

基于多源地理数据的城市突发公共卫生事件风险评估研究

熊励,王思媛

(上海大学 管理学院,上海 200444)

摘要:风险评估是突发公共卫生事件应急管理的关键环节,基于地理数据开展风险评估可有效提高精度.通过采集多源地理数据,结合随机森林算法和地理探测器、核密度分析、空间自相关等空间统计分析提出城市突发公共卫生事件风险评估方法,并通过实证分析验证模型可行性.结果表明:采用随机森林算法构建的风险评估模型表现良好;餐饮美食、公司企业和交通设施等场所是影响疫情的主要因素,疫情流行具有因子交互性,其中餐饮美食与其他因子的交互作用最强;空间传播上位于城市中心区域的社区风险等级较高并呈现向外围逐渐减弱的趋势,同时有明显的高值或低值聚集.

关键词:多源数据;机器学习;空间统计分析;公共卫生;风险评估

中图分类号:TP311.13;R181.8

文献标志码:A **文章编号:**1000-2367(2024)05-0091-10

公共安全是城市发展和治理的重要组成部分^[1],其中突发公共卫生事件因其扩散趋势快、危害程度大、复杂程度高,成为应急管理中最难应对的重大问题.随着信息社会的不断发展,传统定性或定量研究方法已无法适应高度复杂的突发公共事件风险治理需求,利用新兴信息技术实现风险的精准认知,提高风险防范能力对于维护城市公共卫生安全、实现高效能治理、推进应急管理体系和能力现代化具有重要意义^[2].

近些年来诸多学者围绕突发公共卫生事件风险防控进行了相关研究,一方面从医学及病毒学角度探讨其致病机理^[3]、药物防治^[4]、疫苗研发^[5]等;另一方面是探寻非药物防控,如病例的早期识别和隔离、限定地理范围内的区域封锁等^[6].在手机信令数据、遥感数据和传感器网络数据等海量地理空间数据不断产生的时代,空间大数据与传统流行病学调查相结合成为突发公共卫生事件监测预警的新方法.当前通过地理数据分析对突发公共卫生事件风险监测与预警的研究大都从宏观层面切入,对其空间扩散特征与影响因素进行探究.而社区作为一定地域内发生各种社会关系和社会活动,并具有归属感的人群所组成的一个相对独立的社会实体,既是城市结构和社会治理的最小单元,也是突发公共卫生事件防控的第一道防线,应引起足够重视.但由于细粒度数据较难获取,当前较少有聚焦城市内部空间要素与突发公共卫生事件风险评估等方面的研究.

为解决上述问题,本文提出了一种融合机器学习和空间统计分析的突发公共卫生事件风险评估方法,包括城市社区风险等级研判及可视化,突发公共卫生事件空间分异驱动因素探测,突发公共卫生事件空间格局演化分析,着力为重大突发公共卫生事件风险防控提供决策依据,提升城市应急处置能力.

收稿日期:2023-07-10;**修回日期:**2023-08-31.

基金项目:国家社科基金重大项目(21ZDA105).

作者简介(通信作者):熊励(1966—),女,湖北武汉人,上海大学教授,博士,研究方向为信息管理与信息系统、公共卫生与应急管理,E-mail:xiONGLI0616@126.com.

引用本文:熊励,王思媛.基于多源地理数据的城市突发公共卫生事件风险评估研究[J].河南师范大学学报(自然科学版),2024,52(5):91-100.(Xiong Li,Wang Siyuan.Research on risk assessment of urban public health emergencies based on multi-source geographic data[J].Journal of Henan Normal University(Natural Science Edition),2024,52(5):91-100.DOI:10.16366/j.cnki.1000-2367.2023.07.10.0005.)

1 国内外研究现状

1.1 城市突发公共卫生事件风险评估

风险评估是重大突发事件中风险研判和应急管理的关键环节,是应急决策的根本依据^[7].国外发达国家的相关研究起步较早,当前在重大突发事件的预警及风险研究、模型及技术研究等方面的成果较为丰富^[8].GOODE等^[9]为提升突发公共卫生事件早期评估能力,开发了在线快速风险和影响评估工具以促进远程工作专家之间的合作.CHEN等^[10]采用长短期记忆网络方法对加拿大新冠病毒感染疫情进行时间序列预测.国内学者陈秋玲等^[11]以突发事件视角基于突变理论构建指标体系以测定上海市的城市安全度,是较早从风险管理视角展开的研究.刘清华等^[12]借助系统动力学理论建立一类肺结核和新冠疫情共发感染的传播动力学模型,讨论了新冠疫情对肺结核控制的可能影响.陈云松等^[13]基于网络大数据、统计年鉴、疫情公开信息构建疫情风险指数,对中国城市疫情风险指数进行测量和可视化.目前国内突发公共卫生事件应急管理风险评估研究仍处于发展状态,统计分析、可视化分析、时空轨迹分析等大数据治理技术可应用到风险识别、风险分析、风险评价、风险决策等治理流程,基于数据驱动的风险评估体系及预警体系建设是未来的新机遇.

1.2 地理信息技术在城市公共卫生领域的应用

随着全球卫星导航定位技术、地理信息系统和通信网络技术的发展,产生了越来越多可被感知、存储和分析的时空数据.时空分析是流行病学研究的重要组成部分,通过分析疾病的三间分布研判可能进一步传播或者已经存在的隐性传播、潜在传播的风险和强度,从而有效控制疫情传播.Dijst等认为心血管疾病的发生与长期暴露于空气污染有关^[14],通过构建荷兰空气质量数据集为全国不同空间层次的个人污染暴露评估提供了基础.ZHANG等^[15]在SIR(susceptible infected recovered, SIR)模型的基础上利用有关地点、住宅和办公室密度及公共交通数据建立了综合时空风险评估模型,计算流行病在城市环境中的传播.新冠病毒感染疫情暴发后,更多学者将地理信息系统应用于疫情传播与扩散分析、区域规划与应急管理、数据与决策支撑等研究中,从国家、省域或城市群视角等宏观层面和市域层面进行深入研究^[16-19],其中多采用空间自相关等方法分析疫情时空演化过程^[20],采用地理加权回归等方法研究疫情发展影响因素^[21],采用地理探测器模型度量空间分层异质性^[22].

