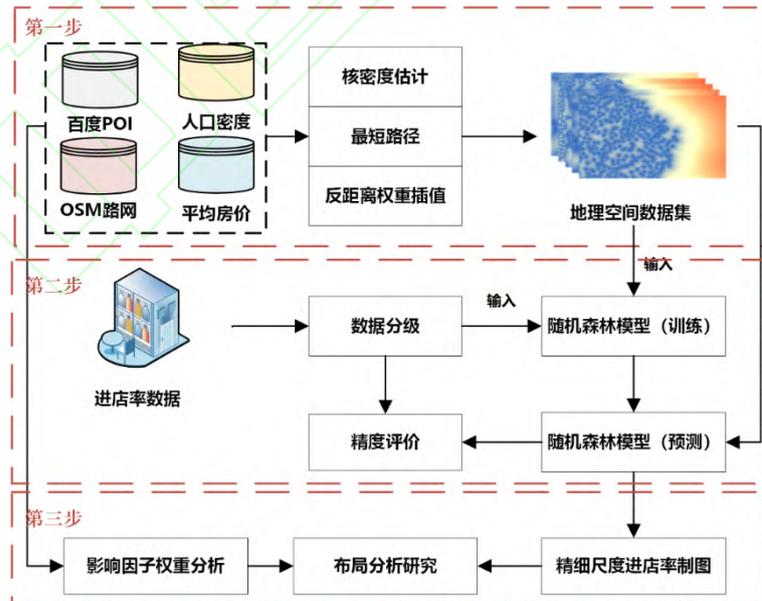


图 6 本文技术路线

Fig. 6 Experimental Flowchart of This Study

### 2.1 基于随机森林模型的户外广告选址定位

本研究基于随机森林 (Random Forest, RF) 模型，

刻画广告到店转化率与对应位置的地理特征及商业经济特征间的复杂相关关系。随机森林是目前广泛应用于文献和应用领域的一种集成学习方法，前人的研究证明

随机森林模型用于空间分布预测研究具有很好的效果，与其他预测模型相比精度更高<sup>[32-34]</sup>。随机森林模型在训练过程中采用的是 Bootstrap 重新抽样的方法，从原始样本中抽取多个样本，对每个 bootstrap 样本构建决策树，然后将所有决策树的预测平均值作为最终预测结果<sup>[35]</sup>。具体的随机森林模型构建过程如图 7 所示。

在构建模型时，一些不相关的变量可能会对模型预测精度产生负面影响。变量选择技术，也称特征选择，能够消除不相关的变量，达到模型精度改进，规避过拟合现象<sup>[36]</sup>。递归特征消除法（Recursive Feature Elimination）是用一个准则来消除特征的循环过程，它包括三个步骤：(1)训练分类器；(2)计算排序准则；(3)去除排序得分最小的特征<sup>[37]</sup>。

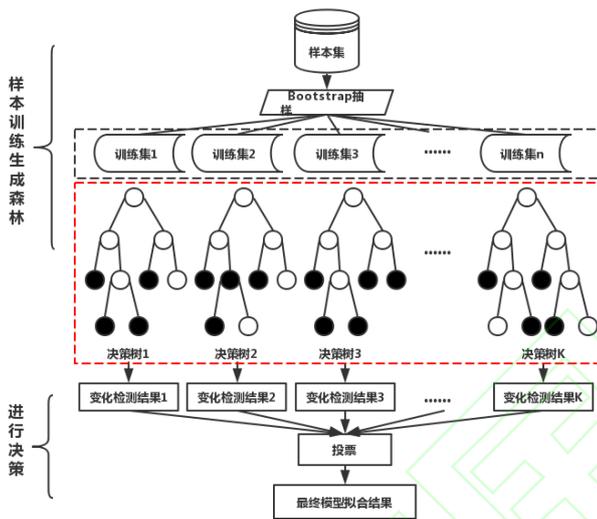


图 7 随机森林模型构建流程

Fig. 7 Construction Process of Random Forest Model

对商家店铺位置，与 OSM 路网叠加，计算其到道路的最近欧式距离，并与 1km×1km 的空间网格匹配，得到道路距离特征数据集。其余特征集均与 1km×1km 的细尺度网格进行匹配，POI 数据、房价数据分别以网格内的核密度和平均房价作为影响特征集。本研究通过递归特征消除法建立基模型进行多轮训练，从多源空间数据集中构建出 19 类特征集，分别为公司企业密度、餐饮服务密度、购物服务密度、交通设施密度、金融服务密度、科教文化密度、汽车服务密度、汽车维修密度、汽车销售密度、商务住宅密度、生活服务密度、体育休闲密度、停车场密度、政府机构密度、房价、路网、道路距离、商场专卖店密度和人口流动。

