

引用格式: 柳林, 纪佳楷, 宋广文, 等. 基于犯罪空间分异和建成环境的公共场所侵财犯罪热点预测[J]. 地球信息科学学报, 2019, 21(11): 1655-1668. [Liu L, Ji J K, Song G W, et al. Hotspot prediction of public property crime based on spatial differentiation of crime and built environment[J]. Journal of Geo-information Science, 2019, 21(11): 1655-1668. ] DOI: 10.12082/dqxxkx.2019.190358

# 基于犯罪空间分异和建成环境的公共场所 侵财犯罪热点预测

柳林<sup>1,2,3,4,\*</sup>, 纪佳楷<sup>1,2</sup>, 宋广文<sup>3</sup>, 廖薇薇<sup>1,2</sup>, 余洪杰<sup>1,2</sup>, 刘文娟<sup>1,2</sup>

1. 中山大学地理科学与规划学院, 广州 510275; 2. 广东省公共安全与灾害工程技术研究中心, 广州 510275;  
3. 广州大学地理科学学院公共安全地理信息分析中心, 广州 510006; 4. 辛辛那提大学地理系, 辛辛那提 OH 45221-0131

## Hotspot Prediction of Public Property Crime based on Spatial Differentiation of Crime and Built Environment

LIU Lin<sup>1,2,3,4,\*</sup>, JI Jiakai<sup>1,2</sup>, SONG Guangwen<sup>3</sup>, LIAO Weiwei<sup>1,2</sup>, YU Hongjie<sup>1,2</sup>, LIU Wenjuan<sup>1,2</sup>

1. School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China;  
2. Guangdong Provincial Engineering Research Center for Public Security and Disaster, Guangzhou 510275, China;  
3. Center of Geographic Information Analysis for Public Security, School of Geographic Sciences, Guangzhou 510006, China;  
4. Department of Geography, University of Cincinnati, Cincinnati OH 45221-0131, Ohio, USA

**Abstract:** Machine learning is the mainstream method for crime hotspot prediction. As a popular machine learning algorithm, the random forest algorithm is widely used in the construction of crime hotspot prediction models because of its ability of handling sparse data, and reliable predictive capability and accuracy. A number of studies use multi-source data representing the geographical environment and built environment to train and construct crime hotspot prediction models. Some are theory-driven, while others more data-driven. Most crime prediction models are global models, by constructing a single model for the entire study area. These models do not fully consider the spatial variations of crime and the built environment, as well as the varying relationship between crimes and the built environment. This paper aims to fill in this gap, using public property crime as an example to demonstrate that crime prediction models can be improved by incorporating the aforementioned spatial variations and spatially varying relationship. Firstly, according to the distribution of historical crime events and the distribution of past crime hotspots, the research area was divided into four subareas: stable high-heat grids, high-heat grids, even-hot grids, and non-hot grids. Then, according to the social disorganization theory, routine activity theory, and crime pattern theory, the three covariates including the urban village, the road network, and POI (catering, entertainment and shopping malls as crime attractors and generators) were used as

收稿日期: 2019-07-05; 修回日期: 2019-10-08.

**基金项目:** 国家重点研发计划项目(2018YFB0505500、2018YFB0505503); 国家自然科学基金重点项目(41531178); 广州市科学研究计划重点项目(201804020016); 广东省自然科学基金研究团队项目(2014A030312010)。[ **Foundation items:** National Key R&D Program of China, No.2018YFB0505500, 2018YFB0505503; Key Program of National Natural Science Foundation of China, No.41531178; Key Project of Science and Technology Program of Guangzhou City, China, No.201804020016; Research Team Program of Natural Science Foundation of Guangdong Province, China, No.2014A030312010. ]

**作者简介:** 柳林(1965-), 男, 湖南湘潭人, 博士, 教授, 博导, 主要从事犯罪空间模拟、多智能体模拟、GIS应用等研究。  
E-mail: lin.liu@uc.edu

the covariates representing the surrounding built environment. The random forest prediction model also used historical crime data for training and validation. Different models were created for the whole study area and each of the four subareas. The results of 26 bi-week crime hotspot prediction experiments in 2017 were compared, showing that, after adding the three covariates representing the built environment, the prediction accuracy of the entire study area, stable high-heat grids, and high-heat grids were all improved. More importantly, the subarea models were substantially more accurate than the whole model. These findings strongly endorse that incorporating spatial differentiation of crime and the built environment plays a critical role in improving the performance of the prediction models. The majority of the hotspots coincide with commercial facilities that serve as crime generators or attractors. Thus, crime prevention and control should target urban villages and the areas where road densities are high. Further, the differences in subarea based models also suggest any crime fighting strategies should be adjusted to fit each local subarea, to achieve the greatest efficiency.

**Key words:** crime hotspot prediction; public property crime; random forest; built environment; spatial differentiation; stable high-heat grids; police strategy

**\*Corresponding author:** LIU Lin, E-mail: lin.liu@uc.edu

**摘要:**机器学习是当前犯罪热点预测的主流方法,随机森林算法因需要的数据量较小、有较好的预测能力和预测精确度、且有较高的可理解度,更是被广泛应用,代表地理环境和建成环境的多源数据也被广泛用于模型改进的尝试实践中,但这些实践都只考虑研究区整体的预测精度变化情况,并未区分不同区域犯罪热点预测结果的差异及其原因。因此,本文以公共场所所侵财犯罪为例,根据历史犯罪分布情况及过往犯罪热点分布规律,将研究区分为稳定高发热点网格、较高高发热点网格、偶发热点网格及非热点网格这4类,并依据社会失序理论、日常活动理论和犯罪模式理论,选取城中村范围、路网密度及POI(餐饮、娱乐、商场3类设施)密度这3个具有代表性的协变量加入到随机森林预测模型中,探讨预测结果精度的变化情况。根据2017年26个双周的犯罪热点预测实验的预测结果,得到以下结论:加入协变量后,研究区整体、稳定高发热点网格及较高高发热点网格的预测精度都有不同程度的提高,分区模型的精度显著高于整体模型的精度,说明考虑空间分异对提高模型精度起重要作用。

**关键词:**犯罪热点预测;公共场所所侵财犯罪;随机森林;建成环境;空间分异;稳定高发热点网格;警务策略

## 1 引言

犯罪的发生一直是影响城市公共安全及社会治安环境的主要问题,犯罪防控是警务工作的重点。传统警务策略时效性差,不能随犯罪活动的发生实时做出响应与调整,对犯罪活动的震慑及控制力度有限,公安部门在工作实践中逐渐认识到数据分析对现代警务策略制定的重要指导作用,并注重多源数据的收集管理,同时在当前大数据和机器学习发展普及的背景下,犯罪预测研究迎来新的发展,并逐渐推向实际应用。

根据历史犯罪的分布分析,犯罪的发生在空间上呈现一定的集聚和离散特征<sup>[1]</sup>,在空间上的分布存在热点与非热点区域,识别与预测空间上的犯罪的热点区域是犯罪预测研究的重要内容。传统的犯罪热点预测通常采用核密度估计法,根据历史犯罪数据进行犯罪风险制图,寻找犯罪热点分布规律,并以此判断下一个时间周期的犯罪热点区

域<sup>[2]</sup>。这种方法得到的结果是热点的一般代表,适用于长期稳定热点区域的识别,但不能识别热点区域的准确位置和具体边界,不适合空间尺度较小的犯罪热点预测研究。随着大数据和机器学习方法的探索和发展,包括随机森林、神经网络、贝叶斯模型等多种方法被运用于犯罪热点预测的实证研究中<sup>[3-4]</sup>。作为机器学习的一种主流算法,随机森林算法对比深度学习算法具有训练数据量较小、可理解度较高等优点,且因有较好的非线性关系数据处理能力及挖掘数据信息效率而广泛应用于各个领域<sup>[5-6]</sup>,在一些最新的犯罪热点预测实证研究中也同样得到了较好的预测结果<sup>[7]</sup>,也优于神经网络等其他机器学习算法得到的预测结果<sup>[8-9]</sup>,因此也被用于包括不同时空尺度的预测、使用多源数据对预测结果优化情况的验证以及预测结果评价指标等多方面的犯罪热点预测的探索研究中<sup>[10-11]</sup>。

