

基于随机森林的流动摊贩分布模型*

戴若颖¹, 党雪薇¹, 冯兆¹, 李海霞², 柳林^{1,3}

(1. 中山大学地理科学与规划学院, 广东 广州 510275;

2. 中山大学岭南学院, 广东 广州 510275;

3. 美国辛辛那提大学地理系, 俄亥俄 辛辛那提 45221-0131)

摘要: 流动摊贩聚集是我国城市管理的难题。相关研究多针对流动摊贩形成与管理的机制, 对特定街道进行实证分析, 而缺少针对区域流动摊贩分布的定量建模。流动摊贩形成与许多因子有关, 各因子关系复杂, 所建-模算法随机森林建立适用于特征值关系复杂的分类器, 不需要依赖贡献度大的单个因子, 并且能够一定程度回避其他分类算法常有的噪声、异常值和过拟合造成的问题, 得到高精度的预测模型, 同时能够在此基础上得到各因子的贡献度, 为进一步研究流动摊贩问题提供依据。根据文献和实地调查分析, 选择房价、街道段长度、路口岔口数量、段内外公交线路数量等因子作为建模的特征值。以广州海珠区为例, 将流动摊贩数量分为 4 个等级, 等级 1 对应 0 个流动摊贩, 等级 2 对应 1~10 个流动摊贩, 等级 3 对应 11~20 个流动摊贩, 等级 4 对应 20 个以上流动摊贩。通过模型训练和参数调整, 使用卡帕系数和整体精度作为标准, 选择性能最优的预测模型。基于随机森林的流动摊贩分布模型预测流动摊贩的空间位置, 一定程度揭示流动摊贩形成和分布的规律, 可用于城市规划和管理, 同时辅助流动摊贩的相关研究。

关键词: 随机森林; 流动摊贩; 广州

中图分类号: D630 **文献标志码:** A **文章编号:** 0529-6579 (2018) 01-0076-07

Random-forest-based street vendors distribution model: A case study of Haizhu District

DAI Ruoying¹, DANG Xuewei¹, FENG Zhao¹, LI Haixia², LIU Lin^{1,3}

(1. School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China;

2. Lingnan (University) College, Sun Yat-sen University, Guangzhou 510275, China;

3. Department of Geography, University of Cincinnati, Cincinnati 45221-0131, USA)

Abstract: Street vendors problem is a challenge to city management in China. Vendors on the street provide convenience for citizens on one hand, and to some extent disturb the daily order of cities on the other hand. Most studies on street vendors focus on the mechanics of vendors' aggregation with certain management methodology, and focus on vendors of particular streets. A quantitative modeling on vendor distribution is rare. A number of factors with complex relationship to one another decide the aggregation of street vendors, therefore, the algorithm of random forest is suitable for its application to building classifiers with complex features. With this algorithm not relying on particular critical factor, problems occurred frequently in other algorithms, such as outliers, noises and overfitting, can be well avoided. Further analysis can

* 收稿日期: 2016-08-27

基金项目: 国家自然科学基金重点项目 (41531178); 广东省自然科学基金研究团队项目 (2014A030312010); 国家自然科学基金 (41171140); 广东省科技计划项目 (2015A020217003)

作者简介: 戴若颖 (1994 年生) 女; **研究方向:** 地理信息和地貌学; E-mail: r.dai@students.uu.nl

通信作者: 柳林 (1965 年生), 男; **研究方向:** 地理信息和公共安全; E-mail: lin.liu@uc.edu

be made since the algorithm provides magnitude for each street vendors concerned factors. According to literature and field investigation, average house price per square, length of street segments, folks number, bus line number are chosen for modeling. Haizhu district of Guangzhou has been investigated. Taking kappa coefficient and overall accuracy as standard, the optimum predict model of street vendors is generated with training and parameters adjustment, setting no vendor as grade 1, 1 ~ 10 vendors as grade 2, 11 ~ 20 vendors as grade 3, 21 or more vendors as grade 4. The proposed model of street vendors distribution which predicts the appearance of street vendors, not only facilitates city planning and management, but also throws new light on relevant research.