## 2.2 广告转化率的驱动因素分析

变量在回归中的重要性是应用统计学中的一个重要课题。随机森林模型能够处理大量变量并且提供对变量重要性的评估<sup>[38, 39]</sup>。随机森林模型进行变量重要性评

估具体过程如下：每个采样集建立一棵回归决策树，用袋外数据（out of bag, OOB）测试集来预测，得到 OOB 残差均方；变量 $x_i$ 在  $b$  个 OOB 样本中随机置换形成一个新的 OOB 集合，用已经建立的随机森林对新的 OOB 进行测试，得到随机置换后的 OOB 残差均方  $MSE_{ij}$ ，生成矩阵  $A$  为：

$$A = \begin{bmatrix} MSE_{11} & \dots & MSE_{1b} \\ \vdots & \ddots & \vdots \\ MSE_{p1} & \dots & MSE_{pb} \end{bmatrix} \quad \#(1)$$

其中  $p$  为影响因素变量个数， $b$  为样本个数。变量  $x_i$  的重要性评分  $Score_i$  可表示为：

$$Score_i = \left( \sum_{j=1}^n (MSE_j - MSE_{ij}) / b \right) / S_E \quad \#(2)$$

其中  $1 \leq i \leq p$ ， $MSE_j$  为第  $j$  个样本的残差均方， $S_E$  为标准误差。

## 2.3 模型可靠性评估

本研究是基于多源空间数据进行户外广告选址适宜性制图，为了反映户外广告到店转化率空间化模型的精度水平，采用三种精度评价指标，将模型预测的广告到店转化率结果和实际情况进行对比来定量评估本文提出的模型准确性。分别为：拟合优度（Standard  $R^2$ ），均方根误差（RMSE）和平均绝对误差（MAE）。计算公式如下：

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad \#(3)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad \#(4)$$

$$Standard R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad \#(5)$$

其中  $y_i$  和  $\hat{y}_i$  分别是第  $i$  个样本广告到店转化率的实际和预测值， $m$  是研究区域内的总样本。

## 3 结果

### 3.1 模型参数设置和精度分析

本文以北京市某家装品牌户外广告到店转化率作为因变量，公司企业、交通设施等 19 个影响因子作为协变量，构建基于随机森林的该家装品牌户外广告到店转化率拟合模型。对北京市进行网格化，得到 180\*220 个 1 km×1 km 网格，并对相应数据进行预处理共得到

20,871 组样本, 随机划分为 80% 的训练数据, 20% 的验证数据。通过 20 次随机训练和预测, 将北京市某家装品牌广告选址适宜性制图结果与真实有效广告转化率数据进行对比分析, 并进行准确性评估, 结果如表 1 所示。结果具有较高的整体精度, *Standard R<sup>2</sup>* 为 0.758, *MAE* 为 0.099, *RMSE* 为 0.018。表明预测值和实际值的差异小且相关程度高, 预测精度较高, 模型的预测效果较好。说明本研究所提出的方法能够有效的挖掘多源空间大数据与该家装品牌户外广告到店转化率之间的关系, 模型能够为户外广告选址提供适宜性的参考。