综上所述,当前国内外主要通过半定量和定量分析方法对突发公共卫生事件风险进行分析预测.突发公共卫生事件传播耦合了时间过程和空间规律,人地关系是地理学研究的核心,通过地理学视角对时空格局、区域差异、空间治理、影响因素和预测预警等过程进行研究,既可以分析自然环境与疾病传播的内在联系,也可以从人文视角为应急管理提供决策支持.但当前研究大多在国家或者省域层面,城市社区作为突发公共卫生事件防控的底层,在控制外部输入和内部传播方面发挥着不可替代的作用,守牢社区防疫第一线也是实施精准防控的关键.

2 突发公共卫生事件风险评估模型

2.1 基于自然断点法的突发公共卫生事件风险等级划分

突发公共卫生事件风险等级预测可视为机器学习中的多分类问题,需要事先定义风险等级作为模型目标值.本文采用自然断点法划分风险等级,以单位感染人数表示感染风险.自然断点法是减少组内的平均离散方差、增加组间的平均离散方差的自然聚类方法,完全依据数据的分布规律划分,避免了人为因素.首先计算病例总数平均值的偏差平方和(sum of squared deviations for array mean, SDAM),即:

$$SDAM = \sum_{i=1}^n (X_i - \bar{X})^2, \quad (1)$$

式中, X_i 为S市各社区病例数; \bar{X} 为病例数的均值.随后针对依据病例数分组的每种分组方式,在其分出的每一组均值上计算组内偏差平方和并累加所有组结果,即类别平均数的偏差平方和(sum of squared deviations for class means, SDCM),并找到最小的分组方式:

$$SDCM = \sum_{j=1}^{G_i} (G_{ij} - \bar{G}_i)^2, \quad (2)$$

式中, G_{ij} 为第 i 个风险等级的第 j 个病例数值; \bar{G}_i 为第 i 个风险等级中病例数的均值. 最后通过计算方差拟合优度 (goodness of variance fit, GVF) 对每一种风险等级划分方案进行评价, 在取值范围 $[0, 1]$ 中越高越好, 即:

$$GVF = (SDAM - SCDM) / SDAM. \quad (3)$$

当依据某一区域内病例数将风险等级分为低、中、高 3 级时, 自然断点法会首先计算全部病例数平均值的偏差平方和, 其次将全部病例数划为 3 个风险等级并迭代每个范围组合, 计算各个等级均值的平方偏差平方和, 最后采用方差拟合优度对不同分组方式进行评价, 选取 $[0, 1]$ 内值最高的分组作为最佳分组. 将数据集采用自然断点法打上风险等级标签后划分为训练集和测试集, 随后使用随机森林算法对测试集进行风险等级预测.

2.2 基于随机森林算法的突发公共卫生事件风险研判

突发公共卫生事件风险研判是通过机器学习算法学习训练集数据特征对风险等级的判断, 从而根据测试集数据特征做出相应的风险等级预测, 即有监督机器学习中的分类问题. 其中随机森林 (random forest, RF) 是一个高度灵活的算法, 具有较高准确率、避免过拟合、能有效分析高维度数据等特点, 其在以决策树为基学习器构建 Bagging 集成的基础上进一步在决策树的训练过程中引入了随机特征选择. 因此采用随机森林算法构建突发公共卫生事件风险评估模型, 预测突发公共卫生风险等级.

2.3 基于空间统计的突发公共卫生事件空间格局演化分析

风险等级研判是对各个社区相对独立的进行风险评估. 地理学第一定律指出地理事物或属性在空间分布上互为相关, 存在集聚、随机、规则分布^[23]. 因此, 进一步采用空间统计分析对突发公共卫生事件空间传播规律进行探究.

2.3.1 空间自相关

全局空间自相关即分布于不同空间位置的地理事物某一属性值存在统计相关性. 本文以全局莫兰指数 (global moran's I) 衡量空间自相关, 其定义式如下:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (x_i - \bar{x})^2}, \quad (4)$$

式中, n 为 S 市社区个数; W_{ij} 为社区单元 i 和 j 之间空间权重, 用来度量空间单元之间的邻接关系; x_i, x_j 分别为第 i, j 个社区病例数的观测值; \bar{x} 为 S 市病例数量的均值. 全局莫兰指数的值域为 $[-1, 1]$, 正值表示空间正相关性, 且数值越大空间相关性越明显. 负值表示空间负相关性, 数值越小表示空间差异越大, 若取值为零, 表明空间呈随机性. 全局莫兰指数可以表明突发公共卫生事件的发生在空间上呈聚集模式、离散模式或是随机模式, 但无法得知在哪里出现, 因此进一步采用热点分析 (getis-ord G_i^*) 探索局部空间聚类分布特征, 找出各类集聚的空间分布区域, 排除全局空间自相关分析的局部不稳定性.

2.3.2 热点分析

热点分析目的在于识别出具有统计显著性聚类的区域, 即分析病例数量在空间上是否存在随机过程创造的统计显著性空间集聚 (热点) 或分散 (冷点) 现象. 其构建在统计推断中常用的零假设检验思想之上, 表示如下:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j - \bar{X} \sum_{j=1}^n w_{ij}}{\sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - \bar{x}^2} \sqrt{\frac{\lfloor n \sum_{j=1}^n w_{ij}^2 - (\sum_{j=1}^n w_{ij})^2 \rfloor}{n-1}}}, \quad (5)$$

式中, w_{ij} 为社区单元 i 和 j 之间的空间权重, x_j 为社区 j 的病例数, n 为 S 市社区总数.

2.3.3 地理探测器

为探索新冠病毒感染疫情病例的空间分布与城市建成环境要素分布的关系,以及城市建成环境要素对病例分布的驱动力,引入可以探测空间分异性以及揭示其背后驱动力的地理探测器模型^[22].空间分异性是指层内方差小于层间方差的地理现象,本文使用因子探测探究单个因子对突发公共卫生事件病例分布的解释力度,用 q 值度量,表达式为:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2}, \quad (6)$$

式中, h 为因变量或自变量的分类或分区数; N_h 为第 h 个子区域单元数量; σ_h^2 和 σ^2 分别是第 h 个子区域和该市全域确诊病例的方差; N 为全市社区数量. q 的值域为 $[0,1]$,值越大则因变量的空间分异性越明显;若空间分异性是由自变量形成的,则 q 值越大表示该因子对因变量的解释力度越强.