Key words: random forest; street vendor; Guangzhou

随着我国城市发展, 流动摊贩的数量不断增加, 对城市交通、景观等方面造成负面影响^[1]。如何管理流动摊贩是我国城市社会空间矛盾的重要问题^[2]。流动摊贩是非正规经济的重要组成部分^[3], 作为一种非正规从业人员, 流动摊贩通过在街头和其他公共场所摆摊设点来进行零售和服务活动, 往往没有营业执照和固定的经营场所^[4]。国外城市流动摊贩问题研究开始于 20 世纪 60 年代, 主要讨论流动摊贩的经济价值及特征^[5]。在“市民社会”理念的影响下, 一些学者开始研究流动摊贩对社会的影响, 如流动摊贩占道经营引起的城市公共空间使用权的问题等^[6-7]。国内的相关研究起步较晚, 主要是针对流动摊贩的制度化管理的研究, 如有关城管与摊贩之间的“猫鼠大战”^[8-9]和“摊贩自治”^[10]的问题等, 而有关于流动摊贩空间分布的研究相对较少。既有研究仅仅分析了影响流动摊贩分布的因素, 少有研究讨论如何预测流动摊贩分布。鉴于此, 本研究采用随机森林模型预测广州市海珠区流动摊贩的分布。

随机森林算法 (Random Forests) 又称随机决策森林 (Random Decision Forests)^[11], 作为强大的分类算法应用于各个领域。Leo Breiman^[12] 和 Adele Cutler^[13] 首先使用了随机森林算法, 该算法综合了 Breiman 自助聚合 (Bootstrap Aggregating) 的理念, Ho^[11] 的随机特征选取的理念, 以及 Amit 和 Ge-man^[14] 提出的训练单棵决策树的方法, 即在分裂节点搜索可行决策的随机子集。随机森林通过训练建立一系列的决策树, 选择决策树分类结果的众数得到最终分类结果, 克服了决策树算法过拟合的问题^[14]。

本研究以广州市海珠区为例, 基于随机森林模型, 定量预测流动摊贩的聚集与分布。研究成果有助于城市管理和规划以及进一步的相关研究, 促进城市和谐发展。

1 随机森林

1.1 自助法

基于计算机技术的自助法 (Bootstrap) 由 Efron 在 1979 提出^[15], 当时用于估计标准差。自助法不是独立的算法, 它应用于多种统计学方法, 比如非参数回归, 密度估计, 决策树分类和最小中值平方回归。

自助法对样本集进行独立同分布的 m 次取样, 由此得到的样本称为自助样本 (Bootstrap Sample), 利用 m 份样本各自生成模型, 将 m 个模型得到的结果进行平均或投票, 分别得到对数据实现回归和分类。

1.2 多折交叉验证

交叉验证 (Cross Validation) 在保证在没有损失大量信息量的情况下建立模型, 并给出用于模型应对新样本时的精度, 可防止过拟合问题。其中, 交叉验证可选择多折交叉验证和留一验证, 决策树分类时使用多折交叉验证 (K-fold Cross Validation)。作为另一个验证方法, 留一交叉验证得到的决策树之间分类精度接近, 对选择最优决策树没有帮助。

多折交叉验证步骤如下: ① 将数据分成等数量的 k 份; ② 使用 $k-1$ 份数据进行建模, 验证时使用第 k 份数据, 对比真实数据后计算模型误差; ③ 完成以上 k 次计算后, 加和所有模型误差, 得到整体模型误差。

1.3 分类与回归决策树

分类与回归决策树 (Classification and Regression Tree) 由 Breiman 等在 1984 提出^[16], 它提供了基于计算机技术的决策树建模方法, 可用于回归问题和分类问题。

在决策树的每个节点 (Node), 根据该节点的逻辑判断问题将样本分到多个分支, 叶节点是停止继续下分的节点。其中每个节点的逻辑判断问题运

用样本集的一个或多个特征值。对于分类决策树,被分到某个叶节点最多的样本类别即为该结点所代表的类别。

获得最佳分类与回归决策树的步骤如下:

1) 使用计算机技术得到每个节点的最佳逻辑判断设定,比如设定特征值 1 小于 5 的样本在左边的分支,大于等于 5 的样本在右边的分支,以便最大程度区分样本;

2) 在计算精度时,对错分的类别设定惩罚分,比如将类别 1 分错的惩罚分设为 1,将类别 2 分错的惩罚分设为 2,以惩罚分之和最小的模型作为最优模型;

3) 通过多折交叉验证得到精度最高的决策树,防止过拟合问题。过拟合问题表现为过大的决策树,叶节点下的样本太少,这导致模型只对用于建模的数据有效。

1.4 随机选取特征子集构建决策树集

随机选取特征子集构建决策树集由 Ho 在 1995 提出^[11]。对于 m 维特征集,有 2^m 个特征子集可用于构建决策树。几何上,每个决策树的叶节点代表了一定特征子集上的范围。通过平均后验概率,判别式确定了样本点所在的邻域范围。

Ho 使用了如下判别式以综合 t 个决策树的结果。其中, X 为特征值向量, $V_j(X)$ 为第 j 棵决策树的叶节点, C_i 为类别记号 ($i = 1, 2, 3, \dots, n$)。

$$P(C_i | V_j(X)) = \frac{P(C_i, V_j(X))}{\sum_{q=1}^n P(C_q, V_j(X))}$$

以上概率可用 C_i 类的样本点数量与所有到达叶节点 $V_j(X)$ 的样本点数量之比估计。在此,多数叶节点只有一个类的样本点,因此 $P(C_i, V_j(X))$ 常常为 1。

判别式子如下所示, $g_{C_i}(X)$ 最大值对应的类别即为样本点的分类结果。

$$g_{C_i}(X) = \frac{1}{t} \sum_{j=1}^t P(C_i, V_j(X))$$

1.5 随机森林的步骤

1) 抽样。使用自助法从原始样本中抽出训练样本集,每一棵决策树均选择独立同分布的训练样本集。自助法的优越性在于,其一,提高随机森林的精度;其二,使用袋外数据可以估计决策树的泛化误差 (Generalization Error) 等;

2) 选择特征子空间。对于决策树的每一个节点,随机挑选特定个数的特征值作为特征子空间并计算每个特征的最佳分支数值作为该节点分支的准则,完成决策树的构建;

3) 组合所有决策树。输入测试样本进行分类,所有样本分类结果的众数即为随机森林的最终分类结果。

1.6 泛化误差与平均精度减少量的 OOB 估计

边际函数 (Margin Function) 计算决策树集合对一类样本点正确分类的平均次数与最大的错误分类平均次数的差值。如下式所示,其中, av_k 指对第 k 个分类器求平均, $I(\cdot)$ 为指示函数, $h_k(\cdot)$ 为第 k 个分类器, X 为特征值向量, Y 为分类结果,即分类器的输出。

$$\text{mr}(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j)$$

泛化误差衡量随机森林模型错误分类测试样本的概率。如下式所示,其中, θ 为特征值向量和分类结果向量在对应条件下的概率。

$$PE^* = P_{X, Y}(\text{mr}(X, Y) < 0)$$

根据 Breiman 的研究, θ 为分布概率随机向量,泛化误差随着决策树数目增长收敛于:

$$P_{X, Y}(P_\theta(h(X, \theta) = Y) - \max_{j \neq Y} P_\theta(h(X, \theta) = j) < 0)$$

将未建模的数据运用到已有的分类器,对泛化误差进行袋外估计 (Out-of-Bag Estimate) 以此衡量决策树分类器的优劣。

平均精度减少量是指在其他变量不变的情况下,将一个变量取为随机数得到的随机森林精度下降程度。一个变量的平均精度减少量越大代表该变量对分类越重要。平均精度减少量可通过袋外估计得到。