日常活动理论提出,犯罪的发生需要3个条件:潜在的犯罪者、合适的目标及犯罪防范的缺失<sup>[12]</sup>。

社会失序理论认为,在一个社区中,社区组织结构、人口流动性和人群异质性会一定程度上增加失序的可能性,导致起内部控制作用的机构难以建立、起非正式控制作用的社区内部关系被弱化,从而对社区的犯罪率产生影响<sup>[13-14]</sup>。基于日常活动理论和理性选择理论而产生的犯罪模式理论有3个关键点,分别为空间认知、行为路径和活动范围,日常生活的行为路径和活动范围决定犯罪主体对平时生活范围内不同场所形成的空间认知,而空间认知很大程度上影响犯罪主体实施犯罪行为<sup>[15]</sup>。不同场所是由不同的空间要素构成的,根据犯罪学理论,各种客观环境因素及社会环境因素共同影响该场所的人员构成、监管力度等,使各个空间环境中有不同的犯罪机会,影响犯罪主体的空间认知及决策,从而很大程度上影响整体空间格局上的犯罪发生概率的分布情况<sup>[16]</sup>。各类微观空间要素包括交通枢纽等重要空间节点以及不同地块上的空间设施对各类犯罪的影响也得到了研究证明<sup>[17-19]</sup>。因此对于犯罪热点预测来说,除了使用历史犯罪数据外<sup>[20]</sup>,也需要考虑周边环境要素的影响。本文研究区面积较小,研究区整体地理环境基本一致,因此只考虑微观建成环境对犯罪热点预测的影响。在现有的犯罪热点预测研究中,已经从研究区整体的角度证明了加入周边环境数据作为预测模型的协变量对预测结果的精度有明显提升作用,但具有一定面积的地理空间都是非均质模型,对于每个研究区来讲,研究区内不同的单元有不同的地理环境和社会建成环境,微观空间因素对犯罪分布的影响也呈现出明显的空间异质性<sup>[21-26]</sup>。因此,在犯罪热点预测研究中,除了关注加入协变量后研究区整体预测结果精度是否提升外,还要考虑加入协变量后研究区内预测结果精度变化的空间分异情况,即关注不同区域在加入协变量前后预测结果精度的变化,从而对各种数据对预测模型的提升效果有更深入的理解,这部分内容在已有研究中仍比较匮乏。

本文以公共场所侵财犯罪为研究案件类型,即发生在公共场所的以侵犯财产所有权为主要目的的犯罪,包括公共盗窃和两抢两种类型,根据犯罪学理论,这两类犯罪有相似的发生机理,需要有潜在犯罪者、合适目标及防范缺失3个条件,多发生在环境偏僻、安防设施覆盖率低或环境嘈杂、难以监控的地区,同时这两类犯罪案件数量多、占比大、影响范围广,破坏社会治安环境的稳定,影响居民安

全感,需要进行重点防控。因此,本文采用随机森林算法构建关于公共场所侵财犯罪的热点预测模型,实验并探讨,当把代表环境要素的多源数据加入到预测模型中时,不同类型区域的预测效果会发生怎样的变化。

## 2 研究区概况和数据来源

### 2.1 研究区概况

本文的研究区XT街道,隶属于我国东南沿海特大城市ZG市的HT区,总面积约6.5 km<sup>2</sup>,总人口约18万人,本地人口只有5万余人,占比较小,人口组成以外来人口为主。该街道是HT区城中村总面积最大的街道,有几片规模较大的城中村。复杂的城市用地组成、以外来人口为主的人口结构,也一定程度上影响着该街道的总体治安形势,街道案件数量一直排在HT区内前列。因此,XT街道的治安状况和变化情况受到当地公安部门和上级公安部门的重点关注。而由于该街道所在地区经济发达,警务信息化水平高,对案件的数据记录较准确完整,为犯罪热点预测的开展提供了基础数据保证,所以我们选择XT街道作为研究区,希望能通过基础数据处理和比较准确的犯罪热点预测为当地治安提出合理建议,帮助公安部门科学部署有限的警力资源,由被动式警务转变为主动防控,改善当地治安情况,增强辖区内居民的安全感。

### 2.2 数据来源

#### 2.2.1 历史犯罪数据

本文使用的历史犯罪数据来源于ZG市公安局P-GIS数据库2014-2017年接警数据,根据数据库中记录的坐标信息将案件落点后提取出XT街道范围内的案件点数据。

#### 2.2.2 建成环境数据

结合社会失序理论、日常活动理论及犯罪模式理论等犯罪学理论,并根据历史犯罪数据分布模式及实地考察验证后,选取城中村范围<sup>[27]</sup>、路网密度<sup>[28]</sup>以及POI(餐饮、商场和娱乐设施)密度<sup>[29-30]</sup>这3个变量作为预测模型的协变量。由于研究区内没有明确的城中村划分界限,所以城中村范围变量即通过遥感影像和实地考察,解译并矢量化研究区内的主要城中村范围,标记每一个具体区域是否位于城中村范围中;对于路网密度,根据道道通导航公司提供的各级路网数据,对研究区中的道路数据整理且

归并为3个级别,分别为国内道路、县内道路及可通车道路,利用每个单元中的3类路网的长度之和,除以该单元的面积,作为一个变量;而POI(餐饮、商场和娱乐设施)密度,则从道道通导航公司提供的POI数据中选择餐饮、商场、娱乐这几类POI进行空间插值,得到研究区内的POI点密度面,并赋值到每一个单元,作为一个变量。城中村范围和路网密度共同确定了研究区的整体建成环境,而POI(餐饮、商场和娱乐设施)密度描述了微观尺度的建成环境,并间接反映每个地理单元对人流的吸引力和每个单元安保防范的完善情况与有效性等,在一定程度上描述及概括了每个地理单元的建成环境。因此,把这3个协变量加入到犯罪热点预测模型中,实验探讨预测精度的变化情况。

### 3 研究方法

#### 3.1 预测模型

本研究采用的预测模型为随机森林模型。随机森林算法是由Leo Breiman和Adele Cutler提出的利用多棵树对样本进行训练与预测的分类器<sup>[31]</sup>。随机森林算法是一种集成算法,随机森林由多棵决策树构成,各个决策树是根据随机抽样的样本数据建立而来,是综合考虑每棵决策树的投票来决定最终的分类结果,可以提高决策的准确度和灵敏度。使用随机森林算法进行犯罪热点预测,考虑每一个样本数据的纵向联系即与周边样本数据的横向联系,同时,由于决策树的生成过程具有随机性,不容易发生过拟合现象,可以减少甚至避免犯罪现象发生的随机因素给预测带来的不良影响。这种算法学习能力强、效率高,适合犯罪数据的训练与预测,适用于热点与非热点的分类的预测研究中。

本研究使用随机森林算法,构建基于历史犯罪数据的原始预测模型以及基于历史犯罪数据与犯罪理论相结合的改进预测模型。在使用相同的历史犯罪数据的前提下,结合社会失序理论、日常活动理论及犯罪模式理论等犯罪理论,从建成环境中分别选取城中村范围、路网密度和POI(餐饮、商场、娱乐3类设施)密度3个具有代表性的变量,作为犯罪热点预测模型的协变量,构建改进预测模型,并与原始预测模型的预测结果进行对比分析。

#### 3.2 预测结果评价指标体系

对比随机森林模型加入建成环境协变量前后的预测结果,主要通过网格命中率 $HitR_a$ 、案件命中率 $HitR_n$ 、预测精度指数PAI和案件命中效率 $HitE_n$ 这4个指标来评价。

(1)网格命中率 $HitR_a$ 是指预测正确的热点网格数量与实际热点网格总数之比。

$$HitR_a = \frac{a^*}{A} \quad (1)$$

式中: $A$ 是实际热点网格的总数; $a^*$ 是预测正确的热点网格数量。 $HitR_a$ 数值在0到1之间,在预测热点网格数量一定的情况下,网格命中率越高,即在预测的热点网格中有更大比例的实际热点,说明预测效率越高。

