

吴立志, 陈振南, 张 鹏. 基于随机森林算法的城市火灾风险评估研究[J]. 灾害学, 2021, 36(4): 54–60. [WU Lizhi, CHEN Zhennan and ZHANG Peng. Research on Urban Fire Risk Evaluation Based on The Random Forest Algorithm [J]. Journal of Catastrophology, 2021, 36(4): 54–60. doi: 10.3969/j.issn.1000-811X.2021.04.010.]

# 基于随机森林算法的城市火灾风险评估研究\*

吴立志<sup>1</sup>, 陈振南<sup>2</sup>, 张 鹏<sup>3</sup>

(1. 中国人民警察大学, 河北 廊坊 065000; 2. 中国人民警察大学 研究生院, 河北 廊坊 065000;  
3. 中国人民警察大学 智慧警务学院, 河北 廊坊 065000)

**摘 要:** 为定量分析城市火灾风险, 引入随机森林算法。首先设计基于随机森林算法的递归特征消除方法优化初选的14个自变量指标, 剔除无关或冗余变量; 其次基于随机森林算法的平均准确度降低方法计算特征重要度, 将其归一化结果作为指标的权重; 最后结合指标数据, 使用线性加权法评估城市火灾风险。以济南市区为例的实证研究显示: 经过特征选择操作后发现只有10个指标时模型误差最小,  $MSE$  为  $7.43 \times 10^{-4}$ ; 用于计算指标权重的随机森林模型与实际火灾密度拟合的决定系数  $R^2 > 0.85$ , 精度较高, 可用于对指标的客观赋权; 火灾风险评估结果不仅显示历史火灾较多的区域是高风险区, 也可得到火灾发生较少区域的风险值。

**关键词:** 城市火灾; 风险评估; 随机森林算法; 特征选择; 指标权重

**中图分类号:** X928.7; X4; X915.5; TU998.1 **文献标志码:** A **文章编号:** 1000-811X(2021)04-0054-07  
doi: 10.3969/j.issn.1000-811X.2021.04.010

城市火灾风险评估(urban fire risk evaluation)是指对城市用地范围内的建筑、场所、设施等发生火灾的危险性和危害性进行的综合评价<sup>[1]</sup>, 对优化消防力量配置以提高灭火救援效率、指导消防检查以提高城市消防安全水平具有重要作用。

英国较早开展城市火灾风险评估, 将城市划分为A、B、C、D共4个风险等级, 规定了不同等级所对应的消防站响应时间等内容<sup>[2]</sup>; Entec公司通过分析火灾发生数等与环境特征、人群分布、建筑布局等因素之间的关系, 开发了“风险评估工具箱”, 帮助消防站对不同的火灾场景作出有效反应<sup>[3]</sup>。美国发布了“风险、危害和经济价值评估(Risk, Hazard and Value Evaluation)”系统, 从区域建筑和人口数据入手评估火灾风险<sup>[4-5]</sup>; 日本借鉴美国经验, 结合本国特点提出了城市等级法<sup>[5-6]</sup>。

国内从研究对象上说, 城市火灾风险评估研究可分为两大块。一是评估指标体系的构建, 包含指标的选取与体系结构。如苏拓<sup>[7]</sup>在火灾风险评估指标体系中首次加入消防安全形势、社会火灾防控两大一级指标; 刘纪华等<sup>[8]</sup>基于Bow-tie方法构建三级风险评估指标体系; 刘晗等<sup>[9]</sup>引入在美国军事防灾领域广泛应用的Fine Kinney方法, 将风险评估要素分为相似度、暴露度和灾害后果准则, 基于Dematel方法和层次分析法建立火灾风险评估框架。二是火灾风险评估的方法。如张

刚<sup>[10]</sup>从规划学的角度, 将城市划分为不同网格区域, 使用空间分析方法对火灾风险因素打分进而评估各区域相应的火灾风险; 陈国良等<sup>[11]</sup>把火灾起数、伤亡人数、财产损失等纳入评估指标体系, 使用模糊数学理论对城市火灾风险进行评估; 崔冠雄使用灰色理论改进的层次分析法建立了风险评估模型<sup>[12]</sup>。

虽然较多学者开展了城市火灾风险评估研究, 但仍存在一些不足: 一是缺乏开展自变量与因变量之间的相关关系建模, 导致所选指标并不一定是最优集合; 二是指标赋权时有一定主观性。随着人工智能、大数据科学、计算机等技术的发展, 机器学习算法被逐步应用到各个领域。其中, 随机森林算法在森林火灾中应用较多, OLIVEIRA等<sup>[13]</sup>, LIANG等<sup>[14]</sup>都基于随机森林算法开展了森林火灾风险研究, 取得了一定成果。