进一步使用地理探测器中交互作用探测说明因子间交互作用的强度,即首先计算两种因子 X_1 和 X_2 对 Y 的 q 值 $q(X_1)$ 和 $q(X_2)$,其次计算相交时的 q 值 $q(X_1 \cap X_2)$,最后通过比较 $q(X_1)$ 、 $q(X_2)$ 与 $q(X_1 \cap X_2)$ 来评估因子 X_1 和 X_2 共同作用时是否会对各社区病例情况的解释力产生影响.

2.3.4 核密度估计

核密度估计是检测局部密度变化、探索空间热点的有效技术,利用核密度估计能有效识别各类风险因子的空间分布规律,进一步探索其与病例数的关系.该方法是根据待估计点和每一个样本点之间的距离,分别计算核密度值,距离越近得到的核密度值越大,最后将所有核密度值进行加权求和,从而得到该估计点在样本分布中的概率密度值.每个估计点的风险因子核密度估计值 $\hat{f}(x, y)$ 为:

$$\hat{f}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x-x_i}{h}, \frac{y-y_i}{h}\right), \quad (7)$$

式中: n 为该类因子的总数量, h 为搜索半径, $K(x, y)$ 为高斯分布核密度函数, (x, y) 为待估计点空间位置, (x_i, y_i) 为该类因子中第 i 个因子的空间位置.

3 实证分析

本文以S市2022年3至5月的新冠病毒感染疫情为例进行分析.2022年3月开始,S市新冠病毒感染疫情发病区域与感染人群呈现复杂多变的态势,通过风险研判,科学防控,至2022年6月1日起全市基本恢复正常状态,因此以2022年3至5月为研究周期具有一定的代表性.

3.1 数据描述

3.1.1 S市新冠病毒感染疫情病例数据

收集整理S市卫生健康委员会网站通报的2022年3至5月病例居住地信息.通常相关研究会将风险定为高、中、低3个等级^[24-25],同时自然断点法是一种在地理学研究中常见且经典的分类方法^[26-27].因此通过高德开放平台API接口进行正向地理编码,根据式(1)~(3)采用自然断点法将感染人数分为3个级别:0~9人为低风险,10~35人为中风险,35(不含)人以上为高风险.本文以单位感染人数表示感染风险,基于ArcGIS 10.7绘制病例点空间分布情况(见附录图S1).

3.1.2 兴趣点数据

兴趣点数据因具有较好的客观性与时效性,在城市精细化研究中得到广泛应用,也是影响居民公共健康的重要因素之一^[28].使用高德开放平台API接口获取兴趣点数据,根据高德地图兴趣点数据行业分类标准,选取与人群活动密切相关的12类要素共331596条文本数据,每条数据具有名称、大类、中类、地址、经度、纬度、行政区等7个属性(见表1).

3.1.3 人口密度数据

人口密度可以用来衡量社交距离,其会对突发公共卫生事件的传播率造成影响^[29].本文采用WorldPop网站上中国2020年人口分布栅格数据(100 m分辨率)作为人口密度数据来源(<https://www.worldpop.org/>).

表 1 兴趣点数据分类

Tab. 1 Point-of-interest data classification

编号	变量名称	POI 类型	编号	变量名称	POI 类型
X ₁	餐饮美食	中国菜、外国菜、咖啡	X ₇	科教文化	幼儿园、小学、中学、高等院校、职业技术学校
X ₂	公司企业	公司	X ₈	旅游景点	公园、广场
X ₃	购物消费	市场、超市、百货商场、商业街、购物中心	X ₉	汽车相关	加油站
X ₄	交通设施	停车场、公交车站、地铁站、港口码头	X ₁₀	商务住宅	住宅区、商住两用楼宇
X ₅	金融机构	银行	X ₁₁	生活服务	公共厕所、事务所、人才市场
X ₆	酒店住宿	宾馆酒店、经济型连锁酒店	X ₁₂	医疗保健	综合医院、专科医院、诊所、医药保健销售

3.2 评价指标

为了评价基于随机森林算法构建的突发公共卫生事件风险评估模型的有效性,选取准确率(accuracy, A)、精确率(precision, P)、召回率(recall, R)、 $F1$ 分数($f1$ -score)等作为模型评价指标,具体公式如下:

$$A = \frac{TP + TN}{TP + FP + FN + TN}, \tag{8}$$

$$P = \frac{TP}{TP + FP}, \tag{9}$$

$$R = \frac{TP}{TP + FN}, \tag{10}$$

$$F1 = \frac{2}{\left(\frac{1}{P} + \frac{1}{R}\right)} = \frac{2PR}{P + R}, \tag{11}$$

式中, TP 表示正确分类为正例的正样本数, TN 表示正确分类为负例的负样本数, FP 表示错误分类为正例的负样本数, FN 表示错误分类为负例的正样本数。

本实验的研究环境部署在 Windows 10 操作系统,利用 Python 3.7 和 ArcGIS 10.7 进行实验。

3.3 实证结果分析

3.3.1 突发公共卫生事件风险等级研判结果分析

对数据集进行特征工程后将 80% 划分为训练集,20% 划分为测试集,对风险评估模型进行参数调优并根据式(8)~(11)的评价指标对模型结果进行评价(见表 2)。

表 2 基于随机森林算法构建的风险评估模型评价指标结果

Tab. 2 Evaluation metrics for a risk assessment model based on random forest algorithm

指标	准确率	精确率	召回率	$F1$ 分数
值	0.859 356	0.802 045	0.859 356	0.804 189

由表 2 可见随机森林算法在各项评价指标上的表现效果良好,准确率、精确率、召回率和 $F1$ 值分别为 0.859 356、0.802 045、0.859 356 和 0.804 189,模型拟合程度较高,可以较为准确地预测出社区突发公共卫生事件风险等级。