(2)案件命中率 $HitR_n$ 是指预测正确的热点网格中实际发生的案件数量与该时段内研究区总案件数量之比。

$$HitR_n = \frac{n}{N} \quad (2)$$

式中: $N$ 是研究区案件总量; $n$ 是预测热点网格中的实际案件数量。案件命中率越高,说明在被预测为热点的网格中,包含更多的案件,预测准确度越高。

(3)预测精度指数(Prediction Accuracy Index, PAI)<sup>[32]</sup>,指预测的热点区域犯罪密度与整体研究区犯罪密度之比。

$$PAI = \frac{n/N}{a/A} \quad (3)$$

式中: $n$ 是预测热点网格中的实际案件数量; $N$ 是研究区案件总量; $a$ 是预测热点网格的数量; $A$ 是研究区网格总数。 $PAI$ 指数越高,则代表预测的热点区域犯罪密度更高,具有较高 $PAI$ 指数的热点预测方法比另一种具有较低 $PAI$ 指数的方法有更高的每平方面积犯罪案件比。

(4)案件命中效率 $HitE_n$ 是指案件命中率除以预测热点网格数量和实际热点网格数量之比。

$$HitE_n = \frac{HitR_n}{a/A} \quad (4)$$

式中: $HitR_n$ 是案件命中率; $a$ 是预测热点网格数量; $A$ 是实际热点网格数量。案件预测效率值越高,说明在一定数量的预测热点网格数量下,预测精确度更高。

其中,网格命中率 $HitR_a$ 、案件命中率 $HitR_n$ <sup>[33]</sup>和预测精度指数 $PAI$ 是评价犯罪热点预测结果最直观的指标,是网格化犯罪热点预测必不可少的评价指标。但由于加入协变量后,不同类型区域的预测

热点网格数量会发生改变,因此除了使用前3个指标外,在对比不同类型区域加入协变量前后的预测效果时,引进一个新的指标,即案件命中效率  $HitE_n$ 。对某个时段的每类网格来说,实际热点网格数量是固定的,而当预测热点网格数量增多时,就会覆盖更多的网格,当预测热点网格数量增多到与该类网格总数量相同时,则不管案件发生在哪里,该类网格的案件命中率  $HitR_n$  均为1,但此时的预测效果却不一定是好的,因此需要案件命中效率  $HitE_n$  这个指标来评价预测结果。当使用较少的预测热点网格数量,能覆盖更多的案件,则表示案件命中效率高,预测精确度高。当案件命中率  $HitR_n$  及案件命中效率  $HitE_n$  均为1时,则代表预测热点网格数量与实际热点网格数量相等,且正确识别出该类网格中的所有热点网格,达到最好的预测效果。

## 4 预测实验和结果对比

### 4.1 历史案件分析与预测实验设计

#### 4.1.1 时空尺度与历史案件分析

当前我国警力资源紧张,同时单个警员的巡逻范围是有限的,为了达到精确预测以及满足实际警务工作的需要,犯罪热点预测实验的空间尺度应该尽可能的小。因此,采用格网化区域研究的方法<sup>[34]</sup>,依据 Griffith 等<sup>[35]</sup>网格化处理研究区的计算公式及对实际警务工作的调查了解及案件点数据的分布情况,将整个研究区用  $150\text{ m} \times 150\text{ m}$  的网格进行切

割划分。对比  $100\text{ m}$  和  $50\text{ m}$  等更小的空间尺度,以  $150\text{ m}$  划分网格,会使案件点集中地落入某些网格,减少热点网格的偶然性,更稳定的热点网格分布也能反映案件的发生机理和分布规律,可以提高犯罪热点预测的精度和准确度。同时,  $150\text{ m} \times 150\text{ m}$  的范围大致是单个警员在一个时间单位内能覆盖的最大巡逻范围,能更好地与实际警务工作对接。由于研究区面积和单元网格面积较小,当时间单位太小时,研究区内公共场所财产犯罪案件数量少,且案件空间分布更随机,不利于总结犯罪发生机理与规律并进行犯罪热点预测。同时,根据 Rummens<sup>[31]</sup>、柳林等<sup>[32]</sup>的犯罪热点预测实证研究,本研究的时间单位确定为两周。

根据确定的时空尺度,对2014–2017年的历史案件数据进行简单分析。从总体案件量来看,2016年案件量明显比其他3个年份的少,2017年的案件量则稍微多于2014年和2015年。在4年内每两周时段内案件量有所波动,大部分两周时段内的案件量在40–80起之间,平均值为每两周63起。从图1可以看出,4年的案件量曲线有相似的变化趋势,基本是在包含节假日的两周案件量有明显的减少,而节假日后的两周则会回升,每年的1–2月案件量有明显的下降趋势,包含春节假期的两周是每一年中案件量最少的时段。

虽然这四年的年案件总量有所波动,但根据四个年份每年发生案件点的热点制图(图2)呈现的空间分布情况来看,犯罪热点空间位置稳定,没有明显的转移。经统计,若将核密度值前20%的区域定义为犯罪高发区,2014–2017年高发区的重叠度均

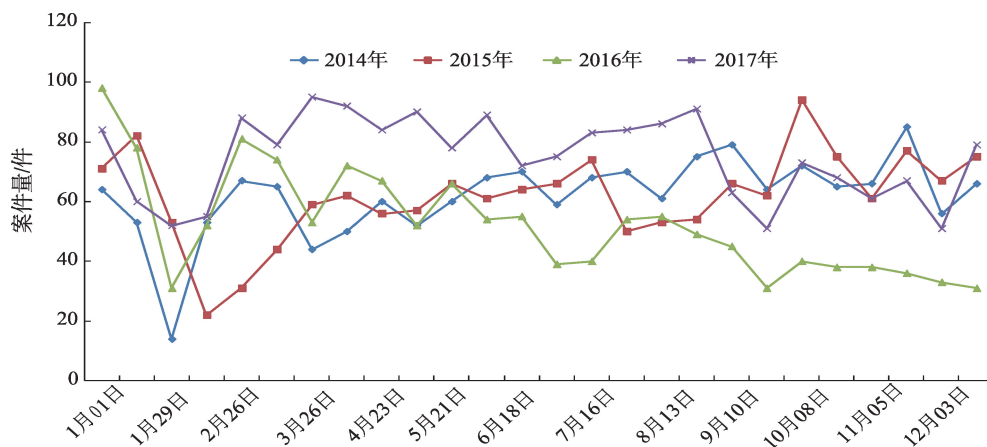


图1 2014–2017年研究区公共场所侵财犯罪两周统计

Fig. 1 Biweekly statistics of public property crime in the study area from 2014 to 2017

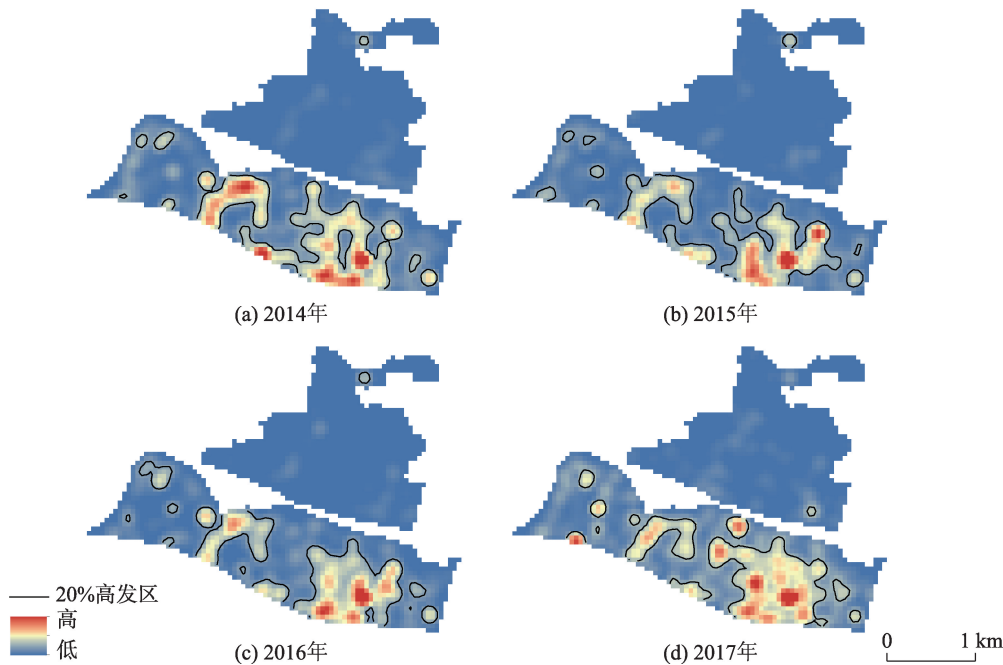


图2 2014–2017年研究区公共场所侵财犯罪热力图  
Fig. 2 Heat map of public property crime in the study area during 2014-2017