考虑到随机森林算法具有鲁棒性强、训练速度快、实现简单和能够进行特征选择并计算指标贡献度等特点, 本文引入随机森林算法开展城市火灾风险评估研究。主要工作有: 一是设计基于随机森林算法的递归特征消除方法(Recursive Feature Elimination, RFE)对初选指标进行特征选择, 识别最优指标集合; 二是基于随机森林算法定量分析自变量与因变量指标间的关系, 实现对指标的客观赋权。

\* 收稿日期: 2021-04-28 修回日期: 2021-07-01

基金项目: 河北省自然科学基金资助项目(D2015507046)

第一作者简介: 吴立志(1968-), 男, 汉族, 浙江金华人, 博士, 教授, 主要从事火灾风险评估、灭火与应急救援研究。

E-mail: wulizhi119@sohu.com

通讯作者: 陈振南(1992-), 男, 汉族, 山东淄博人, 硕士研究生, 主要从事火灾风险评估、灭火与应急救援研究。

E-mail: chenzhennan1992@163.com

## 1 随机森林算法原理

随机森林算法最早由 Breiman 提出, 属于集成学习的一种, 主要思想是将若干弱学习器集合组成强学习器进而获得更优的泛化性能<sup>[15]</sup>。随机森林算法由弱学习器决策树组成。

### 1.1 决策树

决策树是一种基于属性对数据进行分类或回归的树形结构。以分类为例, 其目的是通过一系列的属性划分, 最终将输入数据按照一定规则分开。决策树包含根节点、内部节点和叶子结点。根节点包含所有样本数据; 内部结点的样本数据根据分裂属性划分到子节点中; 叶子结点是决策树对样本数据的分类结果<sup>[16]</sup>。

决策树生长的关键在于找到最优划分属性, 简而言之就是通过一定准则能从属性集合中找到最优划分指标, 使得子节点数据尽量属于同一类别, 即数据纯度越高越好。采用“基尼值”(Gini ( $D$ ))表示节点样本数据的纯度<sup>[16]</sup>, 即:

$$\text{Gini}(D) = 1 - \sum_{i=1}^n p_i^2 \quad (1)$$

式中:  $D$  为数据集,  $n$  为样本类别数,  $p_i$  为  $D$  中第  $i$  类样本所占比例。Gini( $D$ )表示从数据集  $D$  中随机抽取两个样本, 其类别标记不一致的概率。Gini( $D$ )越小, 则数据集  $D$  的纯度越高。

采用基尼系数法选择划分属性。该方法选择划分后基尼系数最小的属性作为最优分裂属性。若数据集  $D$  以离散属性  $a$  划分子节点,  $a$  有  $v$  个取值  $\{a^1, a^2, a^3, \dots, a^v\}$ , 则划分后产生共  $v$  个分支, 第  $v$  个分支包括  $D$  中属性  $a$  取  $a^v$  的所有样本, 记为  $D^v$ , 基尼系数定义为:

$$\text{Gini\_index}(D, a) = \sum_{v=1}^v \frac{|D^v|}{|D|} \text{Gini}(D^v) \quad (2)$$

### 1.2 随机森林

单棵决策树模型的分类规则较复杂, 容易陷入过度拟合。为克服决策树的缺陷, 引入随机森林的思想。首先是构建一片森林, 森林由相互独立、互不关联的决策树作为基分类器构成, 每棵决策树使用各自的输入数据生长, 形成自己的结果; 其次为保证决策树的多样性, 使用随机抽取样本和随机属性选择改进决策树的生成方式, 避免过拟合和降低决策规则的复杂性; 最后随机森林的输出按一定规则生成。随机森林算法性能较决策树提高很多, 应用更广泛, 主要是引入两个随机性, 具体生成步骤如下:

(1) 随机抽取样本。随机森林算法的训练集由有放回的自助法随机产生。假设数据集有  $N$  个样本, 建立  $ntree$  棵决策树, 从中抽取一个样本放入训练集后返回原始数据集, 待下次抽取时原始数据集仍然是  $N$  个, 每个样本仍会被等概率的抽中。经过  $N$  次抽取形成第一棵决策树的训练集。重复  $ntree$  次形成所有训练集。

(2) 随机属性选择。单棵决策树在训练时, 每个节点的分裂属性从当前剩余所有  $M$  个属性中遍历得到。而在随机森林算法中, 分裂时首先从该节点  $M$  个属性中随机抽取包含  $mtry$  ( $m < M$ ) 个属性的子集,  $mtry$  被称为随机特征变量; 其次从属性子集中找到最优分裂属性进而划分。每棵树在

生长过程中随机生成的分裂属性子集很难相同, 引入  $mtry$  增强了树之间的独立性, 增加了随机森林算法的随机性。

(3) 构建决策树。完成步骤(1)和(2)后, 基于基尼系数的方法完成决策树的构建。决策树的数量  $ntree$  是随机森林算法另一个重要参数, 树过少则分类或回归误差大, 性能较差; 树过多则模型较复杂, 增加了时间成本, 泛化能力差。决策树的数量可通过试探法确定。