表 1 模型精度结果

Tab. 1 Accuracy of the Model

精度评价指标	<i>Standard R<sup>2</sup></i>	<i>RMSE</i>	<i>MAE</i>
结果	0.758	0.018	0.099

### 3.2 家装品牌线下广告选址适宜性空间分布

将提取的北京市特征数据利用该模型拟合北京市

某家装品牌户外广告到店转化率, 最终得到制图结果如图 8 所示。

地理学第一定律表明: 所有事物都与其他事物相关, 但是近处的事物比远处的事物更相关<sup>[40]</sup>。为了研究广告牌位置选择与地理位置间是否具有相关性, 本研究对模型得到的家装品牌户外广告到店转化率引入空间自相关分析。计算得到全局 *Moran's I* 指数为 0.641, 对应的 *p* 值为 0.000, *z* 得分为 93.584。 *p* 值小于 0.01 表明具有统计学上的显著性, *z* 得分远大于对应的临界值 2.58, 说明该家装品牌户外广告到店转化率呈现集聚现象, *Moran's I* 大于 0.6 则表明存在明显的空间正相关。为进一步研究分布是否存在高/低值聚类的现象, 得到 *General G* 指数, 对应的 *p* 值和 *z* 得分分别为 0.000 和 92.368, 表明具有统计学上的显著性。*Observed General G* (0.001) 大于 *Expected General G* (0.000), 说明该家装品牌户外广告到店转化率存在高值区域聚类的现象。

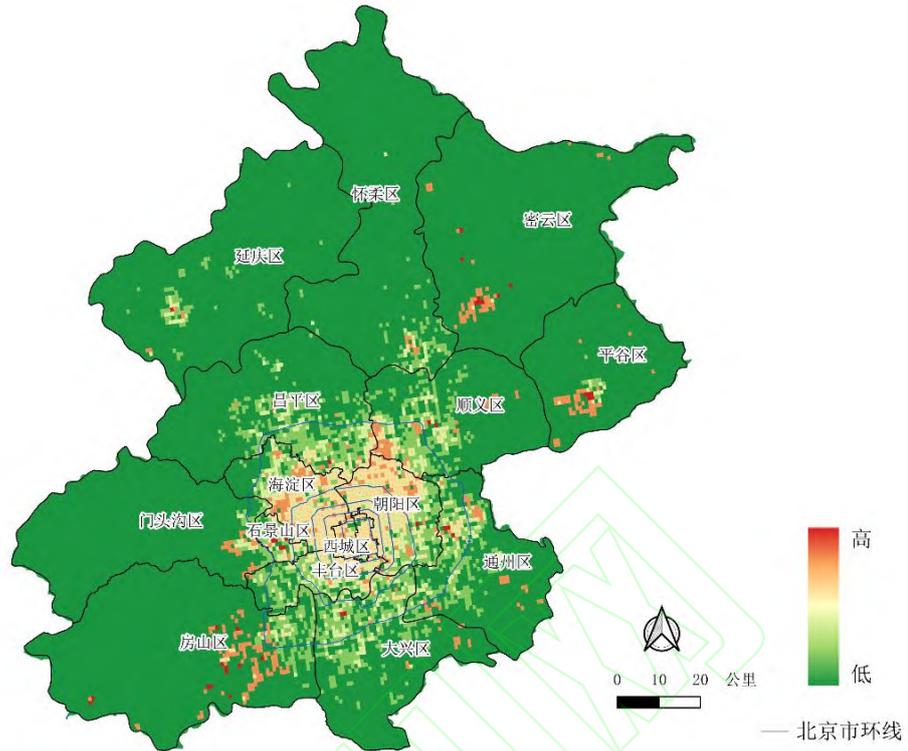


图 8 北京市某家装品牌户外广告选址适宜性制图结果

Fig. 8 Suitability Mapping Results of Outdoor Advertising Site Selection

上述分析结果表明该家装品牌户外广告到店转化率的分布存在较强的空间正相关特征，呈现明显的高值聚集现象。从图 8 中我们可以发现，该家装品牌户外广告到店转化率的空间分布整体呈现“中心高外围低”的格局，对应的中心处呈现高值聚集现象。这和地区实际经济情况符合。转化率高的区域主要位于北京市的中心主城区以及近郊区内靠近主城区的区域，而外围的远郊区和近郊区内远离中心城区的部分转化率则较低，并且高低分界与北京市环线基本一致。

弗里德曼将经济系统划分为中心和外围两部分，中心区发展条件较优越，经济效益较高，处于支配地位，而外围区发展条件较差，经济效益较低，处于被支配地位<sup>[41]</sup>。北京也是如此。2018 年的统计数据显示，2017 年北京市五个中心主城区的地区生产总值之和约占当年全市生产总值的 68%，而远郊区仅占 3%<sup>[14]</sup>。

从空间分布上看，由于自然地理条件不同、政府资源配置差异和部分区县优势产业<sup>[42]</sup>，北京的经济发展差异较大，五环线内经济发展良好，交通发达，人流量巨大，金融业高度聚集在北京中心城区的核心地带，并且聚集程度持续提升，科技、信息等生产者服务业也高度