基于该模型对 S 市市域范围内 16 780 个社区进行风险评估,绘制全市风险等级分布图(见附录图 S2)和中心城区典型社区空间分布图(见附录图 S3)。

风险等级由深浅程度不同的红色进行区分,风险等级越高则突发公共卫生事件发生概率越大,由图 S2 可见本轮疫情在 S 市绝大多数地区都有病例,其中风险等级较高的社区主要集中在中心城区,这些行政区域内公共交通便利,大型商圈较多,同时人口较为密集,流动性高,这也与突发公共卫生事件的传播与人口密度、餐饮、商圈、交通设施的空间相关性较大相对应。

选取图 S3 所示的 3 个不同风险等级的社区进行深入分析,社区 1 风险等级较高,经过分析发现该社区外来人口多、人口密度高、高龄老人多,相对复杂的情况导致病毒极易在空间留存并在高密度的人群中传播扩散,在疫情初期超四分之一的居民感染,后期通过外部驰援、区、街道、社区居民的共同努力成功阻止疫情

继续蔓延.社区 2 风险等级为中风险,该社区所处地理位置公共交通便利,人员的频繁流动增加了风险,但居民相对年轻化,自身防控意识较强的同时能够更好地利用新媒体平台,通过创立微信公众号将居民关心的相关信息及时、透明发布,信息的公开和快速传递对于疫情的控制起到了积极作用.社区 3 风险等级较低,因该社区在组织管理体系建设方面通过调动高素质高学历居民积极性,建立志愿者队伍,明确责任分工,提高疫情防控效率,从而保持疫情传播处于低风险.

3.3.2 突发公共卫生事件地理空间分布分析

为进一步了解疫情的聚类特征,根据式(4)、(5)对 S 市病例进行空间自相关分析和热点分析.首先对 S 市病例进行全局空间自相关分析,全局莫兰指数为 0.134 0($P < 0.001$),表明病例分布存在空间正相关,但关联程度不是十分明显,这也与疫情总体呈现多点暴发相对应.为进一步揭示局部地区疫情的分布特征,采用热点分析得到统计意义上疫情冷点和热点分布格局(见图 1),进一步了解本轮疫情病例点在空间分布上是否明显存在统计显著性空间集聚(热点)或分散(冷点)现象.

由图 1 可见,本轮疫情的热点区域主要在中心城区,即市中心和东北部疫情较为严重,主要由于这些区域是全市人口最密集、人口流动性最大的地区.地理位置相对较偏的行政区域是疫情冷点区,总体与病例总量的空间分布具有一致性.

为探测突发公共卫生事件风险等级的空间分异性以及各类场所对该分异性的解释情况,利用地理探测器对 S 市各社区病例数进行单因子探测和多因子交互作用分析(见表 3).

表 3 S 市新冠病毒感染疫情社区病例数驱动因素探测结果

Tab. 3 The detection outcomes regarding the factors influencing the case count within the community during the COVID-19 outbreak in S city

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
X_1	0.059 0											
X_2	0.064 8	0.018 6										
X_3	0.064 4*	0.018 9	0.000 9									
X_4	0.063 3	0.044 7*	0.018 1	0.017 8								
X_5	0.059 1	0.019 1*	0.001 3	0.018 4*	0.000 3							
X_6	0.059 3*	0.018 8*	0.001 0	0.018 0*	0.000 4	0						
X_7	0.060 3	0.026 0*	0.006 0	0.025 5*	0.005 6	0.005 3	0.005 2					
X_8	0.059 2	0.019 9*	0.001 8	0.019 2*	0.001 3	0.001 0	0.006 3*	0.000 9				
X_9	0.061 3*	0.019 1	0.001 9	0.018 3	0.001 2	0.000 9	0.006 0	0.001 8	0.000 8			
X_{10}	0.066 2*	0.018 7	0.001 5*	0.017 9	0.000 7	0.000 4	0.005 4	0.001 3	0.001 3	0.000 4		
X_{11}	0.059 1	0.020 7*	0.001 9	0.020 1*	0.001 4	0.001 1	0.006 6*	0.002 0	0.001 9	0.001 3	0.001 0	
X_{12}	0.059 4*	0.019 1*	0.001 1	0.018 4*	0.000 5	0.000 2	0.005 4*	0.001 1	0.001 0	0.000 5	0.001 2	0.000 1

注:除 X_6 和 X_{12} 外,其他因子均通过 0.05 置信度检验,* 表示非线性增强,其余为双因子增强.

从单因子作用看,各因子对 S 市新冠病毒感染疫情社区病例数的影响程度由大到小依次为:餐饮美食

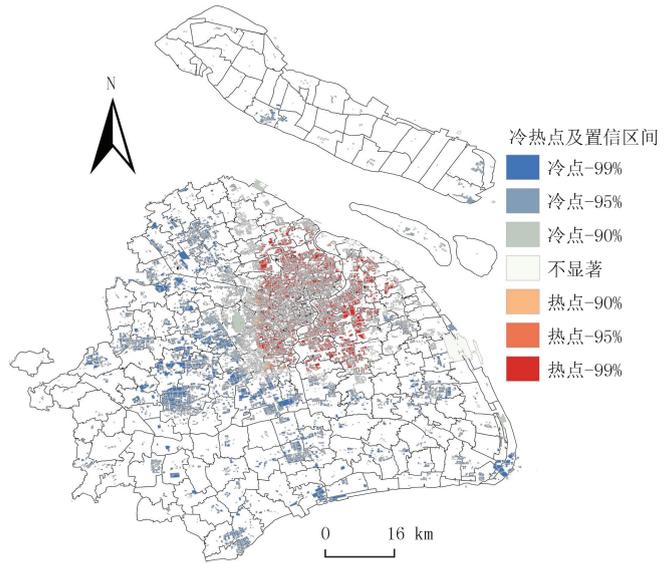


图 1 S 市突发公共卫生事件热点分析

Fig.1 Hotspot analysis of public health emergencies in S city

(X_1), 公司企业(X_2), 交通设施(X_4), 科教文化(X_7), 生活服务(X_{11}), 购物消费(X_3), 旅游景点(X_8), 汽车相关(X_9), 商务住宅(X_{10}), 金融机构(X_5). 餐饮美食、公司企业和交通设施场所对社区疫情病例的空间分异有较大影响, 这些场所都具有人流量大、人群密集、空间密闭等特点.