(4) 结果输出。随机森林算法结果输出最简单且常用的方法是线性集成。对所有决策树的结果进行投票, 每棵决策树占一票, 按照少数服从多数的原则, 被投票最多的那类就是样本的输出类别。

(5) 模型评估。随机森林算法使用自助法抽样的同时有约  $1/3$  的数据一次也未抽中, 被称为“袋外数据”(out-of-bag, OOB)。袋外数据是无偏的, 可直接用于测试评估模型。

## 2 城市火灾风险评估

### 2.1 指标选取

#### 2.1.1 指标选取的原则

一般来说, 火灾风险指标力求满足科学性、系统性、简洁性、可操作性、可比性等原则。

(1) 科学性。在科学分析的基础上选择指标, 不能凭主观判断。选取的指标要与火灾发生相关联, 富有依据。

(2) 系统性。从影响火灾发生的各方面选取指标, 能够全面地反映火灾风险。

(3) 简洁性。指标能够代表某一方面的特征, 相互关联但不包含, 避免增加评估难度。

(4) 可操作性。指标是为火灾风险的量化服务, 所选取的指标要具有可操作性, 能获取数据和资料, 不可操作的指标是无意义的。

(5) 可比性。指标有横向可比性时才有意义。评价对象具有共同影响因素, 在不同条件下具有不同状态才能互相比, 从而区分火灾风险。

#### 2.1.2 选取指标变量

本文采用专家咨询、频率分析、综合分析等方法初选了 15 个火灾风险评估指标<sup>[17-23]</sup>, 如表 1 所示, 结果满足科学性等 5 项原则。其中火灾密度作为因变量, 其余 14 个指标作为自变量。

表 1 15 个火灾风险评估指标

代码	指标名称
TEM	温度/℃
RAIN	降水/mm
DEM	海拔/m
SLOPE	坡度
NDVI	归一化植被指数
SHADE	山体阴影
POP	人口密度/(人/km <sup>2</sup> )
ROAD	道路密度/(km/km <sup>2</sup> )
RGDP	区域 GDP/(万元/km <sup>2</sup> )
CROWD	人员密集及火灾荷载大类场所密度/(个/km <sup>2</sup> )
FRAGILITY	人群脆弱类场所密度/(个/km <sup>2</sup> )
IMPORTANCE	重点保护类场所密度/(个/km <sup>2</sup> )
CHEMISTRY	易燃易爆有毒类场所密度/(个/km <sup>2</sup> )
ORDINARY	一般消防类场所密度/(个/km <sup>2</sup> )
FIRE	火灾密度/(起/km <sup>2</sup> )

## 2.2 基于随机森林算法的特征重要度计算与特征选择过程

以往研究中,选择指标后较少开展自变量与因变量之间的关系建模研究,但指标数量与结果精度并不一定成正相关。在随机森林算法中将指标称为“特征”,能够影响火灾风险分布的特征称为“相关特征”,不影响或作用很小的称为“无关特征”,与其他特征相关系数较高的称为“冗余特征”。评价特征对风险分布作用关键与否不能依据数量多少而定,需要结合研究区域特点定量分析。为此,本文设计了基于随机森林算法的特征选择过程,这是非常重要的数据预处理步骤。此举一是可以剔除无关特征和冗余特征,降低模型维度和训练的时间成本;二是也能够避免过拟合造成变量间的重复解释,提高预测精度。

### 2.2.1 特征重要度计算过程

随机森林算法计算特征重要度有多种方法,最常用的一种是基于“平均准确度下降(mean decrease accuracy)”的思想,即若随机打乱某个特征  $F$  的数据排列后,如果袋外数据的准确率大幅度降低,则说明这个特征对于样本分类或回归的结果影响很大,也就是说它的重要程度比较高。计算流程如图1所示。

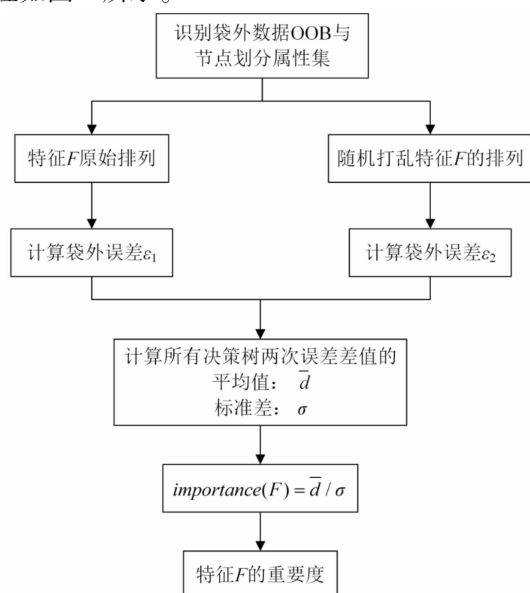


图1 特征重要度计算流程

具体操作如下:

(1)使用袋外数据计算每棵决策树的误差  $\varepsilon_1$ ;

(2)随机打乱某个特征  $F$  在样本中的顺序,再次计算袋外误差  $\varepsilon_2$ , 计算两次误差的差值  $d = \varepsilon_1 - \varepsilon_2$ ;

(3)计算所有决策树差值  $d$  的平均值  $\bar{d}$ , 计算标准差  $\sigma$ ;

(4)计算特征  $F$  的重要度,  $\text{importance}(F) = \bar{d} / \sigma$ 。

如果某个特征  $F$  与火灾风险分布关联大,重要程度高,则在随机排列后,袋外数据的准确率会大幅度降低,因此可依据此方法衡量特征重要度。

### 2.2.2 特征选择过程

所谓特征选择是指在得到特征重要程度后进行排名,从中选出比较重要的相关特征子集的过程。

本文基于递归特征消除策略进行特征选择,简而言之,是从完整的特征集开始,每次去掉一个最不重要的特征直至只剩一个特征,最终选择具有最大正确率的特征集合。具体步骤如下:

(1)假设初始样本特征集合为  $\{F_1, F_2, F_3, \dots, F_n\}$ , 将此样本集作为第一轮特征候选子集,基于随机森林算法计算每个特征的重要度并排序,使用袋外数据评估模型;

(2)从当前特征集合中删除最不重要的特征,剩余  $n-1$  个特征作为第二轮候选子集重复前述操作;

(3)重复第(1)和(2)步,直至最后一个特征;

(4)假定第  $i$  轮特征子集的误差最小,则将此集合中的特征作为最优特征子集输出,删除其余特征。

### 2.3 指标赋权与火灾风险的计算

使用特征重要度  $\text{importance}(F)$  归一化的结果作为指标的权重。此权重一是经过定量分析得到,不涉及主观经验的缺陷;二是将重要度转换成一个相对的权重并不改变其中的关系,同时通过这样类似于归一化的操作,可以避免权重过于离散,从而可以实现数据更好的收敛。指标  $F_j$  的权重计算公式如下:

$$w(F_j) = \frac{\text{importance}(F_j)}{\sum_{i=1}^n F_i} \quad (3)$$

使用线性加权法计算区域  $i$  的火灾风险  $r$  为:

$$r(i) = \sum_{i=1}^n w(F_i) x(F_i) \quad (4)$$

式中:  $x(F_i)$  为指标  $F_i$  处的值。

## 3 实例验证

本文实例研究区域是济南市区,包括历下区、市中区、天桥区、槐荫区、历城区和长清区的全部,面积 3 257 km<sup>2</sup>,所使用的地理数据均在 WGS 1984 UTM Zone 50N 坐标系下处理。

本文用于评估城市火灾风险的随机森林模型基于 MATLAB 2020a 设计。

### 3.1 数据来源

温度、海拔、坡度、归一化植被指数数据来自地理空间数据云对应产品;从全球天气和气候数据库 WorldClim 获取 2018 年降水数据;山体阴影基于 Arcgis10.6 表面分析工具箱计算;从人口动态统计分析数据库 LandScan 中获取 2018 年人口密度数据;道路密度原始数据是从百度地图获取的路网信息,之后基于 Arcgis10.6 线密度分析工具获取区域道路密度;区域生产总值来自中科院资源环境科学与数据中心的 2015 年度网格数据;以区域内“兴趣点”(Point of Interest, POI)数据密度表征人员密集及火灾荷载大类等 5 类建筑密集程度,数据来源于高德地图在线服务,获取时间为 2020 年 1 月 27 日;从济南市消防救援支队获取 2013—2016 年火灾数据,基于 Arcgis10.6 核密度分析工具获取火灾密度。

### 3.2 数据预处理

由于各指标数据的量纲和数量级不同, 建模前分为两步对 15 个指标进行无量纲处理。一是正、负向指标的判断。正向自变量指标值越大, *FIRE* 越大; 负向指标相反。通过分析原始指标间的相关性得出 *DEM*、*SLOPE*、*NDVI* 与 *FIRE* 都是负相关, 划分为逆向指标; 其余指标划分为正向指标。二是使用式(5)、式(6)的归一化函数对指标进行无量纲处理。

$$\text{正向指标: } x_j = \frac{x - x_{\min}}{x_{\max} - x_{\min}}; \quad (5)$$

$$\text{负向指标: } x_j = \frac{x_{\max} - x}{x_{\max} - x_{\min}}. \quad (6)$$

式中:  $x_j$  是指标  $j$  无量纲处理后的值;  $x_{\max}$ 、 $x_{\min}$  分别是原始指标的最大与最小值。

### 3.3 随机森林算法参数的确定

随机森林算法有多个重要参数。一是随机选择进入每棵决策树的特征数 (*NumPredictorsToSample*), 本文采用程序推荐的  $n/3$ ; 二是最小叶节点样本数 (*MinLeafSize*) 与决策树数量 (*NumTrees*), 前者取值越小与后者取值越大越好, 但时间成本较大且容易陷入过拟合。本文通过尝试, 最终取 *MinLeafSize* = 5, *NumTrees* = 30 进行随机森林的建模。