在80%左右,而2014、2015、2016年3个年份与2017年对比,重叠度也在75%左右,高发区范围相似。

4.1.2 网格分类与预测实验

根据150 m×150 m的空间尺度把研究区划分成369个网格后,根据2014–2016年共78个双周的历史犯罪分布情况,统计每个网格发生案件的频数,通过K均值聚类法,确定最优聚类个数为3个或4个,于是所有网格划分为稳定高发热点网格、较高发热点网格、偶发热点网格、非热点网格4类,具体网格分布图见图3。结果表明,稳定高发热点网格共有18个,在78个双周中发生案件的频数超过40次,最多的网格共在其中64个双周中都发生了案

件;较高发热点网格共有34个,在78个双周中发生案件的频数大于23次,小于40次;偶发热点网格共有48个,在78个双周中发生案件的频数大于11次,小于23次;剩下的269个网格,在78个双周中,发生案件的频数基本少于10次,甚至有部分网格从未发生案件,因此把这类网格归类为非热点网格。各类型网格数量及案件数量见表1。

在对所有网格进行分类后,分别为研究区整体和各类网格构建随机森林预测模型。随机森林把犯罪热点网格预测当作一个二分类问题,从所有目标网格中预测 $n$ 个在预测时段内有案件发生的热点网格。统计2014、2015、2016年与目标时段同期及4个临近时段内每个网格的案件数量,并计算每个网格对应的3个建成环境协变量的值。分别使用只有历史犯罪数据和有历史犯罪数据与3个协变量的两份数据作为输入数据,2种算法使用相同的输入数据作为训练样本及待预测数据集的变量,根据徐冲等研究DP半岛街头抢劫案件的邻近重复发生模式的结论<sup>[36]</sup>,选取前3个时段对应的输入数据和已有的热点/非热点标签作为训练样本,学习热点和非热点标签的分类规则,并根据预测目标时段对应的输入数据(即待预测数据集)与该时段的预测热点网格数量 $n$ ,从全部网格中选出 $n$ 个得分最高的网格作为预测热点网格。本文以2017年最后一个双周实验为例对实验操作进行说明如表2所示。

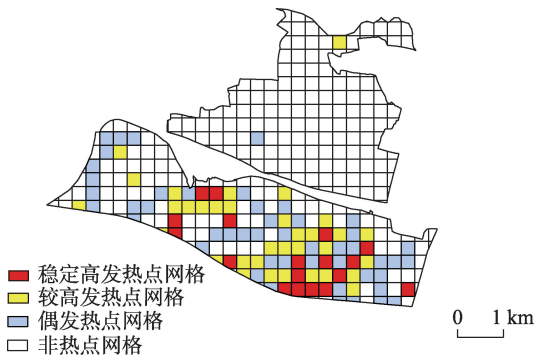


图3 研究区网格历史案发频数分类结果  
Fig. 3 Grid classification by historical crime events in the study area

表1 2014-2017年研究区分类网格案件数量

Tab. 1 Number of classified grid cases in the study area during 2014-2017

类别	稳定高发热点网格	较高发热点网格	偶发热点网格	非热点网格
网格总数/个	18	34	48	269
实际热点网格平均数/个	10	14	13	18
案件数平均值/件	16.346	19.577	17.077	22
案件数标准差/件	5.411	5.147	6.099	7.054

表2 2017.12.17-2017.12.30 犯罪热点预测实验说明

Tab. 2 Description of the crime hotspot prediction experiment (2017.12.17-2017.12.30)

	周期性	邻近性
训练数据集 (2017.12.03-12.16)	2014.12.03-12.16	2017.10.08-10.21
	2015.12.03-12.16	2017.10.22-11.04
	2016.12.03-12.16	2017.11.05-11.18
		2017.11.19-12.02 (建成环境变量)
	输入热点/非热点分类标签	
待预测数据集 (2017.12.17-12.30)	2014.12.17-12.30	2017.10.22-11.04
	2015.12.17-12.30	2017.11.05-11.18
	2016.12.17-12.30	2017.11.19-12.02
		2017.12.03-12.16 (建成环境变量)
	输出热点/非热点分类标签	

4.2 预测结果分析

4.2.1 研究区整体预测结果

加入协变量前后的随机森林模型的研究区整体预测结果主要通过网格命中率  $HitR_a$ 、案件命中率  $HitR_n$  及  $PAI$  3 个指标来进行对比。根据这 3 个评价指标,2017 年 26 个双周实验的预测结果见表 3,表中的模型 A、B 分别表示只使用历史犯罪数据的随机森林模型、使用历史犯罪数据与建成环境协变量数据的随机森林模型。

从 26 个实验预测结果精确度评价指标的平均值来看,模型 B(即加入建成环境协变量数据的随机森林模型)预测结果的网格命中率  $HitR_a$ 、案件命中率  $HitR_n$  及  $PAI$  均高于只使用历史犯罪数据的模型 A(图 4、图 5、图 6),且在 0.05 水平下被检验有显著性差异。并且从标准差来看,模型 B 比模型 A 在 26 个实验中有更稳定的表现。

通过 3 个指标折线图的对比分析,在 2017 年全年 26 个两周预测实验中,从研究区整体预测效果来看,加入建成环境协变量数据的随机森林模型有更好的表现,同样的实验数据和实验要求下,网格命中率 and 案件命中率均高于只使用历史犯罪数据的随机森林模型,并且具有更高的  $PAI$  值,即能正确预测到更多的热点网格,而且预测正确的热点网格有更高的犯罪密度,能覆盖更多的案件。在本研究中,每个双周实验预测的热点网格数量与实际热点网格数量相同,以加入建成环境协变量数据的随机森林模型为例,从整个研究区中寻找每个时段的犯罪热点网格,平均的热点网格命中率可以达到 46.6%,在这接近一半预测正确的网格中,平均可以覆盖到整个研究区 52.3% 的案件量。

从图 4 和图 5 的变化趋势可以发现,不管是模型 A 还是模型 B,预测实验得到结果的案件命中率也随着网格命中率的升降而波动,即一般情况下,网格命中率高,对应该时段的案件命中率也会较高。

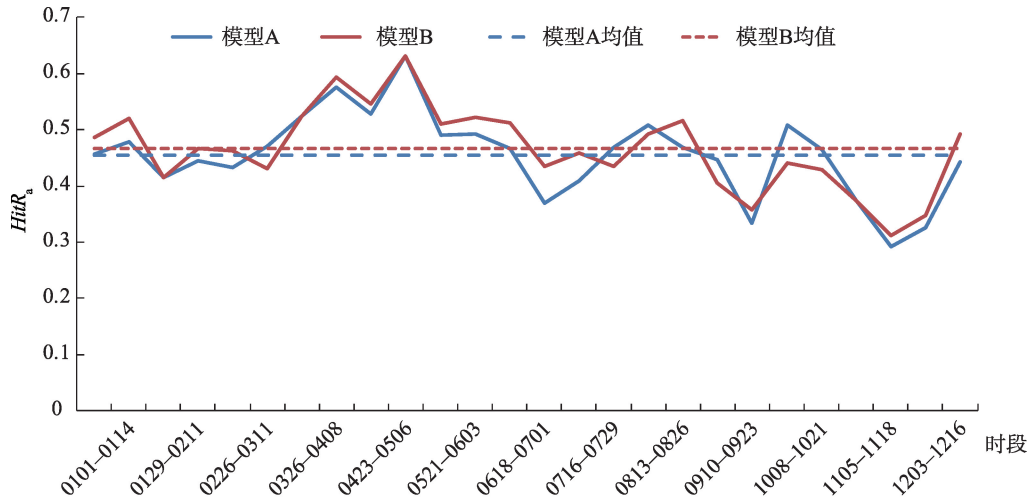
4.2.2 分类网格预测结果

首先对比各类网格的预测热点网格数量,在每个时段整个研究区预测热点网格总数相同的情况下,当加入协变量后,预测热点网格总数分配到 4 类网格的数量发生了变化:当只使用历史犯罪数据时,稳定高发热点网格、较高发热点网格、偶发热点网格、非热点网格的预测热点网格数量平均值分别为 15、18、12、10 个;当加入代表建成环境要素的 3 个变量作为预测模型的协变量后,4 类网格的预测热点网格数量平均值为 16、20、11、8 个。所以,当加入 3 个协变量后,在研究区预测热点网格总数不变