2 数据与模型

2.1 研究区概况

据统计,目前广州市流动摊贩从业人员约有 23 万 ~ 25 万人,最多时超过 30 万^[17],是整个城市的重要组成部分,也是造成城市空间矛盾的主要群体之一。以往,对于流动摊贩的管理,政府一直采取驱逐和排斥的治理模式,管理手段单一且管理效果不佳。2010 年起,广州提出了“疏堵结合”政策,2011 年先后颁布、实施了《广州市流动商贩疏导区管理试行办法》和《关于划定乱摆卖严禁区的通知》等管理办法,试图通过建设疏导区的方式来管理流动摊贩。截至 2015 年 3 月,全市已有 103 个疏导区建成并投入使用,一定程度上将流动摊贩纳入了有序的城市管理系统中,但仍有诸多流动摊贩分布在城市的缝隙空间中,影响着城市的管理与规划。由于其城市化起步早,吸引了大量外来人口,同时也是流动摊贩广泛分布的区域之

一，具有典型性。本研究选取海珠区 14 个街区共 324 个道路段作为研究对象。

2.2 因子筛选与数据获取

流动摊贩的区位选择具有与一般零售业类似的消费者指向性^[18]，但与正规的零售业的区位选择不完全相同。庞宇琦等^[1]研究了城市典型地段中流动摊贩的分布特征和空间上的影响因素，认为流动摊贩的分布受时间、人流量、营业空间等因素的影响，并且在每天的 16 - 19 时流动摊贩的数量最多。张磊^[19]在研究北京市朝阳区流动摊贩的类型及经营规律时也提到流动摊贩主要分布在人流密集的公交车站、居民小区门口、超市及市场周边，集中在市民上下班时间贩卖商品。张延吉等^[20]探究了以北京市甘露园社区为例的流动摊贩的空间分布规律及其影响因素，认为人流量大小、街道物质环境和政府管制强度均对流动摊贩的区位选择产生影响，但政府管制强度仅对流动摊贩的产生起到暂时的抑制作用，其效用仅仅是使流动摊贩的位置产生了变化，并未从根本上影响流动摊贩的产生。鉴于此，本研究不考虑政府管制强度。

在既有研究中，人流量始终是吸引流动摊贩聚集的核心因素^[21]，庞宇琦、黄耿志等人的研究得出流动摊贩在人流交汇的结点位置附近呈现出段状分布^[1-2]，且公交线路因有利于吸引人流而受到流动摊贩的青睐^[2]，因此本研究采用路口岔口数、街道段段内段外公交线路数和道路等级这 3 个变量测度人流动因素对摊贩分布产生的影响：路口岔口数越多、公交线路数越多，可以带来越多的潜在消费者；道路等级的数值越大，相应道路越窄，行人通行慢，在摊贩前停留时间越多，潜在交易越多，吸引流动摊贩。

除此之外，道路环境也对流动摊贩的分布有重要影响，本研究采用街道段长度、街道段旁的建筑物市场价格（房价）和社会失序等级这 3 个变量测度道路环境对摊贩分布产生的影响。根据已有的

研究以及实地调研发现，每个流动摊贩所占的面积大约为 1 ~ 4 m² 的面积^[1]，街道段越长，流动摊贩的数量就越多；街道段附近房价则代表消费者的消费能力和需求，流动摊贩会选择具有对应消费能力的群体作为目标对象；街道段的失序等级越高，环境越复杂，则越易形成流动摊贩，具体的社会失序等级说明如表 1 所示。

综合以上影响因素，本研究最终筛选出 7 种流动摊贩分布的影响因子：道路功能等级，路口岔口数，道路段长度（m），房价（元/m²），混乱程度，段内线路数，段外线路数。考虑到由于公交站的辐射范围不同所以对流动摊贩的分布的影响也有所不同，分别设置各个道路段的半径为 50，100 和 150 m 的公交缓冲区，将各个道路段的“段外线路数”进一步细分为道路段 50 m 范围内公交线路数、道路段 100 m 范围内公交线路数、道路段 150 m 范围内公交线路数。具体影响因子的说明及数据来源如表 2。