### 3.4 特征选择过程

基于 2.2 节构建的递归特征消除法进行特征选择, 即在随机森林算法计算得到的特征重要性基础上, 每次删除一个最不重要的特征直到只有 1 个特征时为止, 均方差 (Mean Squared Error, MSE) 最小的特征个数与集合就是最优结果<sup>[24]</sup>。特征选择过程如图 2 所示, 各步骤具体内容如表 2 所示。

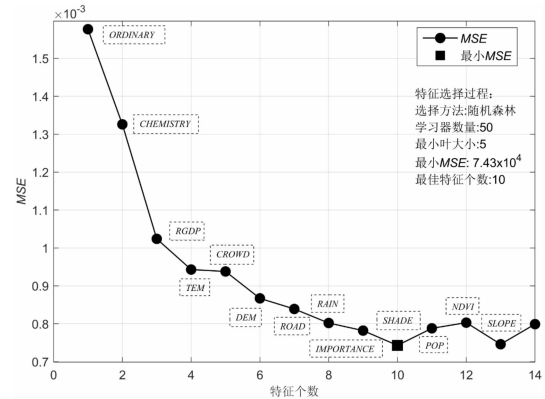


图2 特征选择过程

从特征选择的结果看, 随机森林模型在 10 个特征时精度最高, 误差最小, 此时  $MSE = 7.43 \times 10^{-4}$ , 剔除的 4 个指标为 *SLOPE*、*SHADE*、*NDVI* 和 *POP*。结合特征之间的 Pearson 相关系数可知, 如图 3 所示, *SLOPE*、*NDVI*、*SHADE* 与 *FIRE* 之间的相关系数分别为 0.17、0.22、0.02, 说明在济南市区范围内, 3 个指标对火灾分布影响不大, 删除之后可以简化模型。虽然 *POP* 与 *FIRE* 的相关系数是 0.76, 但 *ROAD*、*CROWDED*、*FACILITY*、*IMPORTANCE* 和 *ORDINARY* 与其相关性都大于 0.7, 一定程度上反映了 *POP* 数据, 且这些指标与 *FIRE* 的相关性都大于 0.8, 高于 *POP*, 模型从多重共线性角度出发将此指标剔除。虽然 *RAIN*、*DEM* 与 *FIRE* 的相关性也较低, 但由图 2 和表 2 知, 在特征选择的过程中剔除两项指标后, 模型的 *MSE* 会增大, 因此最终保留了 *RAIN* 和 *DEM* 指标。

表2 特征选择过程具体步骤

步骤	剔除特征	保留特征	MSE( $\times 10^{-4}$ )
1	无	TEM, RAIN, DEM, SLOPE, NDVI, SHADE, POP, ROAD, RGDP, CROWD, FRAGILITY, IMPORTANCE, CHEMISTRY, ORDINARY	7.99
2	SLOPE	TEM, RAIN, DEM, NDVI, SHADE, POP, ROAD, RGDP, CROWD, FRAGILITY, IMPORTANCE, CHEMISTRY, ORDINARY	7.46
3	NDVI	TEM, RAIN, DEM, SHADE, POP, ROAD, RGDP, CROWD, FRAGILITY, IMPORTANCE, CHEMISTRY, ORDINARY	8.03
4	POP	TEM, RAIN, DEM, SHADE, ROAD, RGDP, CROWD, FRAGILITY, IMPORTANCE, CHEMISTRY, ORDINARY	7.88
5	SHADE	TEM, RAIN, DEM, ROAD, RGDP, CROWD, FRAGILITY, IMPORTANCE, CHEMISTRY, ORDINARY	7.43
6	IMPORTANCE	TEM, RAIN, DEM, ROAD, RGDP, CROWD, FRAGILITY, CHEMISTRY, ORDINARY	7.82
7	RAIN	TEM, DEM, ROAD, RGDP, CROWD, FRAGILITY, CHEMISTRY, ORDINARY	8.02
8	ROAD	TEM, DEM, RGDP, CROWD, FRAGILITY, CHEMISTRY, ORDINARY	8.39
9	DEM	TEM, RGDP, CROWD, FRAGILITY, CHEMISTRY, ORDINARY	8.67
10	CROWD	TEM, RGDP, FRAGILITY, CHEMISTRY, ORDINARY	9.38
11	TEM	RGDP, FRAGILITY, CHEMISTRY, ORDINARY	9.43
12	RGDP	FRAGILITY, CHEMISTRY, ORDINARY	10.24
13	CHEMISTRY	FRAGILITY, ORDINARY	13.26
14	ORDINARY	FRAGILITY	15.77

