

DOI: 10.5846/stxb202005081143

姚尧, 尹瀚琦, 李歆艺, 郭紫锦, 任书良, 王若宇, 关庆锋. 基于多源地理数据精细尺度的武汉市人居环境新型冠状病毒肺炎传播风险评估. 生态学报, 2021, 41(19): 7493-7508.

Yao Y, Yin H Y, Li X Y, Guo Z J, Ren S L, Wang R Y, Guan Q F. Fine-scale risk assessment of COVID-19 in Wuhan based on multisource geographical data. Acta Ecologica Sinica, 2021, 41(19): 7493-7508.

基于多源地理数据精细尺度的武汉市人居环境新型冠状病毒肺炎传播风险评估

姚 尧¹, 尹瀚琦¹, 李歆艺², 郭紫锦¹, 任书良¹, 王若宇^{3,*}, 关庆锋¹

1 中国地质大学(武汉)地理与信息工程学院, 武汉 430078

2 武汉大学测绘遥感信息工程国家重点实验室, 武汉 430072

3 爱丁堡大学地球科学学院, 英国

摘要: 新型冠状病毒肺炎的迅速传播和扩散警示着疾病风险评估的重要性。但现有的风险评估方法受数据限制, 缺少实时性和准确性。此外, 多数研究以行政统计单元作为分析尺度, 存在可变面元问题。为解决这些问题, 耦合精细尺度下武汉市疫情数据及多源地理数据, 基于随机森林算法构建社区尺度的市域疫情传播风险评估模型并进行了疫情风险制图。模型测试精度达到 0.85, Kappa 系数达到 0.70。此外, 本研究还建立基于随机森林算法的社区及场所尺度的“空间变量-感染风险”模型, 评估了不同场所设施疫情传播的风险程度。研究表明, (1) 武汉中心区域感染风险最高并呈现出向外围递减的趋势; (2) 感染风险排名前五的一级场所类型分别为购物服务、医疗服务、金融服务、交通设施以及公共设施; (3) 小学、中学的疫情传播风险较低, 而高等院校传播风险较高; (4) 社区尺度下的疫情风险程度, 预测购物场所与交通场所是疫情传播风险最高的驱动因子。本研究基于精细尺度提出风险评估新方法, 可为未来疾病风险评估提供新思路, 为疫情防控提供决策支持, 人民群众提供安全保障。

关键词: 风险评估; 精细尺度; 多源地理数据; 新型冠状病毒肺炎; 驱动因子

Fine-scale risk assessment of COVID-19 in Wuhan based on multisource geographical data

YAO Yao¹, YIN Hanyu¹, LI Xinyi², GUO Zijin¹, REN Shuliang¹, WANG Ruoyu^{3,*}, GUAN Qingfeng¹

1 School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, China

2 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China

3 Institute of Geography, School of GeoSciences, University of Edinburgh, Edinburgh, UK

Abstract: The severe outbreak of coronavirus disease 2019 (COVID-19) demonstrates the importance of disease risk assessment. The existing risk assessment methods are limited by the real time and accuracy of data. Most of them take the administrative statistical unit as the analysis scale, which has modifiable areal unit problem (MAUP). First, based on a random forest method, we integrated COVID-19 transmission data at community scale and multisource geospatial data to map COVID-19 disease outbreak risks at fine scale. The experimental results (overall accuracy = 0.85, Kappa = 0.70) indicated the feasibility of the model. Second, we built a spatial variable-infection risk model at community and place scale to assess the risk degree of epidemic spread in different places and facilities. Last, we analyzed the possibly spatial drivers of disease transmission. The results show that (1) the central area of Wuhan city has the highest risk of infection and the

基金项目: 国家自然科学基金青年基金项目(41801306)

收稿日期: 2020-05-08; 修订日期: 2021-04-16

* 通讯作者 Corresponding author. E-mail: r.wang-54@sms.ed.ac.uk

risk map presents a trend of decreasing from the center to the periphery; (2) The top five facilities with the highest risk of COVID-19 infection are shopping, medical, financial, transportation and public facilities; (3) The transmission risk of the epidemic is low in primary and middle schools, but high in colleges and universities; (4) The model determines the degree of epidemic risk at the community scale and predicts that shopping and traffic places are two most significant driving factors with the epidemic outbreak. In conclusion, this study suggests a new method of disease risk assessment based on a fine scale, which can pave the way for future disease risk assessment.

Key Words: Disease risk assessment; Fine scale; Multisource geospatial data; Coronavirus disease 2019; Spatial driving factors

在全球范围的突发性公共卫生事件发生时,及时准确的风险评估在提高大众警惕^[1]、控制疾病传播^[2]、稳定大众情绪^[3]、维稳社会环境^[4]方面具有不可或缺的重要作用。风险评估是国家公共卫生机构的核心职能之一,指联合各主要公共卫生机构,评估突发事件的威胁程度,基于已有证据提出事件预警,做好应对准备^[4]。2020年1月,新型冠状病毒肺炎(COVID-19)开始大规模爆发^[5]。1月20日,国家卫健委发布1号公告,将新型冠状病毒感染的肺炎纳入《中华人民共和国传染病防治法》规定的乙类传染病,并采取甲类传染病的预防、控制措施^[6]。然而情况恶化,该肺炎疫情逐渐在全球扩散,成为国际关注的突发性公共卫生事件^[7]。对该疫情的研究主要集中在流行病学^[8-10]、药理学^[11-12]、病理学^[13-16]方面。从流行病学传播角度,及时研判各地区疫情传播风险,预测疫情发展走向,可为政府防控提供决策支持,为人民群众提供安全保障。

风险评估起初多采用专家会商法^[1-4]进行,邀请各方专家进行评估,得出应对建议。该类方法多基于先验知识的积累,存在一定的主观因素;新型疾病的出现会极大增强专家会商法的局限性。现如今处于大数据时代,大数据具有大容量(Volume)、多样化(Variety)、快速化(Velocity)、价值密度低(Value)、真实性(Veracity)等5V特征^[17],仅靠人力对该量级的数据作分析难度极大。反之,以数据驱动方式进行风险评估的实验可行性大大增强。

近十年来,数据驱动方法逐渐被引入风险评估领域,如相关性分析^[18]、经典时间序列分析^[19]、控制图法^[20]、时间聚类分析^[21]、传染病动力学模型^[22]等。例如,杨维中等人采用控制图法、基于时间聚类分析建立国家层面的预警模型,得到较好的预警结果^[21]。但上述方法多从时间角度回顾疾病爆发特征,总结病理学特征,提出疫情爆发的预测,是一种回溯性的风险评估^[23],并不能对疫情性质进行评估,并且研究尺度较为宏观。Deng等人在对手足口病的风险评估中结合时空聚类的方法建立风险评估模型,得到模型拟合效果较好,提出幼儿人口占比、男女比例、日照总量是影响地区手足口病的主要风险因子^[24]。但该研究以广东省县级行政区划为基本单位,存在“主观划分单元”的可变面元问题(Mdifiable areal unit problem, MAUP)^[25]。随着机器学习的出现,贝叶斯网络也被应用于突发事件风险评估领域。如Jiang等人根据贝叶斯网络^[26]实现的疾病预测系统可以预测疫情发生的规模及持续时间,并估计其爆发时间^[27];Liao等人提出一种基于贝叶斯信念网络的不同尺度疾病爆发风险评估模型,从县级尺度评估手足口病的风险,从乡镇尺度分析麻疹疫情情况^[23]。上述两种方法综合多元变量,基于机器学习算法进行风险评估,分析更精细化,对特定疾病的防控更具有指导意义;但这些方法受数据实时性与准确度的影响,同时尚未考虑精细尺度下人群空间聚集因素对疫情传播的作用,无法预测城市内部的疾病传播风险分布。