从因子交互作用看, S 市新冠病毒感染疫情的流行, 存在双因子增强和非线性增强两种类型, 所有双因子的交互作用都大于单因子的作用. 其中, 餐饮美食场所与公司企业、购物消费、交通设施、科教文化、商务住宅的交互对病例数量有较强的影响作用.

为了更加直观地反映主要场所驱动因子的空间布局, 结合疫情病例数分布, 进一步分析其对疫情风险等级的影响程度. 采用式(7)核密度分析公式得到餐饮美食、公司企业和交通设施 3 类对 S 市新冠病毒感染疫情社区病例数影响较大的场所的空间聚集特征, 为使得结果更加直观, 采用自然断点法对核密度估计结果进行分级, 结果见图 2.

从图 2 中可以看出, 空间分布上 3 类驱动因子呈现出明显的聚集形态, 虽然核密度分布不尽相同, 但核密度高值分布都较为接近, 整体呈现大集中、小分散的空间分布结构, 3 类驱动因子核密度高值大范围分布在市中心, 同时公司企业场所的核密度分布呈现“整体聚集、圈层蔓延”的形态, 这说明中心城区各项功能设施完善, 与图 2 S 市市域范围社区突发公共卫生事件风险评估表现的中心城区风险等级较高相对应. 经过核密度分析可以得到主要驱动因子的聚集分布, 进一步印证其与疫情

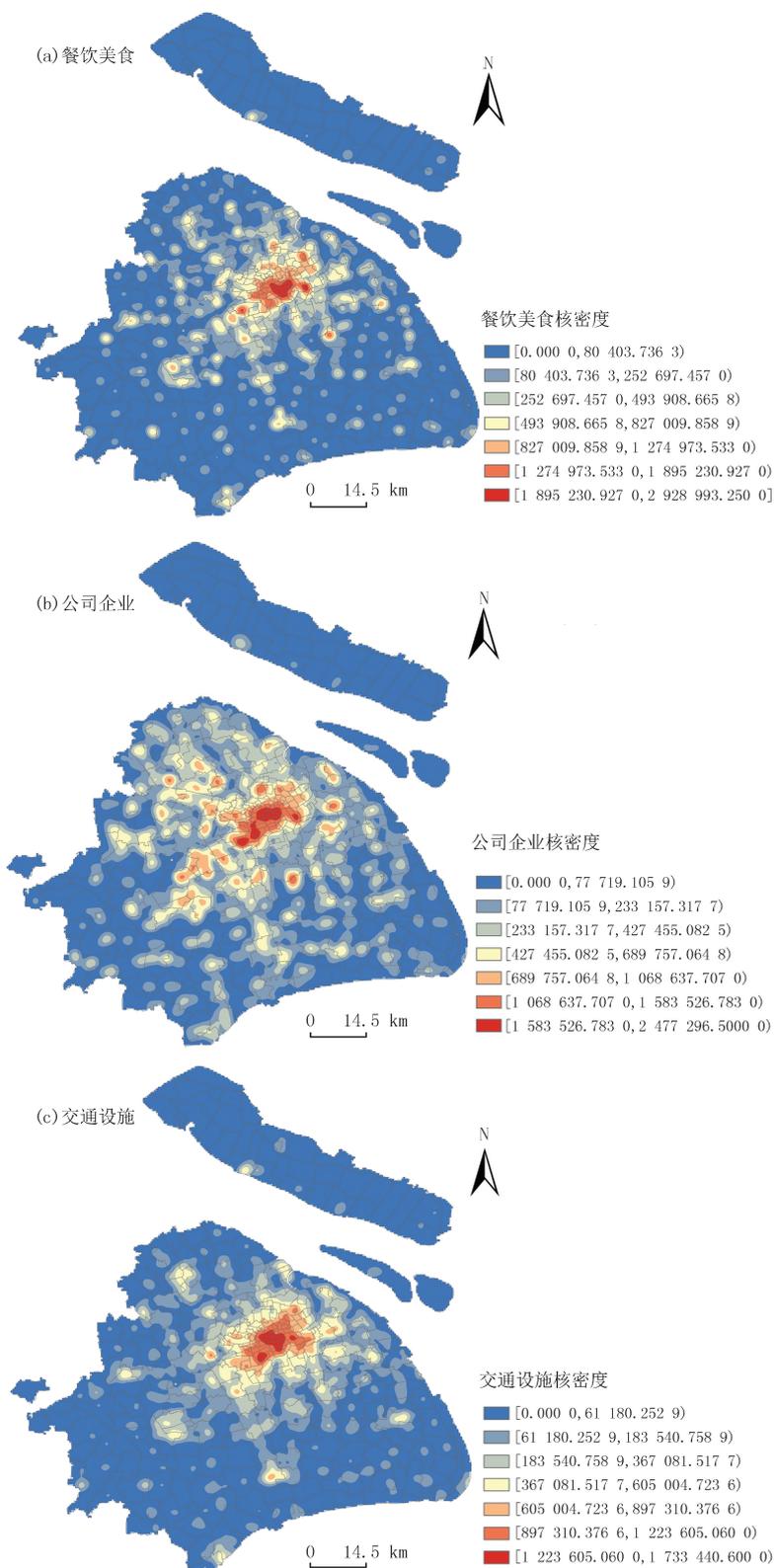


图2 主要驱动因子核密度分布图

Fig.2 Kernel density distribution of the main drivers

风险等级的关联性。

4 结论与建议

本文基于多源地理数据,融合机器学习中的随机森林算法和地理探测器、核密度分析、空间自相关等空间统计分析构建突发公共卫生事件风险评估模型对社区尺度下的城市突发公共卫生事件进行风险评估,并以 S 市新冠病毒感染疫情为例进行实证分析,研究结果如下:

第一,随机森林模型在突发公共卫生事件风险预测方面表现良好,整体上具有较高的性能。

第二,餐饮美食、公司企业、交通设施等场所对疫情的空间分异有较大影响,且双因子的交互作用大于单因子的作用,尤其是餐饮美食场所与公司企业、购物消费、交通设施、科教文化、商务住宅的交互,对病例数量有较强的影响作用。

第三,疫情分布具有空间集聚性,疫情热点区集中在中心城区,郊区大多为疫情冷点区。

在上述结论基础上,为着力提升社区突发公共卫生事件风险防控能力,一是注重开展基于数据驱动的精细尺度下社区突发公共卫生事件风险评估,通过充分有效的数字化、智能化实现基层治理现代化。在突发公共卫生事件发生前期,融合汇聚社区人员、基础设施建设、应急医疗资源配置等实时风险数据,基于事件发展规律快速模拟、评估事态预警等级,实现高效精准的基层风险点预警、应急决策和资源协调及响应处置;事中实现信息互联互通、各方协调联动、资源精准调配;事后及时分析总结,通过构建城市数据资源体系等政策措施,努力打造科学化、精细化的城市“数治”新范式。二是合理规划社区生活圈空间,通过优化城市空间结构设计和调整相关防控政策降低突发公共卫生事件潜在风险,应重点关注餐饮美食、公司企业、交通设施等对社区突发公共卫生事件空间分异影响较大的场所,加强场所内的环境及人流量监测,长期应注重构建韧性城市空间规划,包括完善相关的标准规范体系、平疫结合的空间规划策略、多尺度的设施配套标准、设计指引和管控规程等,增强应对突发公共卫生事件的韧性。三是加强跨区域应急协调联动和城郊地区公共卫生体系建设,增强协同处置能力。中心城区尤其是超大城市中心城区人口密度大、流动性高,各项功能设施完善,是突发公共卫生事件发生的热点区域,通过建立有效的区域联动机制,可以缓解超大城市压力,促进区域协调发展。同时加强中心城区周边地区基础设施建设规划,合理配置各类服务资源,充分利用城郊空间缓解韧性城市建设的资源及用地压力,更好实现中心城区和城郊地区协调联动发展。

本文运用机器学习算法和空间统计分析对突发公共卫生事件外部影响因素进行研究,其创新之处在于从数据科学和地理学的角度出发,基于多源地理数据,融合空间视角实现精细尺度下社区的风险评估,进一步精准划分风险等级,以便及时有效采取防控措施,降低突发公共卫生事件潜在传播风险,增强城市韧性。

时空行为地理学的分析方法如何应用于城市应急管理仍处于探索阶段,本文实证部分采集的主要为突发公共卫生事件病例数据和建成环境空间数据,未来考虑进一步扩大数据源,采集更多突发公共卫生事件相关数据,在一定程度上提高因变量的被解释程度,以更加精准研判风险。

附录见电子版(DOI:10.16366/j.cnki.1000-2367.2023.07.10.0005)。

参 考 文 献

- [1] 熊竞.新冠肺炎疫情影响下的城市规模治理:基于我国 348 个地级以上样本单元的分析[J].城市发展研究,2021,28(4):49-56.
XIONG J.Urban scale governance under the influence of 2019-ncov epidemic:based on the analysis of 348 sample units in China[J].Urban Development Studies,2021,28(4):49-56.
- [2] 张宁,唐嘉仪.大数据在风险治理中的应用场景:基于“人-技术”视角的研究述评[J].电子政务,2021(4):79-89.
ZHANG N,TANG J Y.Application scenarios of big data in risk management;a review based on the perspective of "people-technology"[J].E-Government,2021(4):79-89.
- [3] TANG N,LI D J,WANG X,et al.Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia[J].Journal of Thrombosis and Haemostasis:JTH,2020,18(4):844-847.
- [4] HOFFMANN M,KLEINE-WEBER H,SCHROEDER S,et al.SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor[J].Cell,2020,181(2):271-280.