集中于中心圈层，比如海淀区的东南部集中了较多的大型公司企业，其中包括百度大厦、中软国际、字节跳动等多家大型公司，这对商业发展和经济发展十分有利。

为了进一步验证分析，如图 9 所示，本文选取了几个细尺度区域（包括中心城区海淀区、近郊区房山区以及远郊区怀柔区）进行进一步分析。

可以看出，三个地区的家装品牌广告到店转化率符合整体“中心高外围低”的格局。其中，海淀区中关村软件园处，转化率较高，从地理位置上看，中关村软件园一带是由商务区和研发区组成的中国著名软件基地，园区东临上地信息产业基地，南靠规划绿化带及北大生物城，园区的边缘东南两侧，建有软件广场、信息中心、孵化器、体育场馆、宾馆、公寓等配套设施，各种基础设施配套齐全。

房山区的万科半岛广场，虽然非物质文化遗产丰富，传统文化历史氛围浓厚，但地方人民政府注重于打造社区文化品牌，忽视了公司企业、交通设施等的发展，整体上经济水平稍显不足，广告到店转化率较低；而怀柔区的滨河公园，其以林地、绿地为主体，各类设施配置存在着明显的不足，广告到店转化率更低。



图 9 北京市街区到店转化率分析

(A)海淀区中关村软件园街景；(B)房山区万科半岛广场街景；(C)怀柔区滨河公园街景

Fig. 9 The Analysis Results of in Store Conversion Rate in Beijing

(A) street view of ACEIT in Haidian District; (B)street view of Vanke peninsula square in Fangshan District; (C) street view of Binhe park in Huairou District

### 3.3 广告转化率驱动因素分析

通过对广告适宜性选址进行定性分析，通过制图直观地掌握广告选址的适宜性，但还需要进一步对广告选址到店转化率的各驱动因素定量分析。模型给出的各变量重要性评估如表 2 所示，值越大表明该变量对广告到店转化率的影响越大。

表 2 特征重要性排序

Tab. 2 Ranking of Feature Importance

特征重要性	变量类别	影响因子值
公司企业	POI	0.424
交通设施	POI	0.104
购物服务	POI	0.089
人口流动	社会经济数据	0.052
停车场	POI	0.051
政府机构	POI	0.029
房价	社会经济数据	0.027
商务住宅	POI	0.025
路网	基础地理数据	0.023
汽车维修	POI	0.022
汽车销售	POI	0.022
道路距离	基础设施数据	0.021
金融服务	POI	0.019

特征重要性	变量类别	影响因子值
科教文化	POI	0.019
汽车服务	POI	0.016
商场专卖店	POI	0.016
餐饮服务	POI	0.016
生活服务	POI	0.014
体育休闲	POI	0.011

可以看出，基于随机森林模型构建的户外广告选址适宜性评估方法中，社会经济、商业政治和人群活动较为重要。其中，重要性最高的特征为公司企业(0.424)，广告选址适宜性空间分布的结果也证明了这一点。企业分布较多的区域人口密度相对较大，并且从接触广告的持续性来看，企业分布地是一个可以对工作群体进行长期固定宣传的区域。这增加了户外广告对企业工作人群的曝光度，更容易强化广告的宣传效果。同时，企业较多的区域也有很大的商业潜力，这为广告宣传效果转化为到店率提供了更大的可能。交通设施、购物服务的影响程度分别排在第二(0.104)、第三位(0.089)，两个特征集从侧面反映了社会经济水平以及基础设施建设水平，同时交通设施和购物服务场所往往人流量较大，受众较多，这与重要性程度紧随其后的人口流动(0.052)形成呼应。

影响程度较小的因素，如生活服务(0.014)和体育

休闲 (0.011), 这类场所多有特定的功能, 无法对某一群体进行长期固定的宣传, 广告转化效果较差。

因此我们可以得出, 在进行户外广告选址或商业选址时, 除了地理区位上的商业聚集区域和人群活动活跃区域, 还应关注广告位置的持续影响时间效果。公司企业这种既满足商业经济发达区位条件, 又可以对工作群体进行持续宣传的位置能够强化广告宣传效果, 提升广告的到店转化率, 更好的发挥户外广告的作用。