为积极响应国家恢复经济社会秩序的号召,各地政府致力于标定市域内各地区的不同风险等级,以实施精细尺度下的精准施策。故而,城市内部范围的疫情传播风险评估研究被提上日程。然而目前关于疾病风险评估的研究大多基于城市甚至国家尺度^[19,23,27-30],尚未有具体至社区或建筑场所尺度的疾病风险评估研究。其主要原因在于缺少描述精细尺度的可靠数据和相应模型。描述风险评估需要的精细尺度数据包括高精度人口数据及发病情况^[31],发病情况可通过每日社区通报记录收集,而囿于技术原因^[32],精细尺度的人口数据制图一度成为学界的研究痛点,导致精细尺度的疾病风险评估研究的发展也停滞不前。

随着地理大数据时代的到来,基于位置的服务服务技术(Location based service, LBS)的发展以及多源地理数据的繁盛使得精细尺度下的空间研究分析成为可能^[33]。地理基础服务(LBS)技术由于具有可移动定位、定位精度高、交互性强等特点,被广泛应用于动态地理空间信息相关的研究中^[34]。多源地理数据是指由大众采集并向大众提供的开放地理数据(Volunteered geographic information, VGI)^[35-37]。诸如兴趣点(Points of interest, POIs)、GPS 行驶数据、用户签到数据、手机信令数据等^[38]。研究表明^[33,37,39-42],多源地理数据凭借其在微观层面与人类活动的相关性,可以有效反映人口分布状况及人类活动状态。目前,公共健康领域同样也有多源地理大数据的应用^[43-45]。但是该类数据多用于心理健康的分析研究,尚未有针对精细尺度下疫情风险评估的相关研究。

综上所述,已有研究存在以下不足:(1)面对新型疾病的暴发,传统的基于先验知识的风险评估方法^[1,4]在数据处理量级和评估准确性上均存在极大局限性;(2)现有的根据数据驱动方法建立的风险评估模型^[22-24,27-29],多从时间角度与先验知识回顾总结疾病暴发特征,并不能对新型疾病性质进行评估,且实验精度受数据实时性与准确度影响,多未考虑精细尺度下人群空间聚集因素对疾病传播的风险,无法预测城市内部的疾病传播风险分布研究;(3)随着地理大数据时代的到来,多源地理数据凭借其在微观层面与人类活动的相关性,使得精细尺度下的空间研究分析成为可能,但目前精细尺度下的公共健康领域研究多为心理健康或老龄化问题相关,尚未涉及疾病传染风险评估领域。

基于前人研究的不足,为定量分析微观尺度下城市生态对于新型冠状病毒肺炎传播的关系,本研究首先建立了基于数据驱动的地理空间导向风险评估框架,引入多源地理大数据作为精细尺度下人群空间活动的表征,耦合社区尺度的患者数据,利用非线性机器学习算法建立多元决策模型,实现社区尺度的疫情传播风险等级划分。最后根据评估场所的不同权重,分析疫情传播潜在的驱动因素及判断缘由,进而对疫情在社区尺度下的传播规律进行了总结和应用,讨论并提出关于复工复产后疫情防控的相关建议。

1 数据来源及研究方法

1.1 研究区概况

本研究选定武汉市市域空间作为研究区域。以地理位置而言,武汉位于中国中部、湖北省东部、长江与汉江交汇处,是湖北省省会,处于北纬 29°58'—31°22',东经 113°41'—115°05'之间。以行政区划而言,武汉下辖江岸、江汉、硚口、汉阳、武昌、青山、洪山、蔡甸、江夏、黄陂、新洲、东西湖、汉南 13 个行政区。作为新型冠状病毒肺炎疫情(COVID-19)的爆发区,武汉市具有较高研究价值。截至发稿前(2020 年 4 月 29 日 24 时),武汉市累计报告确诊的病例 50333 例,累计治愈出院 46464 例^[46]。该地区感染患者数据量较大,以此为根据的大数据研究具有一定可信度。

1.2 数据来源

1.2.1 社区尺度的新冠病毒肺炎患者数据

在社区尺度上,本研究采集了武汉市共计 983 个小区截至 2020 年 2 月 16 日通报感染的病例数据(含确诊和疑似病例)及其地理位置,整理制作了各小区感染人数气泡图,如图 1 所示。

该图按照自然断点法(Natural breaks)^[47],将小区按照感染人数分为 5 个等级,1—5 人为 0 级(低风险),5—15 人为 1 级(中低风险),15—30 人为 2 级(中等风险),30—45 人为 3 级(中高风险),45 人及以上为 4 级(高风险)。本研究以单位感染人数表示感染风险。

1.2.2 空间辅助数据

城市兴趣点是影响居民公共健康的重要因素之一^[48]。本研究使用 2018 年武汉市范围内的高德地图 POIs 数据,每条 POIs 数据包括 ID、名称、类型、行政区、地址、经度、纬度 7 个属性,共 886415 条。根据已有研究^[49],将 POIs 类型分成 14 大类,包括汽车服务类、餐饮服务类、购物服务类、生活服务类、体育休闲类、医疗服务类、住宿服务类、风景名胜类、科教文化类、交通设施类、金融类、公司企业类、道路附属类、公共设施类等,

其下又包含 40 小类,具体分类如图 2 所示。



图 1 已采集的社区感染患者数量及分布情况,截止至 2020 年 2 月 16 日

Fig.1 Community Distribution of infected patients, statistics by February 16th 2020

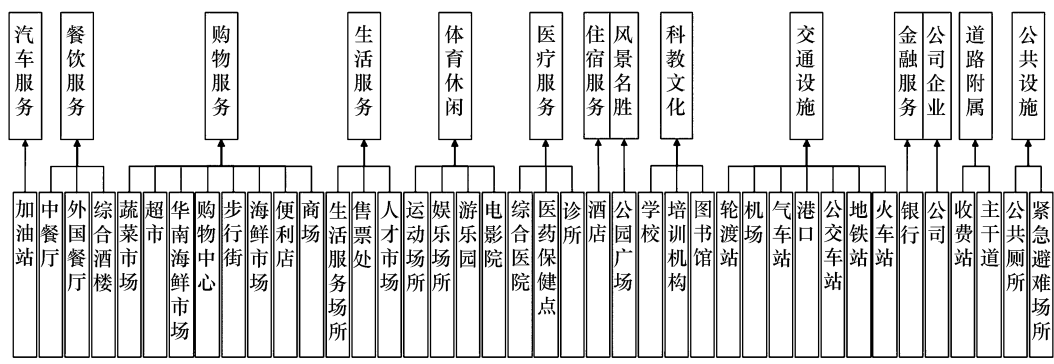


图 2 POIs 分类架构图

Fig.2 The architecture of POIs

POIs: 兴趣点 Points of interest

人群在路网上的空间交互移动对流行病情的时空扩散具有直接作用^[50],因此需要探究其与相邻路网的距离远近。研究区域的路网数据来源于 OpenStreetMap (OSM) 网站 (<http://www.openstreetmap.org>)。根据数据需要,本研究保留了 2019 年 OSM 路网中的一级、二级及干线道路,用以计算各小区与一二级道路的距离远近。

已有流行病学研究认为人口密度与病情的传播具有较强的相关性^[51],数据来自腾讯用户实时定位数据 (Real-time tencent user density, RTUD)^[41]。通过网络爬虫技术,可在腾讯时序数据库 CTSDB 日志存储与监控分析网站 (<https://cloud.tencent.com/act/event/ctsdb>) 获取该类数据。获取的数据时间跨度为 2016 年 9 月 19 日至 9 月 25 日,时间分辨率为 1h,空间分辨率为 0.00025° (约为 26.5m)。本研究基于时序数据压缩算法^[52],将数据按街道尺度进行压缩,并将其以自然间断点标准^[47]进行分级,最后进行归一化处理。从而获得了低信息损失、而高压缩率的 RTUD 数据,作为代表精细尺度下人口密度的特征变量。

1.3 研究方法

为评估不同尺度下疫情传播的风险程度,本研究首先通过网格法对市域范围的社区数据进行空间采样,得到 1986 个基本单元。然后耦合全市域兴趣点 (POIs) 数据、纠偏后的 LBS 人口密度数据以及与主要道路的