2.3 模型训练与参数调整

本模型使用 MatLab2014a 作为运算平台，结合 Abhishek Jaiantilal^[22] 基于 Andy Liaw^[23] 研究得到的运算包进行建模。

随机森林模型中的决策树的数量和建模特征值的数量均影响模型精度，且对于不同建模对象和数据情况有着不同程度的影响。其中，决策树的数量即用自助法抽样的次数，建模特征值的数量即为决策树分裂结点使用的因子数。

为得到最优模型，通过对比不同参数组合得到的袋外估计泛化误差和预测整体精度，选择最优参数组合。泛化误差越小，预测整体精度越大，则模型越好。由于本模型使用五折交叉验证，泛化误差为五折运算的平均结果，由袋外数据得到。比起非交叉运算，交叉运算能同时提供袋外数据的泛化误差和验证样本的精度，综合两者结果，为选择模型参数建立基础。

表 1 社会失序等级说明

Table 1 Social disorder level definition

社会失序等级	说 明
1	最不混乱，治安良好，交通秩序井然，如高档小区，高档商业区附近
2	不混乱，治安较好，交通秩序一般，如高档小区的停车场
3	较不混乱，人流量大，商铺众多，交通一般，类型较统一
4	较混乱，人流量大，商铺众多，交通较乱
5	混乱，人流量大，车流量小，交通复杂，商铺类型比较复杂
6	特别混乱，人流量大，车流量小，交通秩序差，商铺类型复杂，且为低档商铺，如菜市场附近

表 2 影响因子说明及数据来源
Table 2 Sources and definition of factors

自变量	说明	来源
道路功能等级	根据道路中心线编码表确定	道道通公司
路口岔口数	每条道路段两端的道路岔口数之和	2012 年遥感影像
街道段长度	每条道路段的长度	2012 年遥感影像
房价	路段段毗邻的区域平均房价	房天下网 gz.fang.com
社会失序等级	对研究区的格局、商铺类型、治安情况、 交通情况的综合评级	实地调研 (调研时间为 18:00 - 20:00)
街道段段内公交线路数	每条道路段内公交站经过的公交线路数之和	百度地图
街道段 50 m 范围内公交线路数	每条道路段 50 m 范围内的公交站的公交线路数之和	百度地图
街道段 100 m 范围内公交线路数	每条道路段 100 m 范围内的公交站的公交线路数之和	百度地图
街道段 150 m 范围内公交线路数	每条道路段 150 m 范围内的公交站的公交线路数之和	百度地图