## 4 讨论和结论

以往对线下广告的研究由于缺乏数据和模型, 未能对选址的实际布局进行细尺度分析, 同时也未能发现商家的经济效益与地块属性之间的相关性。针对数据和模型缺乏问题, 本研究耦合多源空间数据, 构建了基于多源空间大数据的户外广告有效到店转化率拟合模型, 并对北京市某家装品牌进行户外广告选址适宜性制图。提出的方法有效地融合了自然地理特征和社会经济特征, 具有较高的拟合精度 ( $Standard R^2=0.758$ ), 能够准确地反映出北京市该家装品牌户外广告选址适宜性和转化率的分布差异。

通过驱动因素的重要性分析发现, 地理位置因素对户外广告选址有着非常重要的作用。本研究发现社会经济、商业政治和人群活动对户外广告到店转化率的影响较大。此外, 广告位置的持续影响时间效果也很重要。公司企业这种能够对某一群体持续曝光宣传的位置能够强化广告宣传效果, 提升广告的到店转化率。同时, 户外广告选址适宜性制图结果说明北京市五环线内该家装品牌户外广告到店转化率整体较高, 并且存在明显的高值聚集现象, 进一步说明了社会经济发展和地理位置对户外广告选址的适宜性有很大的影响。

本研究以某大型家装品牌为例, 首次尝试将广告到店转化率应用于户外广告的选址研究中。通过该模型能够定量评估各位置的广告效果, 为户外广告的选址布局提供决策支持, 使广告的效率最大化。该模型可为其他行业或品牌的广告选址提供一定参考。但由于本文所用重要数据源 (该家装品牌的户外广告到店转化率) 存在行业间偏差, 且不同行业的广告选址影响特征还应适当考虑行业独特性。不同行业的客户受众、企业选址偏好和相应的竞争品牌也存在一定差异。这是我们今后需要考虑的重点。因此后续还需针对其他品牌或者行业的户外广告进一步研究, 以得到更为普适的模型和结果。

此外, 本研究中多源空间数据集的类型仍不完备, 由于城区和郊区的社会经济和人群活动因素存在差异, 可能会造成在缺少数据的郊区区域模型拟合精度不高。因此, 后续研究将引入迁移学习模型, 将城区和郊区的共性和区域性特征提取构建迁移模型<sup>[26]</sup>, 力求得到更高

精度的模型为户外广告的选址提供更加精确的参考。

## 参考文献

- [1] Lichtenthal J D, Yadav V, Donthu N. Outdoor advertising for business markets[J]. *Industrial Marketing Management*. 2006, 35(2): 236-247.
- [2] Ren A, Nesrin A E. Interaction of Outdoor Advertising Improved by Innovative Methods with Digital Art[J]. *Procedia Social & Behavioral Sciences*. 2015, 195(3): 799-805.
- [3] Zhu Jun, Chen Jingliang, Zong Liyong. The Impact of Outdoor Advertising Information Attributes on Consumers' Purchase Intention[J]. *Forecasting*. 2019, 38(4): 32-38(朱军, 陈敬良, 宗利永. 户外广告信息属性对消费者购买意愿的影响研究[J]. *预测*, 2019, 38(4): 32-38)
- [4] Tam K Y, Ho S Y. Understanding the Impact of Web Personalization on User Information Processing and Decision Outcomes[J]. *Mis Quarterly*. 2006, 30(4): 865-890.
- [5] Liu D, Weng D, Li Y, et al. SmartAdP: Visual Analytics of Large-scale Taxi Trajectories for Selecting Billboard Locations[J]. *IEEE Transactions on Visualization & Computer Graphics*. 2016, 23(1): 1-10.
- [6] Wilson R T, Till B D. Effects of outdoor advertising: Does location matter?[J]. *Psychology & Marketing*. 2011, 28(9): 909-933.
- [7] Jimenez, Alisha Y. Location Analysis of Alcohol Billboards in Bexar County, Texas[D]. Texas: Texas State University-San Marcos, 2013.
- [8] Cantrell J, Kreslake J M, Ganz O, et al. Marketing Little Cigars and Cigarillos: Advertising, Price, and Associations With Neighborhood Demographics[J]. *American Journal of Public Health*. 2013, 103(10): 1902-1909.
- [9] Rosewarne L. The men's gallery: Outdoor advertising and public space: Gender, fear, and feminism[J]. *Womens Studies International Forum*. 2005, 28(1): 67-78.
- [10] Huang M, Fang Z, Xiong S, et al. Interest-Driven Outdoor Advertising Display Location Selection Using Mobile Phone Data[J]. *IEEE Access*. 2019, 7: 30878-30889.
- [11] Dilchert S, Ones D S, Viswesvaran C, et al. Response distortion in personality measurement: born to deceive, yet capable of providing valid self-assessments?[J]. *Psychology Science*. 2006, 48(3): 209-225.
- [12] Anagnostopoulos A, Petroni F, Sorella M. Targeted interest-driven advertising in cities using Twitter[J]. *Data Mining and Knowledge Discovery*. 2018, 32(3): 737-763.
- [13] Beijing Municipal Bureau Statistics, Survey Office of the National Bureau of Statistics in Beijing. Beijing Statistical Yearbook 2018[M]. Beijing: China Statistics Press, 2018(北京市统计局, 国家统计局北京调查总队. 北京统计年鉴2018[M]. 北京: 中国统计出版社, 2018)
- [14] Beijing Municipal Bureau Statistics, Survey Office of the National Bureau of Statistics in Beijing. Beijing Area Statistical Yearbook 2018[M]. Beijing: China Statistics Press, 2018 (北京市统计局, 国家统计局北京调查