距离数据作为特征向量,将社区感染分级情况作为目标拟合值,基于非线性机器学习方法构建“空间变量-感染程度”的非线性风险程度评估模型。根据特征权重,可进一步探究疫情传播的驱动因素,并分析学校周边的感染风险,以此为武汉市复工、复产、复学等疫情防控安排提出建议。具体技术路线如图 3。

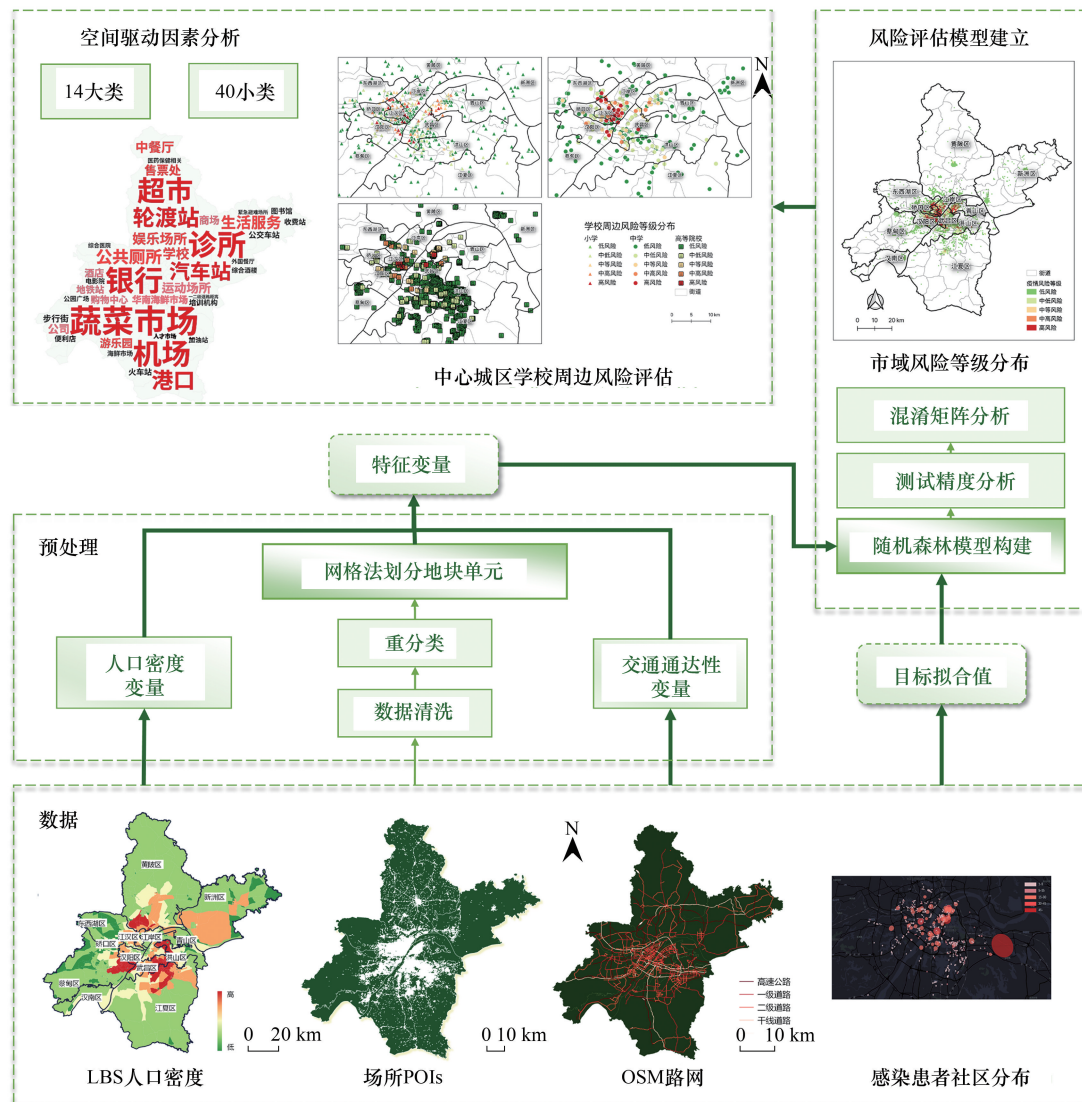


图 3 技术路线图

Fig.3 Work flow chart

LBS: 基于位置的服务 Location based service; POIs: 兴趣点 Points of interest; OSM: 开放街道地图 OpenStreetMap

1.3.1 格网法空间采样

本研究采用规则格网划分武汉市社区。由于收集数据尺度不统一(社区与社区的定义混杂),导致直接用行政统计单元进行空间划分存在一定问题,如聚类质量受影响,结果无法解释;统计数据无法具体至每个社区,部分以社区(社区抱团聚集)形式记录^[53]。而采用规则格网进行重采样,相邻单元间存在隐性空间关系,统计数据依附基本单元而存在,避免了尺度问题^[54],有利于风险评估模型的构建。主要过程如下。

基于已知的行政统计单元进行规则格网的划分,将空间划分成 528m×611m 的单元网格,并给每个单元按照行政数据的位置关系赋予属性数据。本研究共划分 1986 个基本地块单元。

1.3.2 基于随机森林算法的感染风险评估

随机森林(Random forest, RF)是基于集成学习的思路,将多棵决策树集成的一种算法^[55]。相较于大部

分机器学习方法,它具有高精度、能处理高维特征样本、自动生成无偏估计、可控过拟合现象等优点^[56]。

为评估武汉市全域的疫情感染风险,本研究将上述数据包括市域兴趣点(POIs)数据、LBS 人口密度数据以及第一、二级的路网数据作为特征向量 X_i ,并且将社区感染分级情况作为目标值 Y 。通过 RF 模型计算每个因变量的影响权重,对高维因变量进行降维处理,探索特征变量之间的复杂相关性。考虑数据数量限制,设置 130 棵决策树与随机占比 30% 的袋外数据,基于 OOB 交叉验证对 RF 模型进行迭代修正,将随机训练与预测过程重复 100 遍,已获得最可靠的平均净精度结果。

1.3.3 模型评价方法

本研究采用总体精度(Overall accuracy, OA)、均方根误差(Root mean squared error, RMSE)、平均相对误差(Mean relative error, MRE)及 Kappa 系数评价风险传播模型的精度。

Kappa 系数采用一种离散多元统计技术来计算和评价最终指标,基于混淆矩阵,常用来评估二者间的一致性,可以用来描述总体一致性和分类一致性^[57]。该统计值计算方法如下^[57-59]:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (1)$$

式中, p_o 为分类图像与参考图像一致性的概率,即各个类别被划分正确的概率总和; p_e 为随机一致性概率。

$$p_o = \frac{\sum_{i=1}^r x_{ii}}{N} \quad (2)$$

$$p_e = \frac{\sum_{i=1}^r (x_{i+} x_{+i})}{N^2} \quad (3)$$

式中, r 为混淆矩阵行数, x_{ii} 为混淆矩阵中对角线上的样本数,指分类正确的数目, x_{i+} 和 x_{+i} 分别为第 i 行与第 i 列的样本总数, N 为混淆矩阵中的总样本数。

对于 Kappa 系数结果与分类精度的关系,Landis 与 Koch 将计算结果分成五组,用以对应不同的一致性结果:计算结果位于 0.00—0.20 为一致性极低,位于 0.21—0.40 为一致性一般,位于 0.41—0.60 为一致性中等,位于 0.61—0.80 为一致性较高,位于 0.81—1.00 为几乎完全一致^[60]。

1.3.4 空间驱动因子分析

根据风险评估模型,可得到市域范围内各社区的疾病风险等级分级。为探究各社区疾病传播的空间影响变量,可以根据随机森林算法中各决策树的划分情况及其结点的权重分布,计算得到每一个社区单元疾病风险等级的组成参数权重,即为空间驱动因子。本研究重定义后的 POIs 具有二层结构,故而空间驱动因子也相对应有二级权重。将二级权重定义为 POIs 中各场所对其所在社区格网单元评估的影响权重。