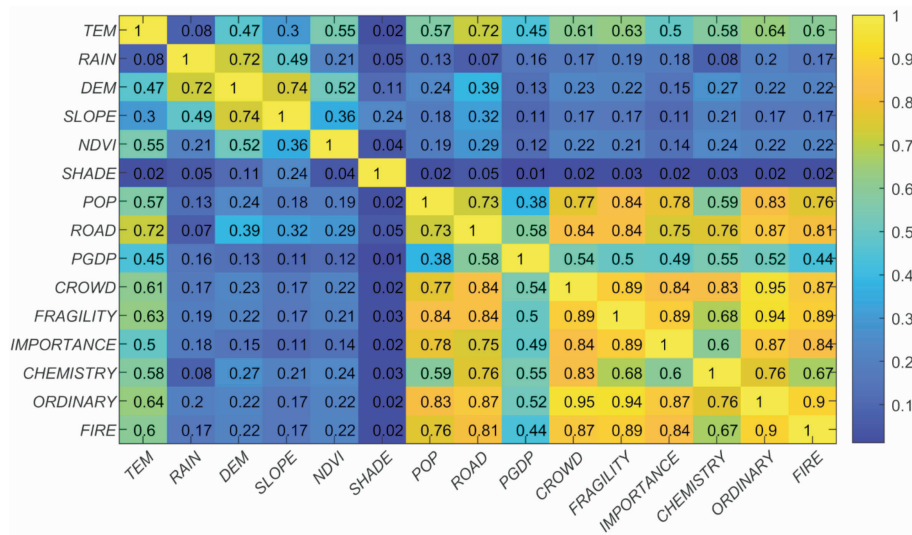


图3 特征之间的 Pearson 相关系数

### 3.5 指标权重计算

使用特征选择过程所保留的 10 个指标重复建立 50 次随机森林模型，将袋外数据作为验证集，误差评价采用  $MSE$ 。计算结果显示，虽然随机森林算法的双重随机性导致 50 次结果不完全一致，但  $MSE$  保持在  $(8.0 \pm 0.4) \times 10^{-4}$  范围内。

以某次结果为例，拟合效果如图 4 所示，黑线是完美预测直线，蓝色点横坐标是实际值，纵坐标是预测值。从图 4 中可知，随机森林模型对大部分样本的拟合误差不大，但面对取值较大的样本时容易低估实际值，其中模型完全低估实际值  $> 0.6$  的样本。由图 5 济南市区火灾密度分布知，火灾密度较大的区域大都集中在中心城区，这些区域在经济、人口、用电等方面高度集中，老城区也一般在集中在这，消防隐患大，是火灾高发区；同时这些高火灾密度样本在总样本中占比较低，样本数少，而其他相对低火灾密度的样本较多，造成模型容易低估 FIRE 取值较大的样本点。通过计算，随机森林模型对总体样本拟合的决定系数  $R^2 > 0.85$ 。

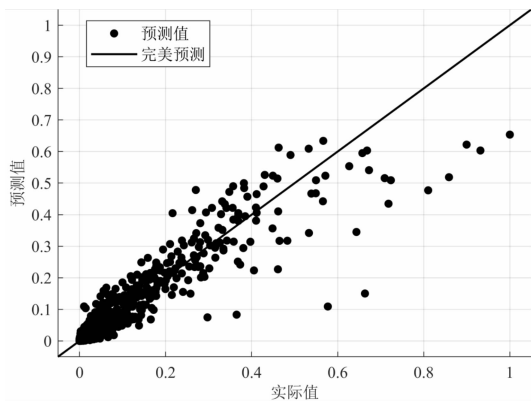


图4 随机森林模型拟合效果

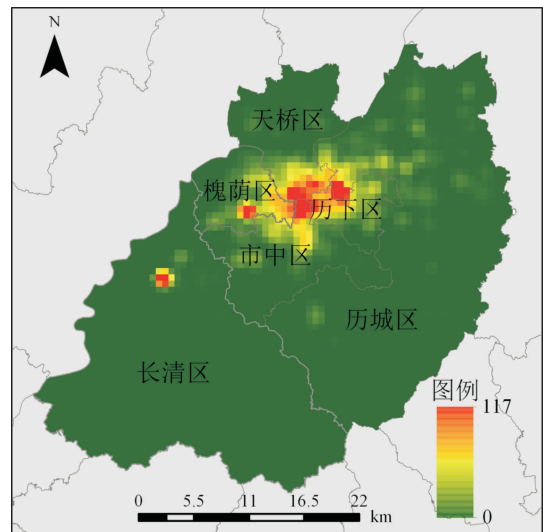


图5 济南市区火灾密度(审图号：鲁 SG(2020)019 号，下同)