- 总队. 北京区域统计年鉴2018[M]. 北京: 中国统计出版社, 2018)
- [15] Chen Y, Liu X, Li X, et al. Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based k-medoids method[J]. *Landscape and Urban Planning*. 2017, 160(160): 48-60.
- [16] Jean N, Burke M, Xie M, et al. Combining satellite imagery and machine learning to predict poverty[J]. *Science*. 2016, 353(6301): 790-794.
- [17] Yao Y, Zhang J, Hong Y, et al. Mapping fine - scale urban housing prices by fusing remotely sensed imagery and social media data[J]. *Transactions in Gis*. 2018, 22(2): 561-581.
- [18] Wang Z, Niu L. A Data Model for Using OpenStreetMap to Integrate Indoor and Outdoor Route Planning[J]. *Sensors*. 2018, 18(7): 2100-2118.
- [19] Xu Zening, Gao Xiaolu. A novel method for identifying the boundary of urban built-up areas with POI data[J]. *Acta Geographica Sinica*. 2016, 71(6): 928-939(许泽宁, 高晓路. 基于电子地图兴趣点的城市建成区边界识别方法[J]. 地理学报, 2016, 71(6): 928-939)
- [20] Zhang Xun, Zhong Ershun, Zhang Xiaohu, et al. Spatial distribution and clustering of commercial network in Beijing during 2004-2008[J]. *Progress In Geography*. 2013, 32(8): 1207-1215(张珣, 钟耳顺, 张小虎, 等. 2004-2008年北京城区商业网点空间分布与集聚特征[J]. 地理科学进展, 2013, 32(8): 1207-1215)
- [21] Han Huiran, Yang Chengfeng, Song Jinping. Impact factors of location choice and spatial pattern evolution of wholesale enterprises in Beijing[J]. *Acta Geographica Sinica*. 2018, 73(2): 219-231(韩会然, 杨成凤, 宋金平. 北京批发企业空间格局演化与区位选择因素[J]. 地理学报, 2018, 73(2): 219-231)
- [22] Gu Yanyan, Jiao Limin, Dong Ting, et al. Spatial Distribution and Interaction Analysis of Urban Functional Areas based on Multi-source Data[J]. *Geomatics and Information Science of Wuhan University*. 2018, 43(7): 1113-1121(谷岩岩, 焦利民, 董婷, 等. 基于多源数据的城市功能区识别及相互作用分析[J]. 武汉大学学报·信息科学版, 2018, 43(7): 1113-1121)
- [23] Yao Yao, Ren Shuliang, Wang Junyi, et al. Mapping the fine-scale housing price distribution by integrating a convolutional neural network and random forest[J]. *Journal of Geo-information Science*. 2019, 21(2): 168-177(姚尧, 任书良, 王君毅, 等. 卷积神经网络和随机森林的城市房价微观尺度制图方法[J]. 地球信息科学学报, 2019, 21(2): 168-177)
- [24] Gallin J. The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets[J]. *Real Estate Economics*. 2006, 34(3): 417-438.
- [25] Yu Hanxiang, Wang Li, Pei Qian. A study on spatial differentiation of commercial housing price and its influencing factors in Dalian[J]. *Territory & Natural Resources Study*. 2018, 177(6): 20-25(于瀚祥, 王利, 裴倩. 大连市商品住宅价格空间分异及其影响因素分析[J]. 国土与自然资源研究, 2018, 177(6): 20-25)