本文从纵向和横向两个角度进行空间驱动因子的分析。

2 结果与分析

本节基于已有框架进行模型构建,并且完成全市域范围的社区尺度疫情传播风险等级划分。再者,对生成模型结果进行性能评估,由局部到整体进行实地结果验证,分析该模型的合理性。并且基于空间可视化方法的疫情传播空间驱动因素分析,定量其权重,定性其成因,根据构成不同场所对精细单元传播风险权重高低,采用词云与柱状图等形式,挖掘社区尺度存在的疫情风险及不同级别驱动因素。

2.1 精细尺度感染风险评估分析

通过感染者社区分布情况、POIs 数据、LBS 人口密度数据及路网数据,基于随机森林算法建立模型,得到了 COVID-19 武汉市全域和中心城区的社区感染风险分布情况。该模型的测试精度达到了 0.85,分类结果的 Kappa 系数达到 0.70,说明模型拟合程度较高,与目标值有很高的一致性。具体精度及相应混淆矩阵如表 1 和表 2 所示。

表 1 精度分析				
Table 1 Precision Analysis				
统计描述 Statistical description	总体精度 OA	均方根误差 RMSE	平均相对误差 MRE	卡帕系数 Kappa
值 Value	0.914	0.159	0.103	0.853

OA: 总体精度 Overall accuracy; RMSE: 均方根误差 Root mean squared error; MRE: 平均相对误差 Mean relative error

表 2 混淆矩阵						
Table 2 Confusion matrix						
		预测 Predicted results				
		低风险 Low risk level	中低风险 Medium to low risk	中等风险 Medium risk level	中高风险 Medium to high risk	高风险 High risk level
实际	低风险	1340	12	0	1	10
Actual results	中低风险	46	223	0	0	8
	中等风险	13	12	32	2	2
	中高风险	3	0	2	28	3
	高风险	14	5	1	2	209

基于已得到的风险评估模型,对武汉市域 1968 个社区进行风险评估,结果如图 4 和图 5 所示。前者为武汉市市域社区的风险等级分布,后者为武汉市都市发展区范围内社区的风险等级分布。可以发现作为这次疫情暴发的中心,汉口的中心区域(江汉区、江岸区、硚口区)不仅是重灾区,也是武汉市感染风险最高的区域。青山区作为工业老区,“熟人社区”导致了疫情扩散。汉阳区明显呈现沿着主干道(汉阳大道)的疫情传播趋势。高校重地洪山区感染风险程度较低,猜测这与今年高等院校放假普遍较早(1 月 10 日左右)有密切关系。同时,洪山区光谷片区的感染主要集中于人群密集的生活社区。相反,年轻人集中生活和工作的关山区,相比于武昌和汉口中心区域感染风险程度较低。

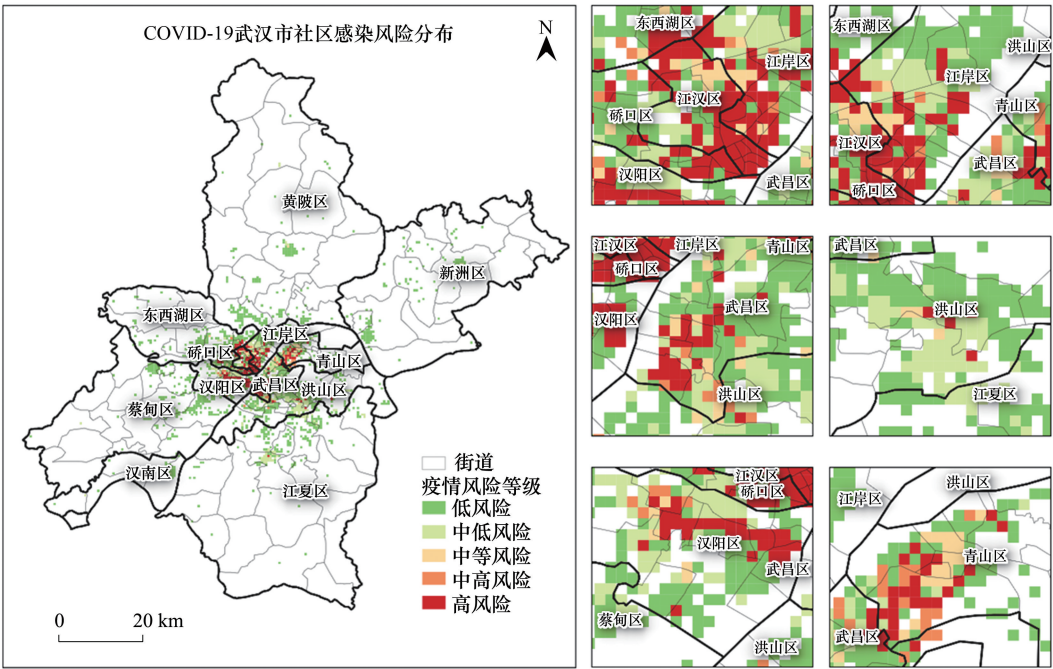


图 4 武汉市域社区感染风险评估

Fig.4 COVID-19 community infection risk assessment in Wuhan

左:市域尺度 右:行政区尺度

<http://www.ecologica.cn>

中餐馆影响权重最高,聚集性聚餐促进了病毒的聚集性传播,分餐而食的外国餐厅影响权重则相对较低。购物服务类中蔬菜市场对病毒传播的影响因素远高于其他购物场所,对病毒防御能力较弱的中老年群众多至蔬菜市场购物,致使病毒更容易通过该人群进行传播。生活服务类通常为密闭环境,空气较难流通,并且该类商店多属于服务业,每日进出人流较多,病毒传播的风险较高。体育休闲类场所多属于封闭环境,且在运动过程中易促进飞沫传播。医疗服务类社区医院(诊所)的感染风险最高,其配套硬件设施的落后以及社区内患者的大量流动导致疫情交叉扩散可能性大。科教文化类场所因环境空气流动性高低,传播风险相应较低或较高。交通设施类场所中封闭环境的轮渡占最高权重,且根据交通工具运载距离远近,风险呈现“高-低-高”的态势。道路附属类场所较易发生接触性感染,公共设施类场所中公共厕所影响因素远高于紧急避难场所,说明及时消毒举措可以降低感染风险。

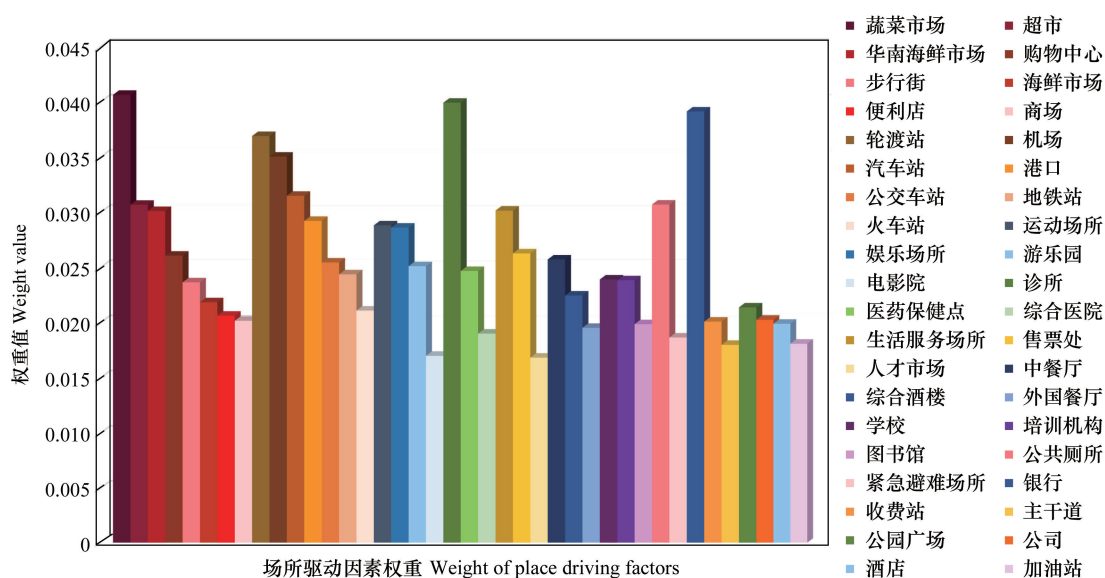


图8 二级场所驱动权重分析

Fig.8 Bar chart for weight analysis of secondary site drivers

2.2 疫情传播空间规律分析及应用

本节主要基于空间聚集性指数总结了疫情传播相关规律,并以此应用于各类院校周边疫情传播风险评估。通过传播规律分析院校周边存在的风险因子,从而对不同院校进行风险传播等级划分。