表3 2017年研究区犯罪热点预测实验结果

Tab. 3 Results of the crime hotspot prediction experiment in the study area in 2017

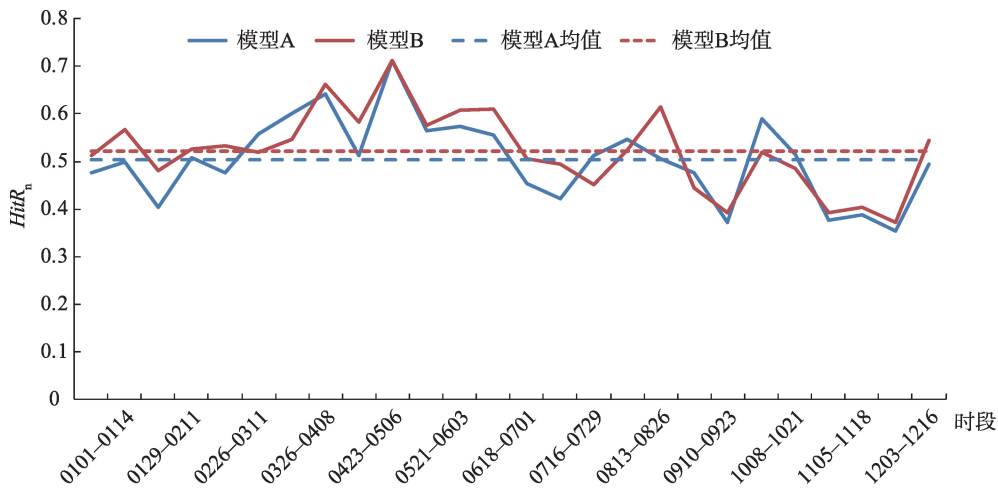
	网格命中率 $HitR_a$		案件命中率 $HitR_n$		$PAI$	
	模型 A	模型 B	模型 A	模型 B	模型 A	模型 B
平均值	0.454	0.466	0.503	0.523	3.395	3.534
标准差	0.075	0.074	0.086	0.084	0.548	0.598



注：横坐标中“0101”表示1月1日，以此类推。

图4 2017年研究区犯罪热点预测结果(网格命中率)

Fig. 4 Graph of crime hotspot prediction experiment in the study area in 2017—Hit Rate of Grid



注：横坐标中“0101”表示1月1日，以此类推。

图5 2017年研究区犯罪热点预测结果(案件命中率)

Fig. 5 Graph of crime hotspot prediction experiment in the study area in 2017 (Hit Rate of Case)

的前提下,稳定高发热点网格和较高发热点网格的预测热点网格数量增多,而偶发热点网格和非热点网格的预测热点网格数量减少,预测的热点网格更集中于稳定高发热点网格与较高发热点网格这两类网格中,而剩下的其他两类网格则更稀疏。

在对比各类网格加入协变量前后的预测效果前,先对评价指标体系进行优化。由于加入协变量后预测热点网格数量会发生改变,因此除了延用前文提及的3个指标外,增加一个新的评价指标,案件命中效率  $HitE_n$ 。表4是4类网格的预测结果精度汇总表,是稳定高发热点网格、较高发热点网格、偶发热点网格及非热点网格在加入建成环境协变量

前后的结果对比,其中A、B分别表示只使用历史犯罪数据的随机森林模型、使用历史犯罪数据与建成环境协变量数据的随机森林模型。

对于稳定高发热点网格来说,加入3个协变量后,预测热点网格数量增多,网格命中率  $HitR_g$ 、案件命中率  $HitR_n$  都有显著的提高(图6),在26个双周实验中,平均可以识别出这类网格89.0%的实际热点网格,覆盖89.7%的案件。同时,从案件命中效率  $HitE_n$  和PAI值来看,模型B的均值也稍高于模型A,也反映了加入协变量后的模型,预测效率高干于只使用历史犯罪数据的原始预测模型,能找到更“热”的实际热点网格。

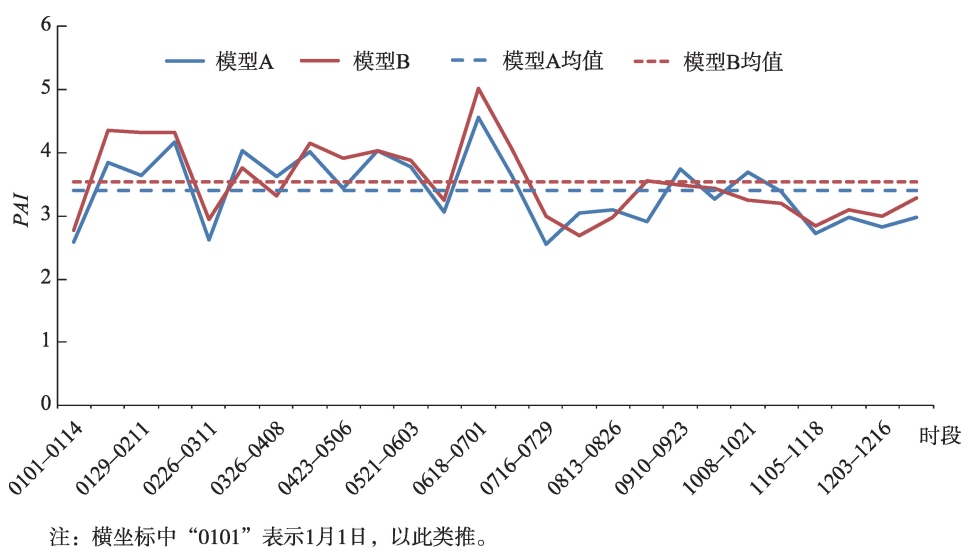


图 6 2017 年研究区犯罪热点预测结果(预测精度指数)

Fig. 6 Graph of crime hotspot prediction experiment in the study area in 2017 Prediction Accuracy Index

表 4 2017 年四类网格预测实验结果

Tab. 4 Results of crime hotspot prediction experiment for four kinds of grids in 2017

	网格命中率 $HitR_g$		案件命中率 $HitR_n$		案件命中效率 $HitE_n$		$PAI$	
	模型 A	模型 B	模型 A	模型 B	模型 A	模型 B	模型 A	模型 B
稳定高发热点网格	0.823	0.890	0.839	0.897	0.584	0.591	1.010	1.025
较高发热点网格	0.604	0.635	0.629	0.677	0.472	0.473	1.179	1.187
偶发热点网格	0.360	0.335	0.395	0.373	0.438	0.424	1.684	1.564
非热点网格	0.184	0.176	0.231	0.223	0.417	0.495	6.906	8.426