- [26] Yao Y, Liu P, Hong Y, et al. Fine - scale intra - and inter - city commercial store site recommendations using knowledge transfer[J]. *Transactions in Gis*. 2019, 23(5): 1029-1047.
- [27] Guo B, Li J, Zheng V W, et al. CityTransfer: Transferring Inter- and Intra-City Knowledge for Chain Store Site Recommendation Based on Multi-Source Urban Data[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. 2018, 1(4):135(1)-135(23).
- [28] Gong Z, Ma Q, Kan C, et al. Classifying Street Spaces with Street View Images for a Spatial Indicator of Urban Functions[J]. *Sustainability*. 2019, 11(22): 6424-6440.
- [29] Guan C, Rowe P G. The concept of urban intensity and China's townization policy: Cases from Zhejiang Province[J]. *Cities*. 2016, 55: 22-41.
- [30] Zhou X, Chen Z, Yeh A G O, et al. Workplace segregation of rural migrants in urban China: A case study of Shenzhen using cellphone big data[J]. 2019. doi:10.1177/2399808319846903
- [31] Zheng Y, Liu F, Hsieh H P. U-Air: When Urban Air Quality Inference Meets Big Data[C]. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Chicago Illinois, USA, 2013.
- [32] Lies M, Glaser B, Huwe B. Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and Random Forest models[J]. *Geoderma*. 2012, 170: 70-79.
- [33] Carranza E J M, Laborte A G. Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random Forests algorithm[J]. *Ore Geology Reviews*. 2015, 71: 777-787.
- [34] Chen W, Xie X, Wang J, et al. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility[J]. *Catena*. 2017, 151: 147-160.
- [35] Breiman L. Random Forests[J]. *Machine Learning*. 2001, 45(1): 5-32.
- [36] Wu Chenwen, Liang Jinghan, W Wei, et al. Random Forest Algorithm Based on Recursive Feature Elimination[J]. *Statistics & Decision*. 2017(21): 60-63(吴辰文, 梁靖涵, 王伟, 等. 基于递归特征消除方法的随机森林算法[J]. 统计与决策, 2017(21): 60-63)
- [37] Mao Y, Xia Z, Yin Z, et al. Fault Diagnosis Based on Fuzzy Support Vector Machine with Parameter Tuning and Feature Selection[J]. *Chinese Journal of Chemical Engineering*. 2007, 15(2): 233-239.
- [38] Gromping U. Variable Importance Assessment in Regression: Linear Regression versus Random Forest[J]. *The American Statistician*. 2009, 63(4): 308-319.
- [39] Li Qiuping, Liu Yishi, Gong Shiyao, et al. Individual Income Level Inference Method Based on Travel Behavior of Urban Residents[J]. *Geomatics and Information Science of Wuhan University*. 2019, 44(10): 1575-1580(李秋萍, 刘逸诗, 巩诗瑶, 等. 基于居民出行活动特征的个体经济水平推断方法[J]. 武汉大学学报·信息科学版, 2019, 44(10): 1575-1580)
- [40] Tobler W R. A Computer Movie Simulating Urban Growth in the Detroit Region[J]. *Economic Geography*. 1970, 46: 234-240.