2.2.1 空间传播规律探究

本研究中以各社区格网单元距离为权重,各社区疫情风险分布数据为属性,得到不同置信区间的 G_i^* 值(图9)。同时,本研究利用标准差椭圆方法体现疫情传播分布的空间特征。标准差椭圆可描述点集相对平均中心的离散程度,反映点的分布特征以及分布的各向异性。图9黄色实线为包含68%数据量、尺寸为1个标准差的椭圆数据,该椭圆长半轴较短,且蔓延主方向与长江流向呈现正交状态,涉及数据主要包含高风险聚集,说明高风险区域分布主要沿长江流向向东西两侧蔓延,且短半轴较长,数据离散程度较大,有向周边蔓延的趋势。根据上章场所因素分析,轮渡、地铁等密闭交通方式为疾病传播影响较大的驱动因子,故而需及时清洁公共交通工具、做好个人防护工作、必要时可采取限制交通等管控措施。橙色实线为包含98%数据量、尺寸为2个标准差的椭圆数据,该椭圆长半轴较长,且向东西方向延伸,能够较清晰地看出东西向的疾病蔓延方向,和一倍标准差椭圆相比,该椭圆包含大部分低风险分布社区,可以较明确地体现疾病可能的发展方向。该图表明东西湖区、洪山区、青山区、蔡甸区、江夏区存在更大疾病蔓延的风险,故而应对上述地区采取更具针对性的防控措施,如以街道为单位,通过核酸试剂检测全面排查患者,及时接收诊治管理疑似病例、确诊病例及

无症状感染者。

通过 G_i^* 统计量分析疾病风险的聚集程度,藉由标准差椭圆分析疾病风险的离散方向,凭借核密度分析探究疫情传播的空间分布情况(图 10),总结新冠病毒肺炎疫情在武汉社区层面传播的规律如下:

(1) 从传播源头分析,武汉市疫情分布呈现“一个中心点,两个分中心”向外辐射的形态,以江岸区华南海鲜市场附近的社区为中心点、半径 10km 的缓冲区,以及以机场为中心点,半径 3km 的缓冲区均为本次疫情高风险区。建议对该处加强管控措施,实时“内防扩散、外防输出、严格管控”。

(2) 从传播方向分析,疫情传播风险严重区大多沿河岸分布,且向东西方向延伸,能够较清晰地看出东西向的疾病蔓延发展方向。故建议对东西湖区、洪山区、青山区、蔡甸区、江夏区等存在更大疾病蔓延风险的地区采取更具针对性的防控措施,如以街道为单位,通过核酸试剂检测全面排查患者,及时接收诊治管理疑似病例、确诊病例及无症状感染者。

(3) 从传播手段分析,购物服务与交通设施两大因素对疫情传播的影响力最为重要。其中,蔬菜市场、超市、购物中心为购物服务中影响因子权重最高的三大场所,而轮渡站、机场和汽车站是交通设施的场所大类中影响权重最高的三类场所。

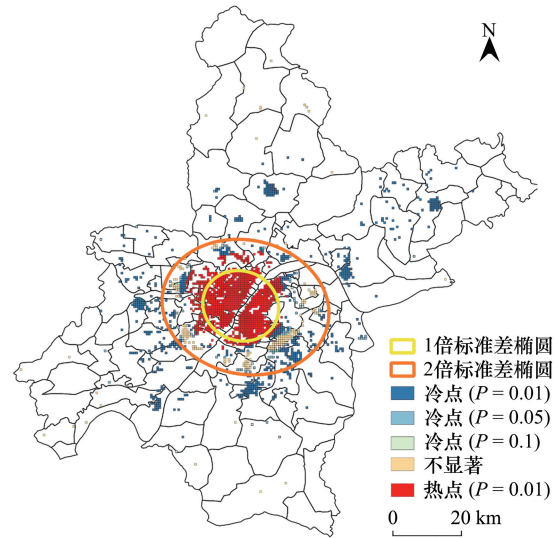


图 9 疫情风险聚集程度指数分析
Fig.9 Analysis of epidemic risk concentration index

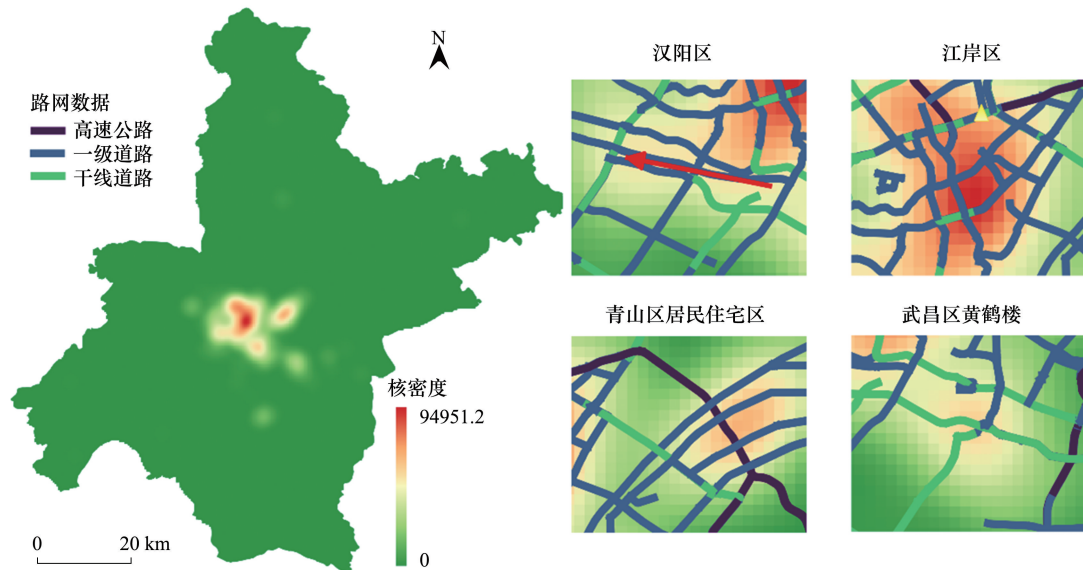


图 10 核密度分析结果
Fig.10 Results of nuclear density analysis(A: Hanyang District
A:汉阳区 箭头所指为琴台大道; B:江岸区 黄色三角为华南海鲜市场; C:青山区居民住宅区 D:武昌区黄鹤楼

2.2.2 传播规律的应用

通过传播规律分析各级各类院校周边存在的疫情传播风险因子,从而对不同院校进行风险传播等级划

分,同时根据疫情传播规律对疫情未结束期间学校开展复学决定提出相应建议。

分别对小学、中学、高等院校进行空间变量点的缓冲区分析,与之前预测的风险区域进行区域统计,求得武汉市所有小学、中学、高等院校空间变量点的风险等级分布。表 4 估了武汉市全域范围内所有学校的传播风险。具体而言,绝大部分小学、中学的疫情传播风险为低风险。这主要是因为中小学处于封闭性管理,人员流动性较低。虽然超过半数高等院校的疫情传播风险为低或中低风险,但其高风险占比以及平均风险等级均偏高,疫情潜在爆发风险不容乐观。这主要是因为高等院校中学生相对自由,学生与校外人群接触多,因此交叉感染可能性大。从图 11 可以看出,武汉市整体学校周边疫情风险等级较低,风险高的区域主要集中在中心城区,并呈现出向外的辐射递减特征,因此复学前期应重点关注中心区域各大高校的消毒防护、人员限流以及肺炎检测工作。图 12 表明,学校周边风险等级高的各类院校主要集中在江汉区、武昌区。中心区人口密集,在复学准备期间可能会由于学生的聚餐、外出购物等集体性行为而产生较高的疫情传播风险,因此应重点关注这些区域的商超、菜市场、中小型作坊以及公共厕所等生活服务设施,对不合格场所及时进行严格整顿。同时,在复学准备期间,应密切关注该区域上学通勤交通工具的消毒工作,并适当增加区域车次,避免上学高峰时期的拥挤现象。