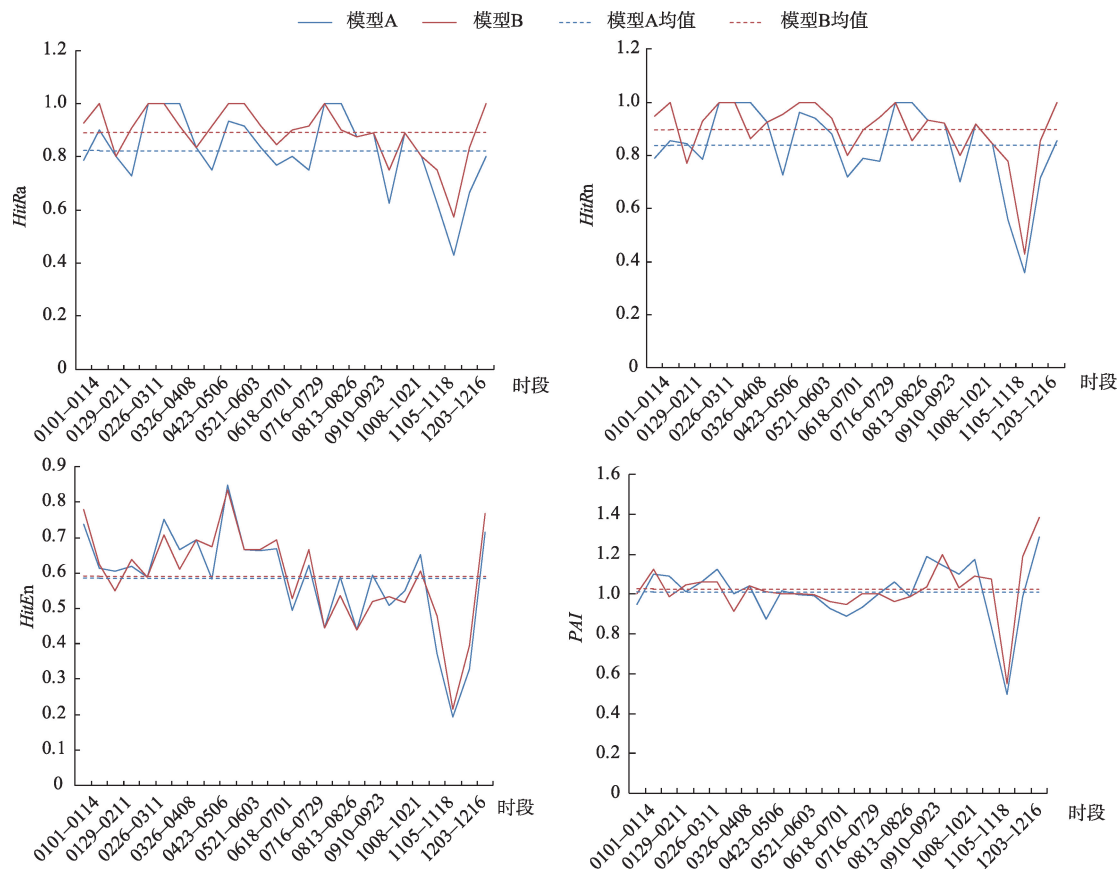
对于较高发热点网格来说,加入3个协变量后,预测热点网格数量增多,网格命中率  $HitR_g$ 、案件命中率  $HitR_n$  都有显著的提高,在26个双周实验中,平均可以识别出这类网格63.5%的实际热点网格,覆盖67.7%的案件。从案件命中效率  $HitE_n$  和  $PAI$  值来看,模型B的均值虽然微微高于模型A的结果,但在统计上并不显著,说明加入协变量后,整体预测效率并没有明显提升。因此,对于较高发热点网格来说,主要是通过增加预测热点网格的数量来提高预测的命中率。

而当加入协变量后,偶发热点网格与非热点网格的预测热点网格数量减少,网格命中率  $HitR_g$ 、案件命中率  $HitR_n$  都有所下降。从案件命中效率  $HitE_n$  和  $PAI$  值来看,偶发热点网格中,模型A的结果稍优于模型B的结果,说明使用历史犯罪数据的原始预测模型的预测效率稍优于加入协变量后的预测模型;非热点网格中,模型B的均值更高,主要是因为这类网格的基数较多,所以即使预测热点网格数量减少,只要命中的网格和案件数量差距不大

的前提下,就会导致  $HitE_n$  和  $PAI$  绝对值明显上升。

对4类网格分别进行分析后,可以看出,当加入代表建成环境要素的3个协变量后,稳定高发热点网格和较高发热点网格的预测结果精确度有所提高,尤其是稳定高发热点网格,有明显的提高,其预测结果精度对比见图7。从图中可以看出,加入协变量后的随机森林模型(模型B)的结果精度均高于只使用历史犯罪数据的原始模型(模型A),且预测结果精度比较稳定,说明历史犯罪数据及3个建成环境协变量的分布能反映这类网格中热点网格的分布规律并满足预测的需要。

稳定高发热点网格和较高发热点网格作为公共场所侵财犯罪集聚高发的区域,需要作为犯罪预测和防控的重点区域。从上述分析结果可以得知,当考虑空间分异时,稳定高发热点网格和较高发热点网格的分区模型的案件命中率分别可以接近0.897和0.677,能根据分区模型的预测结果对这两类高发区域进行更有效的监控管理,对整体研究区的犯罪预测防控具有重要意义。而分区模型的精



注：横坐标“0101”表示1月1日，以此类推。

图7 稳定高发热点网格原始模型与改进模型预测结果对比

Fig. 7 Comparison between original model and improved model prediction of stable high-heat grids

度显著高于整体模型精度,也说明考虑空间分异对提高犯罪热点预测模型精度起重要作用,因此需要根据犯罪空间分异情况,为不同类型区域建立不同的分区模型,通过各分区模型犯罪热点预测模型的优化,促使研究区整体预测精度提高。

4.2.3 具体预测结果展示

经过前期多次实验尝试,4月底到5月的预测

实验效果均比较稳定,同时在2017年的双周实验中,4月23日到5月6日这一双周实验的预测精度居中,有较好的代表性,因此选取这个双周实验的预测结果为例进行展示(图8、图9),这个双周时间内研究区内共有84起案件发生在55个网格中,只使用历史犯罪数据的随机森林模型正确预测其中29个热点网格,覆盖43起案件,使用历史犯罪数据

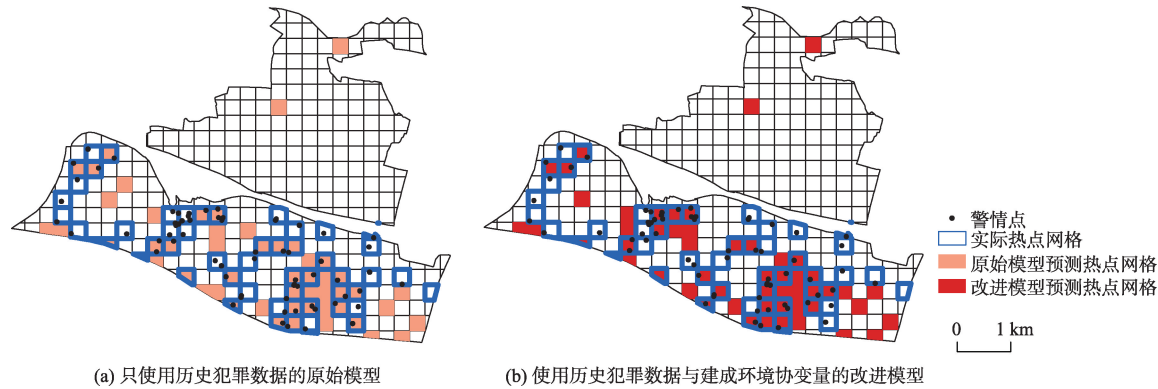
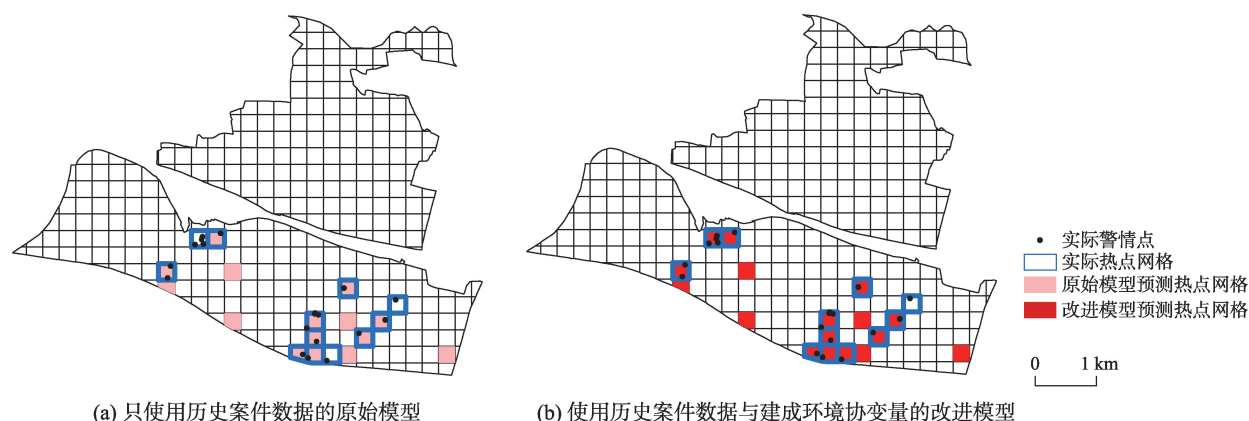