取  $MSE$  最小的 10 个模型计算指标重要度的平均值，结果如表 3 所示。根据式(3)计算得到各指标的权重，结果如图 6 所示。

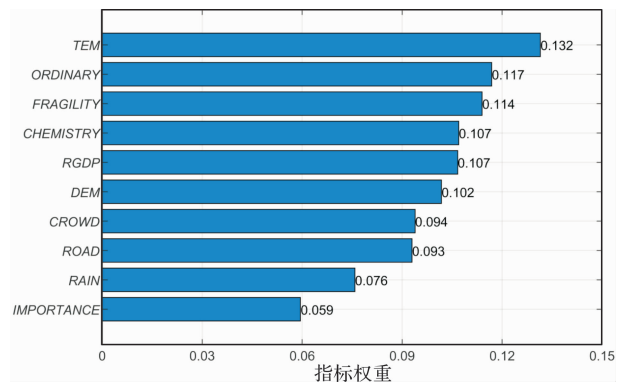


图6 指标权重



表 3 10 次随机森林模型结果

特征	TEM	RAIN	DEM	ROAD	RGDP	CROWD	FRAGILITY	IMPORTANCE	CHEMISTRY	ORDINARY	MSE( $\times 10^{-4}$ )
1	0.875	0.970	0.948	1.218	0.660	0.590	0.857	0.592	1.427	1.180	7.569
2	0.962	0.733	0.705	0.578	0.925	0.673	0.716	0.564	0.837	0.903	7.571
3	0.927	0.501	0.678	0.915	0.985	0.708	0.989	0.621	0.754	0.823	7.589
4	1.237	0.418	0.769	0.762	0.778	1.187	0.899	0.445	0.896	1.151	7.747
5	1.040	0.594	1.000	0.609	0.735	0.618	1.013	0.268	0.585	0.930	7.751
6	1.153	0.560	0.756	0.618	0.952	0.930	0.823	0.377	0.802	0.684	7.849
7	1.056	0.593	0.890	0.993	0.670	0.745	0.902	0.388	0.775	0.720	7.854
8	1.212	0.669	0.655	0.600	0.845	0.543	0.827	0.441	0.995	1.052	7.885
9	0.855	0.535	0.835	0.591	0.905	0.735	1.113	0.414	0.760	0.717	7.990
10	1.160	0.473	0.876	0.524	1.047	0.754	0.944	0.626	0.695	1.156	7.999
平均值	1.048	0.604	0.811	0.741	0.850	0.748	0.908	0.474	0.853	0.931	7.780

### 3.6 济南市区火灾风险评估结果分析

根据式(4)计算得到济南市区的火灾风险分布,结果如图7所示。

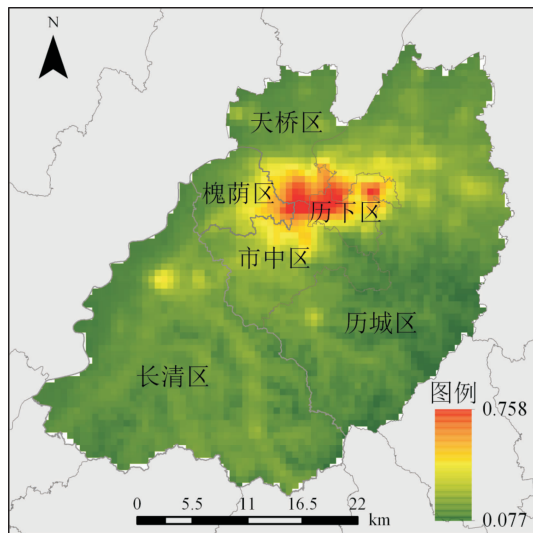


图7 济南市区火灾风险分布

火灾风险评估使用了基于随机森林算法得到的客观权重,不具有主观经验的不足,结果同济南市区火灾密度分布相比:

(1)相同的是,评估结果对较高火灾密度区域有较好的拟合,如中心城区在评估结果中处于红色的较高风险值;长清区城区的历史火灾较高,此区域的评估值高于周边区域。

(2)不同的是,评估结果中风险处于浅黄色的区域较多,而其中多数区域历史火灾很少甚至为0。分析原因,根据文献<sup>[25]</sup>,火灾风险=概率×后果,并不简单由历史火灾反映。虽然上述区域4年内发生火灾较少,但并不代表未来不会发生火灾。文中风险评估使用了10个指标,由于指标值不为0,得到非0风险值可显示其发生火灾的相对可能性。