表 4 武汉市全域学校疫情传播风险评估表

Table 4 COVID-19 infection risk evaluation around schools in Wuhan city

	低风险 Low risk level	中低风险 Medium to low risk level	中等风险 Medium risk level	中高风险 Medium to high risk level	高风险 High risk level
小学 Primary school	81.56%	7.96%	5.05%	2.78%	2.65%
中学 Middle school	74.83%	11.70%	5.08%	3.75%	4.64%
高等院校 University	35.71%	23.81%	19.05%	13.10%	8.33%

3 讨论和建议

如今,新型冠状病毒肺炎疫情已经在全世界范围内传播^[5]。全社会的人们必须要尽最大努力,采取一切方法,用尽一切手段以提高公共健康能力和政府行政执行能力,以减缓甚至逼停新型冠状病毒肺炎的传播^[7]。与此同时,政府为避免经济大衰退,势必要采取措施维护经济发展^[61]。2月21日,习近平总书记主持中共中央政治局会议,指出要以实行分区分级精准防控为抓手,统筹疫情防控与经济社会秩序恢复^[62]。武汉市企事业单位的复工复产即将提上日程,本文的研究意义就在于对疫情分布空间属性的深化认知,特别是对传统研究中有关传染病分布研究建模过程中忽视空间特征以及微观环境的不足做出了补充。

本研究主要有以下几个方面的研究发现:首先,武汉中心区域感染风险最高并呈现出向外围递减的趋势;其次,感染风险排名前五的一级场所类型分别为购物服务、交通设施、体育休闲、医疗服务、生活服务以及公共设施;第三,小学、中学的疫情传播风险较低,而高等院校传播风险较高;最后,模型确定了社区尺度下的疫情风险程度,预测购物场所与交通场所是疫情传播风险最

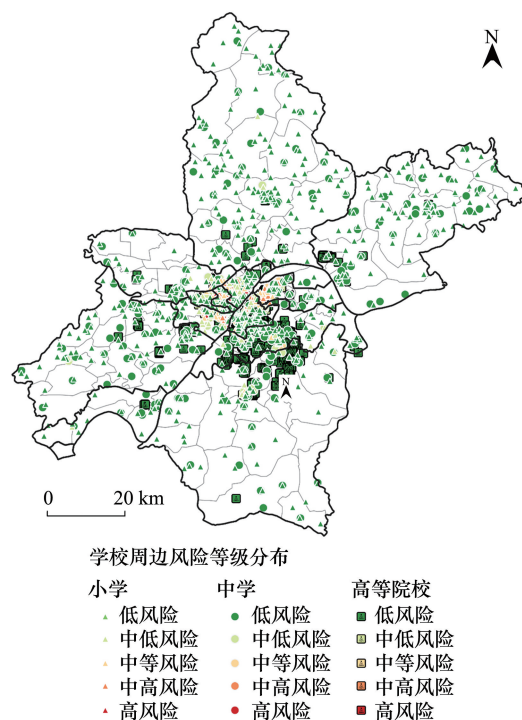


图 11 武汉市全域学校空间分布和周边风险等级

Fig. 11 COVID-19 infection risk spatial distribution around schools in Wuhan city

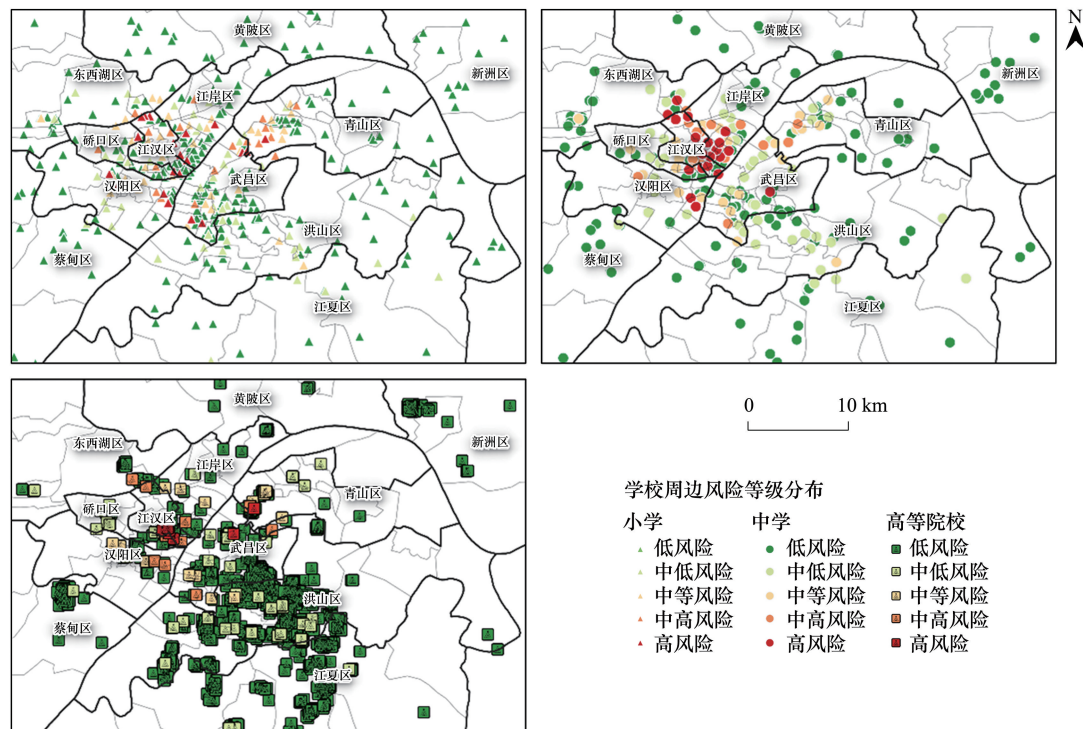


图 12 武汉市中心城区学校周边风险等级分布

Fig.12 COVID-19 infection risk evaluation around schools in downtown

高的驱动因子。针对本研究成果,从政策层面上可提出进行城市治理的相关建议如下:

(1) 加强对公共场所的卫生治理,完善垃圾分类制度,对感染高风险区域加强防疫力度。科学划分疫情等级,重点关注市中心区域,及时更新各地疫情状况,做到分区分级精准施策。对城中村、老旧社区、老年人口较多的社区加强卫生服务和监管,减少感染风险,防控易感人群的暴露和聚集导致疫情扩散的风险。提升社区医院的硬件条件,推广分级治疗。对公共场所进行卫生治理,保证区域中厕所的定期消毒,加强卫生督查工作。增加公共交通消毒频率和力度,保护司乘人员的身体健康,建议建立健全动态的船舶卫生风险评估机制。小型餐馆导致的疫情扩散不可忽视,应增强对小型餐馆的卫生检查管理力度。

(2) 关注学校周边区域卫生环境,错峰开学,限制高风险区域学生外出和跨区流动。建议根据该区域风险评估情况,制定错时错峰开学方案,做好教学衔接。做好各级各类学校周边区域的卫生督查工作,完善垃圾分类制度,严格措施,加大校园管理力度。对于自由度相对较高的高校学生,校内管理人员应做好安全监管工作,实施网格化管理,对发热和密切接触疫情严重区域的学生及时治疗,并在校内定点区域进行隔离。同时也应做好对各级各类学校学生的卫生教育工作。

(3) 合理规划城市功能,加强城市周边区域卫星城居住、医疗和商业服务设施配套工作。适当考虑增加密集居住区的商业和医疗服务配套设施,尤其是城郊大型社区应建立更多便利店以分散大型超市和菜市场承担的商业职能,并可通过菜蔬货品快递上门的方式,避免人群聚集产生疫情感染风险。推动城市服务的自动化和线上化,可有效避免公共服务导致的人群间的病毒的接触传播。在未来的社区服务规划上,根据人口密度、人口年龄、经济水平和教育结构等指标,按需科学配置商业、医疗服务设施和服务人员,强调“以人为本”,合理配置社区各类服务资源。