图8 2017年4月23日-2017年5月6日研究区整体预测结果对比

Fig. 8 Comparison of prediction results of the overall study area during 04/23/2017 and 05/06/2017



(a) 只使用历史案件数据的原始模型

(b) 使用历史案件数据与建成环境协变量的改进模型

图9 2017年4月23日-2017年5月6日稳定高发热点网格预测结果对比

Fig. 9 Comparison of prediction results of stable high-heat grids during 04/23/2017 and 05/06/2017

和建成环境协变量的随机森林模型正确预测其中30个热点网格,覆盖49起案件。使用历史犯罪数据及建成环境协变量的随机森林模型的预测结果在各个评价指标都有更好的表现。如果将预测结果运用于实际警务工作中,根据预测结果合理部署警力资源,在这个两周时段内能对超过一半的热点网格进行有效的防控,对热点网格中可能发生的犯罪现象产生震慑作用,能减少案件发生的可能性,并对发生的案件做出及时的响应。

4月23日到5月6日这个时段内,在稳定高发热点网格这类网格中,共有12个实际热点网格,22起案件。只使用历史犯罪数据的随机森林模型正确预测其中9个热点网格,覆盖16起案件;加入建成环境协变量的随机森林模型能正确预测其中11个热点网格,覆盖21起案件,网格命中率、案件命中率、真实案件命中率以及预测精度指数4个指标都反映预测结果明显优于原始模型的预测结果。

#### 4.3 犯罪防控策略

根据本次犯罪热点预测实验的结果以及对研究区的多次实地考察,本研究从犯罪热点预测结果出发,结合警务工作的实际及智能警务的发展情况,整理提出了以下关于犯罪防控策略的建议。

从研究区整体的预测结果尤其稳定高发热点网格和较高发热点网格这两类区域的预测结果可以看出,使用历史案件数据与犯罪理论结合的改进模型能较好地预测这两类区域的犯罪热点网格。这也说明了当考虑环境要素时,对于公共场所财产犯罪来说,这两类区域的某些网格有极高的发案几率。根据3个协变量所反映的每个地理网格单元的建成环境以及实地考察验证情况,在本研究区中乃

至其他具有相似地理环境和社会建成环境的地区,城中村范围内具有商业设施集聚且附近有可通车道路的区域,应作为公共场所侵财犯罪的重点防控区域,需要布控警力资源,负责巡逻等日常管理工作,对重点区域内的犯罪形成监管与震慑,减少犯罪的发生。

对于像本研究中偶发热点网格及非热点网格这类区域,由于历史案件数据及周边环境数据不能反映犯罪发生的规律并得到稳定有效的预测结果,各地方政府及公安部门仍需大力推动辖区范围内监控设备的安装与覆盖,实现全区域的监视和管控,同时能更方便地获取辖区范围内各种实时数据,为辖区的犯罪预测及智能管理提供数据支持。

除此之外,公安部门可以从警力的派遣和调配角度入手,开发智能管理调度系统,除了根据犯罪预测结果将警力主要部署在对重点区域的管控,当案件发生于其他区域时,在通用系统上发出警报,指挥中心根据案发地点、警员位置以及路网信息,通过空间算法计算并派遣附近的警员即刻到现场进行勘察追踪,提高对案件的及时反映及处理能力,合理利用有限的警力资源,对整个辖区范围形成有效的监管。

## 5 结论与讨论

### 5.1 结论

综合上述的犯罪热点预测实验及结果,主要得出以下2个重要结论:

(1)建成环境协变量为犯罪热点预测研究带来积极效果,基于历史犯罪数据和犯罪理论结合的预

测模型精度更高:当根据社会失序理论、日常活动理论和犯罪模式理论选用城中村范围、路网密度及POI(餐饮、娱乐、商场设施)密度3个建成环境要素作为辅助预测的协变量,在一整年共26个双周的连续时段的公共场所侵财犯罪热点预测中,从研究区整体预测结果精度来看,加入协变量的随机森林模型有更好的表现,预测结果的网格命中率、案件命中率以及预测精度指数等评价指标均优于只使用历史犯罪数据的原始模型。

(2)考虑空间分异对提高模型精度能起重要作用:本研究基于研究区内公共场所侵财犯罪的空间分异建立了不同类型区域的犯罪热点预测模型,并实验探讨了建成环境对分区模型精度的影响情况,发现在加入协变量后,稳定高发热点网格和较高发热点网格这两类重点防控区域的预测结果有更优的精度,证明了分区模型的精度显著高于整体模型的精度。

在国内外其他犯罪热点预测实证研究中,Rumens等<sup>[3]</sup>使用历史犯罪数据,在200 m×200 m的网格单元尺度下,使用逻辑回归、神经网络以及逻辑回归与神经网络相结合的3种模型,在双周时段的预测中,两抢案件类型的最高案件命中率为31.97%,最高网格命中率为32.95%;柳林等<sup>[32]</sup>在50 m×50 m的网格单元尺度下,随机森林模型对公共盗窃犯罪的两周和一个月的预测,案件命中率从分别为10.53%和26.22%,而网格命中率分别为9.72%和18.57%;本研究在总体研究区域范围更小、案件数量更少的情况下,在150 m×150 m的两周时段的多次实验预测中,加入协变量的随机森林热点预测模型的平均案件命中率为52.3%,平均网格命中率为46.6%,对比前人研究成果均有显著提高,且长期实验结果稳定。同时,根据基于犯罪空间分异和建成环境的公共场所侵财犯罪热点预测模型的预测结果及周边环境分析总结可以得出,对于该类型犯罪,需要重点关注城中村中主要道路,尤其是城中村范围内小型商业中心附近等人流密集、人口构成复杂的区域,对这些案发可能性高的区域进行更加严格的监控管理。

## 5.2 讨论

本研究仍存在一些不足和待改进的地方。首先是网格单元的切分,本研究按照150 m×150 m的标准直接将整个辖区范围切分成等面积的正方形网格,没有考虑实际用地情况,若根据实际用地情

况对辖区范围进行合理切分,更能对预测结果进行合理的解释,犯罪热点预测的结果也可能进一步优化,同时150 m×150 m的单个网格面积对于实际警务工作来说仍偏大,只有进一步缩小地理尺度进行犯罪热点预测,才有利于警方更精准详细地布置警务工作。其次是构建预测模型的算法选择,本文基于前人研究经验,使用随机森林算法构建犯罪热点预测模型,获得了不错的预测结果,证明了随机森林算法在犯罪热点预测中的有效性,但机器学习包含深度学习、向量机等多种算法,对于其他算法在犯罪热点预测中的实用性和优劣表现的对比,有待下一步实验研究探讨。最后是训练数据的选择,本文已经验证了相同的输入数据对不同类型区域的预测结果有不同的影响,对于稳定高发热点网格和较高发热点网格来说,基于历史犯罪数据和建成环境协变量数据构建的犯罪热点预测模型能很好地预测出这两类区域的犯罪热点网格,但并不适用于偶发热点网格和非热点网格。对于偶发热点网格和非热点网格来说,可能需要通过人流量等其他实时数据发现案件分布的规律并构建合适的预测模型。因此,对于某个地区来说,在合理划分不同类型区域并准确预测不同类型区域的热点网格数量的前提下,需要通过不同数据构建相应的犯罪热点预测模型,以实现整个研究区的犯罪热点预测结果精度的更大幅度的提升,这也是本研究有待进一步改进和探讨的方向。

基于以上研究,随机森林算法在犯罪热点预测研究中具有重要作用与意义。短期犯罪热点预测对警方的短期警务策略和警力资源调控等具有指导作用,本文的短期犯罪热点预测对犯罪热点预测研究和警方的实际警务工作部署有一定的借鉴参考意义。但不同类型的研究区域、不同的犯罪类型,应该根据什么原理,采用那些数据来进行预测,仍需要更多的实验研究来探索,在实证研究的同时进一步地认识与理解犯罪现象发生的原因与规律,才能通过实际措施进行更有效的犯罪防控,稳定社会治安环境,维护城市公共安全。

## 参考文献(References):

- [1] Ratcliffe J H. A temporal constraint theory to explain opportunity-based spatial offending patterns[J]. Journal of Research in Crime and Delinquency, 2016,43(3):261-291.
- [2] Gorr W, Harries R. Introduction to crime forecasting[J]. International Journal of Forecasting, 2003,19(4):551-555.