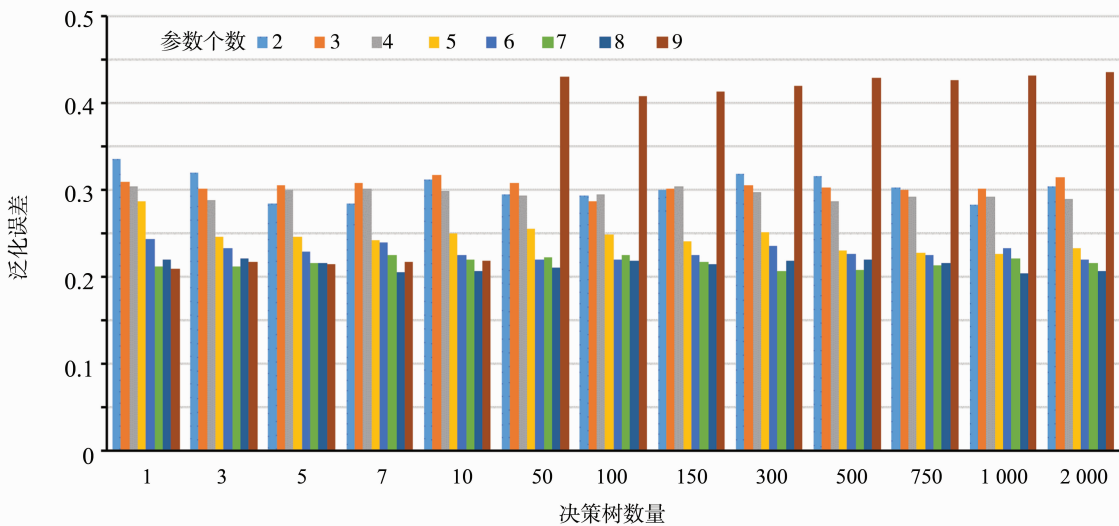


图 1 泛化误差

Fig. 1 Generalization error

经实验,泛化精度结果如图 1 所示,分析可得,在流动摊贩分布预测问题上:①在建模特征值数量一定时,整体上决策树数量对泛化误差的影响较小。在建模特征值数量为 9 时,决策树越多,泛化误差越多。这是因为过多的建模特征值使得决策树之间的相关性太大。②在决策树数量一定时,整体上特征值数量越大,泛化误差越小,其中特征值数量为 6 和 7 时泛化误差最小。

综合对泛化能力、整体精度和计算效率的考虑,我们选择决策树为 100 棵,特征值数量为 8 作为建模的参数组合。

2.4 结果

由上述模型产生的各因子的重要性度量如图 3 所示。从图中可看出街道段附近房价、街道段长度和社会失序等级这三个变量对预测流动摊贩分布的重要性较高,这与张延吉等人提出的道路的物质环

境对流动摊贩的发生具有显著影响的结论相同^[20],其中街道段附近房价最为重要,这是因为街道段附近房价反映了该街道段大部分消费者的消费能力和需求,由于房价低的小区聚集消费水平较低的人群,他们是流动摊贩的主要客源,客源多吸引流动摊贩;街道段长度的重要性略低于街道段附近房价,足够长的街道段为流动摊贩提供了充分的摆摊空间,吸引更多的流动摊贩;社会失序等级可以真实地模拟摊贩和消费者对道路的主观评价,可以为预测流动摊贩的分布提供有效支持。其次是道路岔口数、街道段 150 m 范围内的公交线路数和道路功能等级,道路岔口数越多在一定程度上可以表示人流量越多,从而吸引流动摊贩;道路功能等级的数值越大,车速越小,行人越多,对流动摊贩的产生有一定的影响。街道段段内和段外的公交线路数对预测流动摊贩分布的重要性不同,不同范围的段

外线路数的重要性也不同,其中,段内公交线路数的重要性最小,街道段外的范围越大,重要性越高。

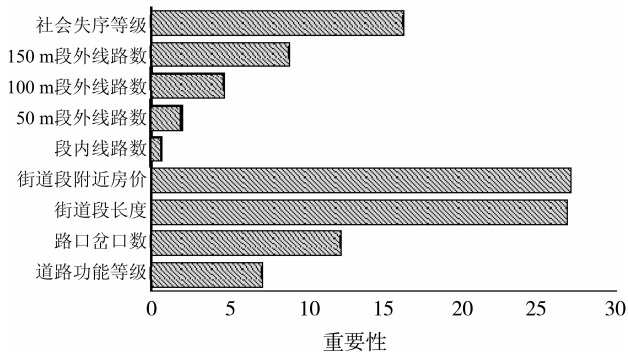


图 2 影响因子重要性度量

Fig. 2 Significance of factors

本研究使用 70% 的研究样本用于建立模型,30% 的研究样本用于验证模型。对比 30% 的真实数据与模型预测值,正确预测的样本占 86.15% (即整体精度),卡帕系数为 0.596 (卡帕系数: (正确预测率 - 偶然预测率) / (1 - 偶然预测率))。