本文的创新点在于运用地理信息技术和大数据,在社区和场所两个不同尺度上对武汉市市域疫情传播风险进行评估,且以可视化方法简洁明了地分析疫情传播的驱动因素,最后依据研究结论提出防控建议。希望本文的研究方法及结论可以为未来结合地理信息技术的传染病研究提供借鉴与参考。

在设计和构建风险评估模型的过程中,本研究仍有不足的地方。首先,收集的社区疫情数据时间跨度不够,仅考虑到空间范围,摘取时刻点上的相应数据进行分析,使得结果可能存在一定偏差。在未来研究中将考虑将时序感染数据加入模型进行分析。其次,武汉主城区外的新区多以自然村形式分布,且大多为手机使用频率较低的老年人口和儿童,故而人口密度数据存在一定偏差,相关地区的评估代表性可能不足。此外,本文未将境外输入病例作为考虑因素,而现阶段国内大部分确诊病例的主要来源均为境外输入病例^[31,63],在今后的研究中会将此作为新的特征变量予以考虑。

4 结论

本文以曾经的疫情高发地武汉作为研究区域,耦合社区疫情传染数据、兴趣点数据、LBS 人口密度数据以及路网数据,结合随机森林算法评估社区和场所两个尺度上的武汉疫情传播风险,模型测试精度达 0.85,且与真值一致性较好。并且对学校周边地区疫情风险进行分析,由此提出错峰上学、重点区域重点防控的建议。以特征向量权重为依据分析疫情传播的驱动因子,发现购物服务和交通设施存在聚集性疫情爆发的可能性,说明模型对引起疫情传播的因子具有一定的识别能力,体现地理信息技术在传染病防控方面的适用性。

参考文献 (References):

- [1] 李昀, 杨静, 靳森, 任婧寰, 涂文校, 向妮娟, 张彦平. 2019 年 11 月中国大陆需关注的突发公共卫生事件风险评估. 疾病监测, 2019, 34 (11): 960-962.
- [2] U.S. Department of Health & Human Services. Coronavirus disease 2019 (COVID-19) situation summary. (2020-04-19) [2020-05-01]. <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/summary.html>.
- [3] Buchanan R L, Gorris L G M, Hayman M M, Jackson T C, Whiting R C. A review of *Listeria monocytogenes*: an update on outbreaks, virulence, dose-response, ecology, and risk assessments. Food Control, 2017, 75: 1-13.
- [4] Kim I, Lee J, Lee J, Shin E, Chu C, Lee S K. KCDC risk assessments on the initial phase of the COVID-19 outbreak in Korea. Osong Public Health and Research Perspectives, 2020, 11(2): 67-73.
- [5] Wang C, Horby P W, Hayden F G, Gao G F. A novel coronavirus outbreak of global health concern. The Lancet, 2020, 395(10223): 470-473.
- [6] 国家卫生健康委办公厅. 新型冠状病毒肺炎防控方案(第五版). (2020-02-21) [2020-03-07]. <http://www.nhc.gov.cn/jkj/s3577/202002/a5d6f7b8c48c451c87dba14889b30147/files/3514cb996ae24e2faf65953b4ecd0df4.pdf>.
- [7] World Health Organization. Coronavirus disease 2019 (COVID-19): situation report-100. (2020-04-29) [2020-04-30]. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200429-sitrep-100-covid-19.pdf?sfvrsn=bbfb3d1_6.
- [8] Wu J T, Leung K, Leung G M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet, 2020, 395(10225): 689-697.
- [9] Boldog P, Tekeli T, Vizi Z, Dénes A, Bartha F A, Röst G. Risk assessment of novel coronavirus COVID-19 outbreaks outside China. Journal of Clinical Medicine, 2020, 9(2): 571.
- [10] Yang X F, Xu T, Jia P, Xia H, Guo L, Ye K. Transportation, germs, culture: a dynamic graph model of 2019-nCoV spread. Preprints, 2020: 2020020063, doi: 10.20944/preprints202002.0063.v1.
- [11] 巩玥, 史志祥, 陈菁, 张宇辉, 赵光慧, 王刚, 伍洲, 岳卫平, 王琳. 冠状病毒的研究现状. 中国生物工程杂志, 2020, 40(S1): 1-20.
- [12] Gao J J, Tian Z X, Yang X. Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. BioScience Trends, 2020, 14(1): 72-73.
- [13] Zhou F, Yu T, Du R H, Fan G H, Liu Y, Liu Z B, Xiang J, Wang Y M, Song B, Gu X Y, Guan L L, Wei Y, Li H, Wu X D, Xu J Y, Tu S J, Zhang Y, Chen H, Cao B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet, 2020, 395(10229): 1054-1062.
- [14] Guan W J, Ni Z Y, Hu Y, Liang W H, Ou C Q, He J X, Liu L, Shan H, Lei C L, Hui D S C, Du B, Li L J, Zeng G, Yuen K Y, Chen R C, Tang C L, Wang T, Chen P Y, Xiang J, Li S Y, Wang J L, Liang Z J, Peng Y X, Wei L, Liu Y, Hu Y H, Peng P, Wang J M, Liu J Y, Chen Z, Li G, Zheng Z J, Qiu S Q, Luo J, Ye C J, Zhu S Y, Zhong N S, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. The New England Journal of Medicine, 2020, 382(18): 1708-1720.
- [15] Chan J F W, Yuan S F, Kok K H, To K K W, Chu H, Yang J, Xing F F, Liu J L, Yip C C Y, Poon R W S, Tsoi H W, Lo S K F, Chan K H, Poon V K M, Chan W M, Ip J D, Cai J P, Cheng V C C, Yuen P K Y. A familial cluster of pneumonia associated with the 2019 novel coronavirus