- [3] Rummens A, Hardyns W, Pauwels L. The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context[J]. *Applied Geography*, 2017,86:255-261.
- [4] 李卫红, 闻磊, 陈业滨. 改进的GA-BP神经网络模型在财产犯罪预测中的应用[J]. *武汉大学学报·信息科学版*, 2017,42(8):1110-1116. [ Li W H, Wen L, Chen Y B. Application of improved GA-BP neural network model in property crime prediction[J]. *Geomatics and Information Science of Wuhan University*, 2017,42(8):1110-1116. ]
- [5] Rodriguez-Galiano V F, Ghimire B, Rogan J, et al. An assessment of the effectiveness of a random forest classifier for land-cover classification[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2012,67:93-104.
- [6] 马慧娟, 高小红, 谷晓天. 随机森林方法支持的复杂地形区土地利用/土地覆被分类研究[J]. *地球信息科学学报*, 2019,21(3):359-371. [ Ma H J, Gao X H, Gu X T. Land use/land cover classification of complex terrain areas supported by random forest method[J]. *Journal of Geo-information Science*, 2019,21(3):359-371. ]
- [7] 卢睿, 李林瑛. 基于随机森林的犯罪预测模型[J]. *中国刑警学院学报*, 2019(3):108-112. [ Lu R, Li L Y. Crime prediction model based on random forest[J]. *Journal of China Criminal Police College*, 2019(3):108-112. ]
- [8] Bogomolov A, Lepri B, Staiano J. Once upon a crime: Towards crime prediction from demographics and mobile data[C]. *International Conference on Multimodal Interaction ACM*, 2014:427-434.
- [9] Levine N. Commentary the "Hottest" part of a hotspot: Comments on "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime"[J]. *Security Journal*, 2008,21:295-302.
- [10] 王雨晨, 过仲阳, 王媛媛. 基于随机森林的犯罪风险预测模型研究[J]. *华东师范大学学报(自然科学版)*, 2017(4): 89-96. [ Wang Y C, Guo Z Y, Wang Y Y. Research on criminal risk prediction model based on random forest[J]. *Journal of East China Normal University (Natural Science Edition)*, 2017(4):89-96. ]
- [11] 孙菲菲, 曹卓, 肖晓雷. 基于随机森林的分类器在犯罪预测中的应用研究[J]. *情报杂志*, 2014,33(10):148-152. [ Sun F F, Cao Z, Xiao X L. Application research of classifier based on random forest in crime prediction[J]. *Intelligence Magazine*, 2014,33(10):148-152. ]
- [12] Cohen L E, Felson M P. Social change and crime rate trends: A routine activity approach[J]. *American Sociological Review*, 1979,44(4):588-608.
- [13] Merleau M. The structure of behavior[M]. Boston: Beacon Press, 1963:188-189.
- [14] Schutz A. The phenomenology of the social world[M]. Chicago: Northwest University Press, 1967:34.
- [15] Brantingham P L, Brantingham A J. Nodes, paths and edges: Considerations on the complexity of crime and the physical environment[J]. *Journal of Environmental Psychology*, 1993(13):3-28.
- [16] 周东平. 西方环境犯罪学: 理论、实践及借鉴意义[J]. *厦门大学学报(哲学社会科学版)*, 2014(3):20-28. [ Zhou D P. Western environmental criminology: Theory, practice and reference significance[J]. *Journal of Xiamen University (Philosophy and Social Sciences Edition)*, 2014(3):20-28. ]
- [17] Qin X, Liu L. Evaluating the relationships of bus transit with street and off-street robberies[J]. *Professional Geographer*, 2015,68(2):1-11.
- [18] Hart T C, Miethe T D. Street robbery and public bus stops: A case study of activity nodes and situational risk [J]. *Security*, 2014,27(2):180-193.
- [19] Blair L, Wilcox P, Eck J. Facilities, opportunity, and crime: An exploratory analysis of places in two urban neighborhoods[J]. *Crime Prevention & Community Safety*, 2017,19(1):1-21.
- [20] 段炼, 党兰学, 胡涛, 等. 融合历史犯罪数据的疑犯社会活动位置预测[J]. *地球信息科学学报*, 2018,20(7):929-938. [ Duan L, Dang L X, Hu T, et al. Prediction of social activity location of suspects with historical crime data[J]. *Journal of Geo-information Science*, 2018,20(7):928-938. ]
- [21] Zhang H, Song W. Addressing issues of spatial spillover effects and non-stationarity in analysis of residential burglary crime[J]. *Geojournal*, 2014,79(1):89-102.
- [22] Arnio A N, Baumer E P. Demography, foreclosure, and crime: Assessing spatial heterogeneity in contemporary models of neighborhood crime rates[J]. *Demographic Research*, 2012,26:449-488.
- [23] Graif C, Sampson R J. Spatial heterogeneity in the effects of immigration and diversity on neighborhood homicide rates[J]. *Homicide Studies*, 2009,13(3):242.
- [24] David C Wheeler, Lance A Waller. Comparing spatially varying coefficient models: A case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests crime rates and their relationships to alcohol outlets and illegal drug arrests[J]. *Journal of Geographical Systems*, 2009,11(1):1-22.
- [25] Malczewski J, Poetz A. Residential burglaries and neighborhood socioeconomic context in London Ontario: Global and local regression analysis[J]. *Professional Geographer*, 2005,57(4):516-529.

- [26] Macbeth E, Ariel B. Place-based statistical versus clinical predictions of crime hot spots and harm locations in Northern Ireland[J]. *Justice Quarterly*, 2019,36(1):93-126.
- [27] 杨刚斌,柳林,何深静,等.广州门禁小区入室盗窃受害率与内部环境分析[J].*人文地理*,2016,31(3):45-51. [ Yang G B, Liu L, He S J, et al. Analysis on the victimization rate and internal environment of burglary in guangzhou access control community[J]. *Human geography*, 2016,31(3):45-51. ]
- [28] 柳林,杜方叶,肖露子,等.不同类型道路密度对公共空间盗窃犯罪率的影响——基于ZG市的实证研究[J].*人文地理*.2017,32(6):32-38. [ Liu L, Du F Y, Xiao L Z, et al. The impact of different types of road density on the crime rate of public space theft: An empirical study based on ZG City[J]. *Human Geography*, 2017,32(6):32-38. ]
- [29] 宋广文,肖露子,周素红,等.居民日常活动对扒窃警情时空格局的影响[J].*地理学报*,2017,72(2):356-367. [ Song G W, Xiao L Z, Zhou S H, et al. The influence of residents' daily activities on the time and space pattern of plagiarism[J]. *Journal of Geographical Sciences*, 2017,72(2): 356-367. ]
- [30] 孙兆瑞.长春市城乡结合部建筑外环境防卫安全设计研究[D].长春:吉林建筑大学,2017. [ Sun Z R. Study on the safety design of the exterior environment of the urban and rural joint zone in Changchun City[D]. Changchun: Jilin University, 2017. ]
- [31] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1):5-32.
- [32] 柳林,刘文娟,廖薇薇,等.基于随机森林和时空核密度方法的不同周期犯罪热点预测对比[J].*地理科学进展*, 2018,37(6):761-771. [ Liu L, Liu W J, Liao W W, et al. Comparison of hotspot predictions of different periodic crimes based on random forest and space-time nuclear density method[J]. *Progress in geography*, 2018,37(6): 761-771. ]
- [33] Chainey S, Thompson L, Uhlig S. The utility of hotspot mapping for predicting spatial patterns of crime[J]. *Security Journal*, 2008,21(1-2):4-28.
- [34] Liu M, Lu T. A hybrid model of crime prediction[J]. *Journal of Physics: Conference Series*. DOI: 10.1088/1742-6596/1168/3/032031.
- [35] A G D. Amrhein C G, Desloges J R. Statistical analysis for geographers[J]. *Journal of the American Statistical Association*, 1999,94(446):654.
- [36] 徐冲,柳林,周素红,等. DP半岛街头抢劫案件的临近重复发生模式[J].*地理研究*,2015,34(2):384-394. [ Xu C, Liu L, Zhou S H, et al. The recurrence pattern of street robbery cases on the DP peninsula[J]. *Geographic research*, 2015,34(2):384-394. ]