- [5] 王秀芹.谁在参与新冠疫苗研发“竞赛”:从参与主体特征到研发模式[J].中国科技论坛,2021(2):1-8.
WANG X Q. Who are competing for the covid-19 vaccines: characteristics of leading-participants and R&D mode[J]. Forum on Science and Technology in China, 2021(2):1-8.
- [6] LAI S J, RUKTANONCHAI N W, ZHOU L C, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China[J]. Nature, 2020, 585(7825):410-413.
- [7] 翟羽佳,许佳,赵明.突发公共卫生事件风险研判信息语义聚合模式研究[J].图书与情报,2021(5):21-30.
ZHAI Y J, XU J, ZHAO Y. Research on the semantic aggregation model of risk research and judgment of public health emergencies[J]. Library & Information, 2021(5):21-30.
- [8] 刘雅妹,梁宇,周红磊,等.基于事理图谱的重大突发事件动态演变研究[J].图书情报工作,2022,66(10):143-151.
LIU Y S, LUAN Y, ZHOU H L, et al. Research on the dynamic evolution of major emergencies based on event knowledge graph[J]. Library and Information Service, 2022, 66(10):143-151.
- [9] GOODE E J, THOMAS E, LANDEG O, et al. Development of a rapid risk and impact assessment tool to enhance response to environmental emergencies in the early stages of a disaster: a tool developed by the European multiple environmental threats emergency NETWORK(EMETNET) project[J]. International Journal of Disaster Risk Science, 2021, 12(4):528-539.
- [10] CHEN Y Y, LI N Z, LOURENÇO J, et al. Measuring the effects of COVID-19-related disruption on dengue transmission in Southeast Asia and Latin America: a statistical modelling study[J]. The Lancet Infectious Diseases, 2022, 22(5):657-667.
- [11] 陈秋玲,张青,肖璐.基于突变模型的突发事件视野下城市安全评估[J].管理学报,2010,7(6):891-895.
CHEN Q L, ZHANG Q, XIAO L. The indexes, model and method for urban safety assessment under emergent situation based on the catastrophe model[J]. Chinese Journal of Management, 2010, 7(6):891-895.
- [12] 刘清华,曹慧,李海燕.肺结核和 COVID-19 共发感染的建模与分析[J].河南师范大学学报(自然科学版),2023,51(1):54-62.
LIU Q H, CAO H, LI H Y. Modeling and analysis of tuberculosis and co-infection with COVID-19[J]. Journal of Henan Normal University(Natural Science Edition), 2023, 51(1):54-62.
- [13] 陈云松,陈步伟,句国栋,等.突发重大疫情下城市系统风险量化评估方法[J].西安交通大学学报(社会科学版),2020,40(4):33-41.
CHEN Y S, CHEN B W, JU G D, et al. A quantitative assessment method for urban systemic risk in the case of a major epidemic outbreak [J]. Journal of Xi'an Jiaotong University(Social Sciences), 2020, 40(4):33-41.
- [14] 王玉圳.面向空间规划的时间地理学研究进展:以 Martin Dijkstra 研究为例[J].城市发展研究,2021,28(11):1-7.
WANG Y Z. Time geography and its application in spatial planning: a case study of the research of martin dijest[J]. Urban Development Studies, 2021, 28(11):1-7.
- [15] ZHANG N, HUANG H, DUARTE M, et al. Dynamic population flow based risk analysis of infectious disease propagation in a metropolis [J]. Environment International, 2016, 94:369-379.
- [16] KANGA S, MERAJ G, SUDHANSHU, et al. Analyzing the Risk to COVID-19 Infection using Remote Sensing and GIS[J]. Risk Analysis: an Official Publication of the Society for Risk Analysis, 2021, 41(5):801-813.
- [17] 曹永旺,刘樱,周春山.城市群视角下新冠肺炎疫情的时空扩散特征与影响因素[J].地域研究与开发,2021,40(3):1-7.
CAO Y W, LIU Y, ZHOU C S. Spatiotemporal diffusion characteristics and influencing factors of COVID-19 epidemic from the perspective of urban agglomeration[J]. Areal Research and Development, 2021, 40(3):1-7.
- [18] 王姣娥,杜德林,魏冶,等.新冠肺炎疫情的空间扩散过程与模式研究[J].地理研究,2020,39(7):1450-1462.
WANG J E, DU D L, WEI Y, et al. The development of COVID-19 in China: spatial diffusion and geographical pattern[J]. Geographical Research, 2020, 39(7):1450-1462.
- [19] 李文辉,邱晓晴.基于社会网络的城市疫情空间扩散及治理研究:以深圳新冠肺炎病例为依据[J].世界地理研究,2022,31(3):637-648.
LI W H, QIU X Q. Research on spatial spread and control of urban epidemic by social network: based on the prevention and control of COVID-19 in Shenzhen[J]. World Regional Studies, 2022, 31(3):637-648.
- [20] RIPLEY BRIAN D. "book-review" Spatial Processes: models and Applications[J]. Journal of the American Statistical Association, 1984, 79(385):238.
- [21] STEWART FOTHERINGHAM A, CHARLTON M, BRUNSDON C. The geography of parameter space: an investigation of spatial non-stationarity[J]. International Journal of Geographical Information Systems, 1996, 10(5):605-627.
- [22] 王劲峰,徐成东.地理探测器:原理与展望[J].地理学报,2017,72(1):116-134.
WANG J F, XU C D. Geodetector: principle and prospective[J]. Acta Geographica Sinica, 2017, 72(1):116-134.
- [23] TOBLER W R. A computer movie simulating urban growth in the Detroit Region[J]. Economic Geography, 1970, 46:234.
- [24] 陈扬,赵漫,欧阳桃花,等.应对重大突发公共卫生事件的敏捷性机制研究:四川省新冠疫情案例[J].公共管理与政策评论,2022,11(1):73-87.
CHEN Y, ZHAO M, OUYANG T H, et al. The agility mechanism of magnitude outburst public health incident: a case study of COVID-19 pandemic in Sichuan[J]. Public Administration and Policy Review, 2022, 11(1):73-87.

- [25] 孟兆敏.社区常态化疫情防控机制研究[J].西北人口,2021,42(2):62-70.
MENG Z M.Study on prevention and control mechanism of community normalized epidemic disease[J].Northwest Population Journal, 2021,42(2):62-70.
- [26] 朱伶俐,任红艳,丁凤,等.陕西省 HFERS 疫情时空分异特征及影响要素研究[J].地球信息科学学报,2020,22(5):1142-1152.
ZHU L L,REN H Y,DING F,et al.Spatiotemporal variations and influencing factors of hemorrhagic fever with renal syndrome in Shaanxi Province[J].Journal of Geo-Information Science,2020,22(5):1142-1152.
- [27] 赵凯旭,张帅兵,黄晓军,等.新冠疫情管控期间西安市人口分布演变及影响因素探测:基于多源时空大数据视角[J].人口与发展,2022,28(1):140-150.
ZHAO K X,ZHANG S B,HUANG X J,et al.Research on evolution of population distribution and influencing factors in Xi'an during the COVID-19 epidemic control period:based on a perspective of multi-source spatio-temporal big data[J].Population and Development, 2022,28(1):140-150.
- [28] WANG R Y,LIU Y,LU Y,et al.Perceptions of built environment and health outcomes for older Chinese in Beijing;a big data approach with street view images and deep learning technique[J].Computers,Environment and Urban Systems,2019,78:101386.
- [29] COBB J S,SEALE M A.Examining the effect of social distancing on the compound growth rate of COVID-19 at the county level(United States)using statistical analyses and a random forest machine learning model[J].Public Health,2020,185:27-29.