- indicating person-to-person transmission: a study of a family cluster. *The Lancet*, 2020, 395(10223): 514-523.
- [16] Bai Y, Yao L S, Wei T, Tian F, Jin D Y, Chen L J, Wang M Y. Presumed asymptomatic carrier transmission of COVID-19. *JAMA*, 2020, 323(14): 1406-1407.
- [17] Narasimhan R, Bhuvaneshwari T. Big data-a brief study. *International Journal of Scientific & Engineering Research*, 2014, 5(9): 350-353.
- [18] 刘双芳, 金华, 靳利梅, 刘志军, 俞晓英, 徐兴华, 刘敏科, 朱飞飞, 张俊鹏. 不同中医证型高血压病患者缺血性心血管病 10 年发病风险评估. *中国中医药信息杂志*, 2020, 27(6): 28-33.
- [19] Travers D, Barnett C, Ising A, Waller A. Timeliness of emergency department diagnoses for syndromic surveillance. *AMIA... Annual Symposium Proceedings. AMIA Symposium*, 2006, 2006: 769-773.
- [20] Hanslik T, Boelle P Y, Flahault A. The control chart: an epidemiological tool for public health monitoring. *Public Health*, 2001, 115(4): 277-281.
- [21] 杨维中, 邢慧娴, 王汉章, 兰亚佳, 孙乔, 胡世雄, 吕伟, 袁政安, 陈裕旭, 董柏青. 七种传染病控制图法预警技术研究. *中华流行病学杂志*, 2004, 25(12): 1039-1041.
- [22] 梅文娟, 刘震, 朱静怡, 杜立. 新冠肺炎疫情极限 IR 实时预测模型. *电子科技大学学报*, 2020, 49(3): 362-368.
- [23] Liao Y L, Xu B, Wang J F, Liu X C. A new method for assessing the risk of infectious disease outbreak. *Scientific Reports*, 2017, 7: 40084.
- [24] Deng T, Huang Y, Yu S C, Gu J, Huang C R, Xiao G X, Hao Y T. Spatial-temporal clusters and risk factors of hand, foot, and mouth disease at the district level in Guangdong province, China. *PLoS One*, 2013, 8(2): e56943.
- [25] Fotheringham A S, Wong D W S. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A: Economy and Space*, 1991, 23(7): 1025-1044.
- [26] Smail L. Uniqueness of the level two bayesian network representing a probability distribution. *International Journal of Mathematics and Mathematical Sciences*, 2011, 2011: 845398.
- [27] Jiang X, Wallstrom G L. A Bayesian network for outbreak detection and prediction//*Proceedings of the 21st National Conference on Artificial Intelligence*. Boston: AAAI Press, 2006: 1155-1160.
- [28] Ribeiro A F, Tengan C, Sato H K, Spinola R, Mascheretti M, França A C C, Port-Carvalho M, Pereira M, de Souza R P, Amaku M, Burattini M N, Coutinho F A B, Lopez L F, Massad E. A public health risk assessment for yellow fever vaccination: a model exemplified by an outbreak in the state of São Paulo, Brazil. *Memórias do Instituto Oswaldo Cruz*, 2015, 110(2): 230-234.
- [29] Li L T, Mills W L, Gutierrez A M, Herman L I, Berger D H, Naik A D. A patient-centered early warning system to prevent readmission after colorectal surgery: a national consensus using the Delphi method. *Journal of the American College of Surgeons*, 2013, 216(2): 210-216.e6.
- [30] Wagner M M, Tsui F C, Espino J U, Dato V M, Sitting D F, Caruana R A, McGinnis L F, Deerfield D W, Druzdzel M J, Fridsma D B. The emerging science of very early detection of disease outbreaks. *Journal of Public Health Management and Practice*, 2001, 7(6): 51-59.
- [31] 国家卫生健康委办公厅. 新型冠状病毒肺炎防控方案(第六版). (2020-03-07) [2020-04-23]. <http://www.nhc.gov.cn/jkj/s3577/202003/4856d5b0458141fa9f376853224d41d7/files/4132bf035bc242478a6eaf157eb0d979.pdf>.
- [32] Lu D S, Weng Q H, Li G Y. Residential population estimation using a remote sensing derived impervious surface approach. *International Journal of Remote Sensing*, 2006, 27(16): 3553-3570.
- [33] Yao Y, Liu X P, Li X, Zhang J B, Liang Z T, Mai K, Zhang Y T. Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *International Journal of Geographical Information Science*, 2017, 31(6): 1220-1244.
- [34] 于翔. 基于城市公交刷卡数据和兴趣点的城市功能区识别研究——以北京市为例[D]. 杭州: 浙江大学, 2014.
- [35] Howe J. The rise of crowdsourcing. *Wired Magazine*, 2006, 14(6): 1-5.
- [36] Heipke C. Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2010, 65(6): 550-557.
- [37] Bakillah M, Liang S, Mobasher A, Arsanjani J J, Zipf A. Fine-resolution population mapping using OpenStreetMap points-of-interest. *International Journal of Geographical Information Science*, 2014, 28(9): 1940-1963.
- [38] Goodchild F M. Commentary: whither VGI? *GeoJournal*, 2008, 72(3/4): 239-244.
- [39] Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and POIs//*Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Beijing, China: Association for Computing Machinery, 2012: 186-194.
- [40] Hu Q W, Wang M, Li Q Q. Urban hotspot and commercial area exploration with check-in data. *Acta Geodaetica et Cartographica Sinica*, 2014, 43(3): 314-321.
- [41] Yao Y, Li X, Liu X P, Liu P H, Liang Z T, Zhang J B, Mai K. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science*, 2017, 31(4): 825-848.
- [42] Yao Y, Liu P H, Hong Y, Liang Z T, Wang R Y, Guan Q F, Chen J M. Fine-scale intra- and inter-city commercial store site recommendations

- using knowledge transfer. *Transactions in GIS*, 2019, 23(5): 1029-1047.
- [43] Helbich M, Yao Y, Liu Y, Zhang J B, Liu P H, Wang R Y. Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in Beijing, China. *Environment International*, 2019, 126: 107-117.
- [44] Wang R Y, Chen H S, Liu Y, Lu Y, Yao Y. Neighborhood social reciprocity and mental health among older adults in China: the mediating effects of physical activity, social interaction, and volunteering. *BMC Public Health*, 2019, 19(1): 1036.
- [45] Wang R Y, Liu Y, Xue D S, Yao Y, Liu P H, Helbich M. Cross-sectional associations between long-term exposure to particulate matter and depression in China: the mediating effects of sunlight, physical activity, and neighborly reciprocity. *Journal of Affective Disorders*, 2019, 249: 8-14.
- [46] 武汉市卫生健康委员会. 武汉市新冠肺炎疫情动态(2020年4月29日). (2020-04-30) [2020-04-30]. http://wjw.wuhan.gov.cn/ztl_28/fk/tzgg/202004/t20200430_1202883.shtml.
- [47] Li J, Long Y, Dang A R. Live-Work-Play Centers of Chinese cities: identification and temporal evolution with emerging data. *Computers, Environment and Urban Systems*, 2018, 71: 58-66.
- [48] Wang R Y, Liu Y, Lu Y, Zhang J B, Liu P H, Yao Y, Grekousis G. Perceptions of built environment and health outcomes for older Chinese in Beijing: a big data approach with street view images and deep learning technique. *Computers, Environment and Urban Systems*, 2019, 78: 101386.
- [49] Wang R Y, Liu Y, Lu Y, Yuan Y, Zhang J B, Liu P H, Yao Y. The linkage between the perception of neighbourhood and physical activity in Guangzhou, China: using street view imagery with deep learning techniques. *International Journal of Health Geographics*, 2019, 18(1): 18.
- [50] 冯明翔, 方志祥, 路雄博, 谢泽丰, 熊盛武, 郑猛, 黄守倩. 交通分析区尺度上的 COVID-19 时空扩散推估方法——以武汉市为例. *武汉大学学报: 信息科学版*, 2020, 45(5): 651-657, 681-681.
- [51] Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. *Mathematical Biosciences*, 2013, 244(2): 125-134.
- [52] Liu Y, Liu X, Gao S, Gong L, Kang C G, Zhi Y, Chi G H, Shi L. Social sensing: a new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 2015, 105(3): 512-530.
- [53] 杨帆, 米红. 一种基于网格的空间聚类方法在区域划分中的应用. *测绘科学*, 2007, 32(S1): 66-69, 5-5.
- [54] 骆剑承, 周成虎, 梁怡, 张讲社, 黄叶芳. 多尺度空间单元区域划分方法. *地理学报*, 2002, 57(2): 167-173.
- [55] Breiman L. Random forests. *Machine Learning*, 2001, 45(1): 5-32.
- [56] Lin Y, Jeon Y. Random forests and adaptive nearest neighbors. *Journal of the American Statistical Association*, 2006, 101(474): 578-590.
- [57] McHugh M L. Interrater reliability: the kappa statistic. *Biochemia Medica*, 2012, 22(3): 276-282.
- [58] Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, 20(1): 37-46.
- [59] Tharwat A. Classification assessment methods. *Applied Computing and Informatics*, 2020, 17(1): 168-192. doi: 10.1016/j.aci.2018.08.003.
- [60] Landis J R, Koch G G. The measurement of observer agreement for categorical data. *Biometrics*, 1977, 33(1): 159-174.
- [61] Anderson R M, Heesterbeek H, Klinkenberg D, Hollingsworth T D. How will country-based mitigation measures influence the course of the COVID-19 epidemic?. *The Lancet*, 2020, 395(10228): 931-934.
- [62] 新华网. 中共中央政治局常务委员会召开会议 分析新冠肺炎疫情形势研究加强防控工作 中共中央总书记习近平主持会议. (2020-02-12) [2020-04-30]. http://www.xinhuanet.com/politics/2020-02/12/c_1125565831.htm.
- [63] 国家卫生健康委员会. 截至4月30日24时新型冠状病毒肺炎疫情最新情况. (2020-05-01) [2020-05-01]. http://www.xinhuanet.com/politics/2020-05/01/c_1125931400.htm.