3 讨 论

根据实际调研,部分预测错误的道路段附近有农贸市场、学校、广场、地铁口等人流量较高的公共场所,吸引流动摊贩聚集,如客村地铁站。客村地铁站是广州 3 号地铁线和 8 号地铁站的中转站,靠近丽影广场,人流量大且有足够的空间供流动摊贩摆摊售卖,由于本次研究未考虑广场和地铁出口等影响因子,因此在客村地铁站附近的预测值低于真实值。也有部分道路段由于道路宽度太窄,不利于摆摊,位于新港西路南部的康乐村和轻纺城,由于这两个区域内道路狭窄、岔路口多,且有大量三轮车快速穿行,环境复杂,尽管人流量很大,但少有充足的空间供流动摊贩摆摊,因此在这两个区域的预测结果高于真实值。鉴于此,在今后的研究中可考虑将农贸市场、广场、大中小学、地铁站口等公共场所量化后纳入影响流动摊贩分布的因子中从而完善模型,并根据实际道路情况优化模型,使其预测精度得以提高。

流动摊贩形成的原因复杂,单个因子对每个摊贩的影响程度不同,因此单个因子不足以区分流动摊贩数量等级。例如公交线路数量较少的区域不一定意味着人流量少,有的研究区域不存在公交线

路,却因为有三轮车等公共交通而人流量大,流动摊贩多。虽然数据中含有一定噪音和异常值影响模型精度,但随机森林对数据中的离散点不敏感,可以一定程度避免这些数据对最终分类结果的干扰。此外随机森林设置了袋外数据,建模所使用的训练样本不会过拟合,充分运用了研究区数据,使得流动摊贩模型最大程度地泛化,对于建模数据以外的地区也有较高精度的预测功能。最后根据随机森林袋外数据进行估计,可得到各个因子重要性的定量分析,为揭示流动摊贩形成的机理提供科学依据。

流动摊贩定量建模仍有改善空间。城市流动摊贩的分布不仅受制于诸如人流量,房价,公交线路数的可观因素,还受经济、政府管制等难以量化的因素影响。本研究主要考虑了影响流动摊贩分布的确定性因子,而政府政策、管制以及经济环境的变化往往在很大程度上影响流动摊贩的分布甚至决定其是否存在,所得模型未考虑这些因素,只是对现实世界进行了初步的模拟。因此,为提高模型精度和进一步揭示海珠区流动摊贩分布规律,接下来的定量研究还可纳入更多的影响因子并结合所研究区的政治政策及经济环境。

4 结 论

本研究在前人的定性研究基础上,选取 7 个影响因子,用随机森林模型模揭示广州市海珠区流动摊贩空间分布规律。主要结论如下:

- 1) 提出社会失序是影响流动摊位的重要因素,并得以验证;
- 2) 随机森林模型的预测结果稳健,预测的流动摊位分布与实际情况基本吻合。其度量的各个因子进行重要性度量便于解释各因子对流动摊贩分布的影响;
- 3) 街道段附近房价、街道段长度和社会失序等级这 3 个变量对预测流动摊贩分布的影响程度大于路口岔口数量和段内外公交线路数量。

本研究的模型和结果不仅为城市管理和城市规划提供科学依据,同时也为流动摊贩和非正规经济相关研究提供新视角。

参考文献:

- [1] 庞宇琦,薛德升,王德. 流动商贩的分布特征及营业空间界面影响分析——以广州中山大道西路段为例[J]. 人文地理, 2012, 27(2): 56-61.
PANG Y Q, XUE D S, WANG D. The distribution and spatial impact factors of street hawker[J]. Human Geog-