[41] Huang Lu. A comparative study on the economic radiation capacity of Beijing and Shanghai to the surrounding regions[J]. *Times Finance*. 2015, 607(33): 235-236(黄露. 北京、上海对周边区域经济辐射能力的比较研究[J]. 时代金融, 2015, 607(33): 235-236)

[42] Hu Wangshu, Sun Wei. Regional economic disparity in Beijing based on Theil index[J]. *Journal of Graduate University of Chinese Academy of Sciences*. 2013, 30(3): 353-360(胡望舒, 孙威. 基于泰尔指数的北京市区域经济差异[J]. 中国科学院大学学报, 2013, 30(3): 353-360)

## Research on Location Selection of Outdoor Advertisement of Home Decoration Brand Based on Multi-source Spatial Big Data

ZHANG Jiaqi<sup>1</sup> DU Kaihu<sup>1</sup> REN Shuliang<sup>1</sup> WANG Ruifan<sup>1</sup> GUAN Qingfeng<sup>1</sup> CHEN Wenhui<sup>2</sup> YAO Yao<sup>1,3</sup>

1 School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China;

2 School of Information Engineering, Zhejiang A&F University, Hangzhou 311300

3 Alibaba Group, Hangzhou 311121, China

**Abstract: Objectives:** Reasonable outdoor advertisement site selection plays a positive role in improving brand awareness and expanding market. Because commercial data is difficult to obtain, previous studies mostly focused on the macro theoretical level and failed to conduct detailed analysis on the actual layout of offline advertising site selection. We propose a framework to solve the above problems. **Methods:** This study chose Beijing as the study area, and built a prediction model of in-store conversion rate (*Standard*  $R^2=0.758$ ) based on random forest. We combined road network density, point of interest (POI) and other multi-source spatial data. These data can represent geographic and economic characteristics. We finally got the spatial distribution results of advertising suitability of this home decoration brand in Beijing. And then we analyzed affecting factors. **Results:** The results show that the in-store conversion rate of the home decoration brand is high in the center and low in the periphery in Beijing. It has the phenomena of strong spatial autocorrelation and high value aggregation. Meanwhile, the in-store conversion rate has a strong correlation with social economy, commercial politics and crowd activities. The location of continuous advertising exposure to the same group has a great influence on the in-store conversion rate. When select outdoor advertisement or commercial site, attention should also be paid to the places that can provide continuous exposure. **Conclusions:** The fine mapping results of the model constructed in this study can quantitatively evaluate the advertising effect of each location and maximize the efficiency. It also can provide a reference and theoretical basis for relevant studies on outdoor advertisement layout or commercial site selection.

**Key words:** multi-source data; outdoor advertisement; random forest; block feature set; in-store conversion rate

**First author:** ZHANG Jiaqi, postgraduate, specializes in urban computing. ningzetaolover@cug.edu.cn

**Corresponding author:** YAO Yao, PhD, associate professor. yaoy@cug.edu.cn

**Foundational support:** National Key Research and Development Program of China (No. 2019YFB2102903); National Natural Science Foundation of China (Grant No. 41801306, Grant No. 41671408, Grant No. 41901332); Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University (Grant No. 18S01); Natural Science Fund of Hubei Province(No. 2017CFA041); Natural Science Fund of Zhejiang Province(No. LY18D010001)

**网络首发:**

**标题:** 多源空间大数据场景下的家装品牌线下广告选址布局研究

**作者:** 张嘉琪, 杜开虎, 任书良, 王瑞凡, 关庆锋, 陈文辉, 姚尧

DOI:10.13203/j.whugis20190468

**引用格式:**

张嘉琪, 杜开虎, 任书良, 等. 多源空间大数据场景下的家装品牌线下广告选址布局研究[J]. 武汉大学学报·信息科学版, 2020,

DOI:10.13203/j.whugis20190468 ( ZHANG Jiaqi, DU Kaihu, REN Shuliang, et al. Research on Location Selection of Outdoor Advertisement of Home Decoration Brand Based on Multi-source Spatial Big Data[J]. Geomatics and Information Science of Wuhan University, 2020, DOI:10.13203/j.whugis20190468 )

**网络首发文章内容和格式与正式出版会有细微差别, 请以正式出版文件为准!**

**您感兴趣的其他相关论文:****海量监视数据云存储服务模型的设计与实现**

孙超, 肖文名, 曾乐, 白金婷

武汉大学学报·信息科学版, 2020, 45(7): 1099-1106

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20180404>

**高速公路多源数据融合下的层次拥堵区间探测方法**

方志祥, 黄守倩, 苏荣祥, 肖和平

武汉大学学报·信息科学版, 2020, 45(5): 682-690

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20190117>

**基于 Phoenix 的地理空间大数据管理系统**

陈勉, 李龙海, 谢鹏, 付少锋, 何列松, 周校东

武汉大学学报·信息科学版, 2020, 45(5): 719-727

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20180435>