Research on risk assessment of urban public health emergencies based on multi-source geographic data

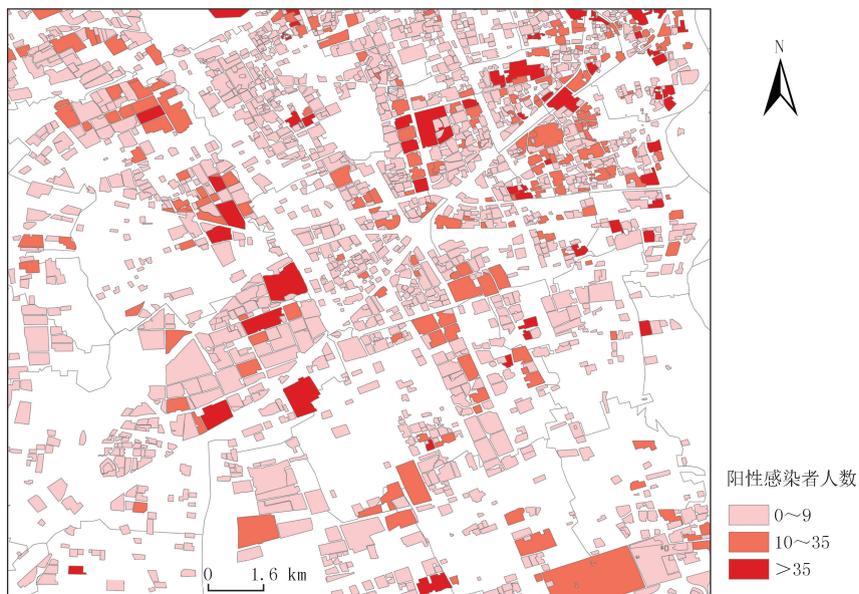
Xiong Li, Wang Siyuan

(School of Management, Shanghai University, Shanghai 200444, China)

Abstract: Risk assessment is a crucial component in the emergency management of sudden public health incidents, and geospatial data-based risk assessment can effectively enhance accuracy. By collecting multi-source geospatial data and employing spatial statistical analyses such as the random forest algorithm, geographic detector, kernel density analysis, and spatial autocorrelation, a methodology for risk assessment of sudden public health incidents in urban areas is proposed. The feasibility of the model is validated through empirical analysis. The results indicate that the risk assessment model is constructed well by using the random forest algorithm performs. Places such as dining establishments, corporate enterprises, and transportation facilities are identified as the major factors influencing spatial differentiation of the epidemic. The epidemic spread exhibits factor interactions, with dining establishments having the strongest interaction with other factors. In terms of spatial dissemination, communities located in the central areas of the city exhibit higher risk levels, showing a gradual reduction towards the periphery, accompanied by distinct patterns of high or low-value aggregation.

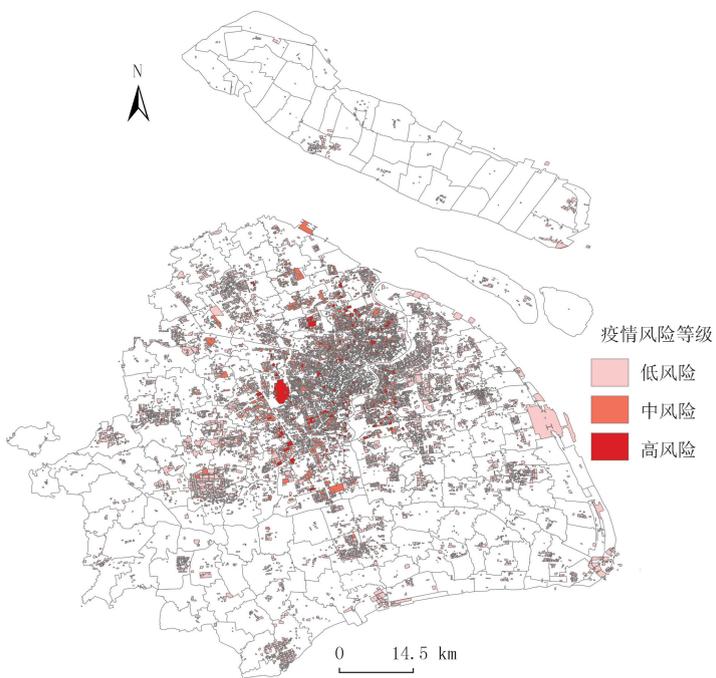
Keywords: multi-source data; machine learning; spatial data analysis; public health; risk assessment

[责任编辑 陈留院 赵晓华]



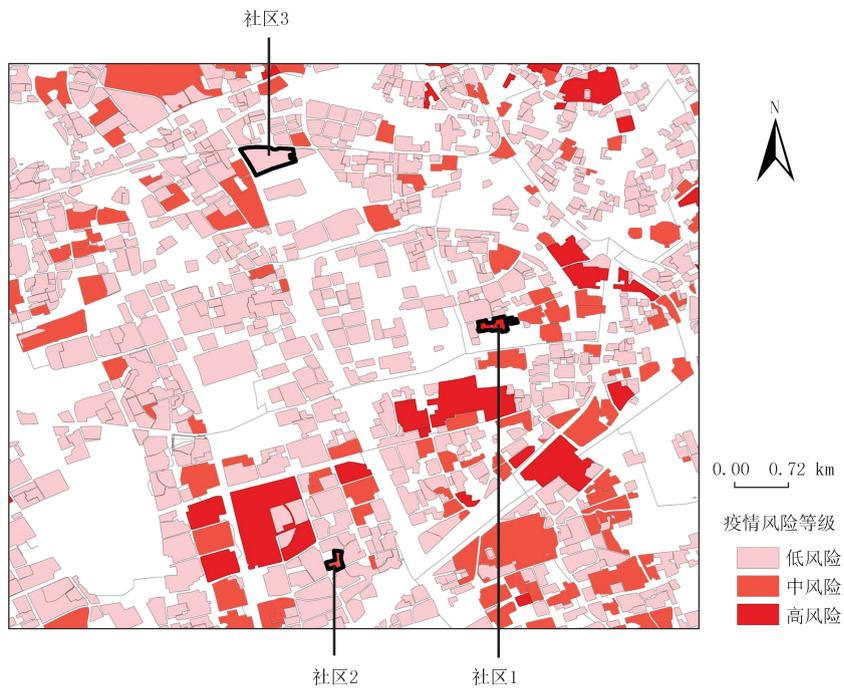
图S1 S市新冠病毒感染疫情局部病例数量及分布情况

Fig.S1 Number and distribution of local COVID-19 cases in S city



图S2 S市市域范围社区突发公共卫生事件风险评估

Fig.S2 Risk assessment of community-based public health emergencies within the municipal area of S city



图S3 S市中心城区典型社区空间分布

Fig.S3 Spatial distribution of typical communities in the central urban area of S city