- raphy, 2012, 27(2): 56–61.
- [2] 黄耿志,徐孔丹,薛德升. 广州摊贩空间疏导模式及其成功效应的影响因素——基于沥滘村疏导区的实证研究[J]. 人文地理, 2013,28(6):74–79.
HUANG G Z, XU K D, XUE D S. The spatial inclusion of street-vending and influence factors of its effect in Guangzhou: A case study of Lijiao[J]. Human Geography, 2013, 28(6): 74–79.
- [3] ILO. Women and men in the informal economy: A statistical picture [R]. International Labour Organization, 2002.
- [4] SHARIT K B. Street vendors in Asia: A review[J]. Economic and Political Weekly, 2005, 40(22): 2256–2264.
- [5] 何丹,朱小平,钱志佳. 城市流动摊贩研究述评——兼论上海市摊贩的特征[J]. 城市问题, 2013(3):73–79.
HE D, ZHU X P, QIAN Z J. Review on the researches on city street vendors: characteristics Shanghai's city street vendors[J]. Urban Problems, 2013(3):73–79.
- [6] STAUDT K. Struggles in urban space: street vendors in El Paso and Ciudad Juarez[J]. Urban Affairs Review, 1996, 31:435–454;
- [7] DONOVAN M G. Informal cities and the contestation of public space: The case of Bogotá's street vendors, 1988–2003[J]. Urban Studies, 2008, 45(1): 29–51
- [8] 吴佳丽. 当权力遭遇弱者:城市流动摊贩治理中的行为逻辑[D]. 南京:南京理工大学, 2009.
WU J L. When the power encountered the weak: the logic in the vendors administration-A case study of the team of city H's urban management[D]. Nanjing: Nanjing University of Science and Technology, 2009.
- [9] 黄文芳. 试论摊贩管理中的治理与包容[J]. 环境卫生工程, 2008(5):55–58.
HUANG W F. Administer and comprehension in Hawkers' management[J]. Environmental Sanitation Engineering, 2008(5):55–58.
- [10] 杨介聪. 摊贩组织化及其自我管理问题研究[D]. 上海:复旦大学, 2009.
YANG J C. On street vendors organization and autonomy issue[D]. Shanghai: Fudan University, 2009
- [11] HO T K. Random decision forests [C]. International conference on document analysis and recognition, 1995.
- [12] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45: 5.
- [13] LIAW A. Documentation for R package random forest [EB/OL]. <https://cran.r-project.org/web/packages/randomForest/index.html>,2013.
- [14] AMIT Y, GEMAN D. Shape quantization and recognition with randomized trees [J]. Neural Computation, 1997, 9(7): 1545–1588
- [15] EFRON B. Bootstrap Methods: Another look at the Jackknife[J]. Annals of Statistics, 1979, 7(1): 1–26.
- [16] BREIMAN L. Classification and regression trees[M]. Wadsworth & Brooks/Cole Advanced Books & Software, 1984.
- [17] 杨芳,张倩,林静,等. 创新管理模式 治理城市顽疾——广州市流动商贩疏导区建设成效和经验[J]. 城市管理与科技, 2015,17(6):42–44.
YANG F, ZHANG Q, LIN J, et al. Solving urban issues—The achievement and experience of street vendors placement in Guangzhou [J]. Urban Management and Science & Technology, 2015,17(6):42–44.
- [18] 曹嵘,白光润. 交通影响下的城市零售商业微区位探析[J]. 经济地理, 2003(2):247–250.
CAO R, BAI G R. The study on the retail micro-location in city under the influence of transportation[J]. Economic Geography, 2003(2):247–250.
- [19] 张磊. 流动摊贩类型及其经营规律研究——以北京市朝阳区为例[J]. 城市问题, 2013(5):72–76.
ZHANG L. On street vendors classification and their management pattern: A case study of Chaoyang district of Beijing city [J]. Urban Problems, 2013(5):72–76.
- [20] 张延吉,张磊,吴凌燕. 流动摊贩的空间分布规律及其影响因素——以北京市甘露园社区为例[J]. 城市问题, 2014(8):81–85.
ZHANG Y J, ZHANG L, WU L Y. On street vendors spatial distribution and its factors[J]. Urban Problems, 2014(8):81–85.
- [21] HAYSMITCHELL M. Streetvending in Peruvian cities: The spatio-temporal behavior of ambulantes [J]. The Professional Geographer, 1994, 46(4): 425–438.
- [22] JAIANTILAL A. Randomforest-matlab [EB/OL]. <https://github.com/ajaiantilal/randomforest-matlab>,2016.
- [23] LIAW A. Random forest: Breiman and Cutler's random forests for classification and regression [EB/OL]. <http://cran.r-project.org/web/packages/randomForest/index.html>,2015.