## 4 结论与讨论

(1)初始构建的指标并非最佳指标集合,其中包含无关或冗余指标,既增加了评估的难度也影响准确度。随机森林算法可在评估城市火灾风险前定量分析指标自变量与因变量关系,只保留最佳指标集合,方法简单,实用性强。

(2)不同专家对指标权重判断不同。引入随机森林算法建立自变量与火灾因变量的关系模型后,可对指标客观赋权,拟合精度较高,能解决层次分析等方法中有主观经验的问题。

(3)仅以历史火灾数据表示城市火灾风险是片面的。使用随机森林算法进行指标筛选与赋权后,依据线性加权法计算的结果会得到历史火灾较少或未发生火灾区域的风险得分。虽然这些历史火灾不多,但风险值表示了未来发生火灾的可能性,能提示有关部门采取针对性防灾减灾措施。

(4)因篇幅有限,没有与其他评估方法进行对比,下一步可加强对比研究。

## 参考文献:

- [1] GB 51080—2015 城市消防规划规范[S]. 北京:中国工业出版社,2015.
- [2] MICHAEL S W. Dwelling Risk Assessment[M]. Toolkit, 1999.
- [3] Entec. Review of high occupancy risk assessment[M]. Toolkit, 2000.
- [4] 连旦军,董希琳,吴立志. 城市区域火灾风险评估综述[J]. 消防科学与技术, 2004, 23(3): 240-242.
- [5] 杜霞,张欣,刘庭全,等. 国外区域火灾风险评估技术及应用现状[J]. 消防科学与技术, 2004, 23(2): 137-139.
- [6] YOSHIYUKE Y. Evaluating building fire safety through egress prediction: a standard application in Japan[J]. Fire Technology, 1995, 31(2): 158-174.
- [7] 苏拓. 城市消防安全影响因素及评估体系研究[D]. 合肥:中

- 国科学技术大学, 2018.
- [8] 刘纪华. 宝鸡市区域火灾风险评估方法研究及应用[D]. 西安: 西安科技大学, 2018.
- [9] 刘晗, 王健. 基于 Fine Kinney 和 Dematel 理论的城市火灾风险评估[J]. 消防科学与技术, 2018, 37(12): 1730–1734.
- [10] 张刚. 基于空间分析的城市火灾风险评估与应用——以西安为例[J]. 城市规划, 2016, 40(8): 59–64.
- [11] 陈国良, 胡锐, 卫广昭. 北京市火灾风险综合评估指标体系研究[J]. 中国安全科学学报, 2007, 17(4): 119–124, 180.
- [12] 崔冠雄. 基于灰色层次的城市火灾风险评估[J]. 科技信息, 2012(20): 154–155.
- [13] OLIVEIRA S, OEHLER F, SAN-MIGUEL-AYANZ J, et al. Modeling spatial patterns of fire occurrence in mediterranean Europe using multiple regression and random forest[J]. Forest Ecology & Management, 2012, 275(4): 117–129.
- [14] LIANG H L, LIN Y R, YANG G, et al. Application of random forest algorithm on the forest fire prediction in Tahe area based on meteorological factors[J]. Scientia Silvae Sinicae, 2016, 61(1): 89–99.
- [15] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5–32.
- [16] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [17] 谭常春, 谭景宝, 朱华亮. 城市火灾次数与气象因素的 Adaptive-Lasso 分析[J]. 应用数学与计算数学学报, 2013, 27(3): 408–414.
- [18] 贾旭, 高永, 魏宝成, 等. 基于 MODIS 数据的内蒙古地形因子对火灾分布的影响分析[J]. 北京林业大学学报, 2017, 39(5): 34–40.
- [19] 马文苑, 冯仲科, 成竺欣, 等. 山西省林火驱动因子及分布格局研究[J]. 中南林业科技大学学报, 2020, 40(9): 57–69.
- [20] SONG C, MEI P K, SONG W G, et al. A comparison between spatial econometric models and random forest for modeling fire occurrence[J]. Sustainability, 2017, 9(5): 819.
- [21] 李国辉, 王颖, 原志红, 等. 火灾发生率与经济因素的时空相关性分析[J]. 灾害学, 2016, 31(2): 111–115.
- [22] 祝明明, 罗静, 余文昌, 等. 城市 POI 火灾风险评估与消防设施布局优化研究——以武汉市主城区为例[J]. 地域研究与开发, 2018, 37(4): 86–91.
- [23] 徐智邦, 周亮, 蓝婷, 等. 基于 POI 数据的巨型城市消防站空间优化——以北京市五环内区域为例[J]. 地理科学进展, 2018, 37(4): 535–546.
- [24] 侯晓静, 明金科, 秦荣水, 等. 基于随机森林模型的交界域火灾风险分析[J]. 林业科学, 2019, 55(8): 194–200.
- [25] 郑红梅, 陈国良, 王玮, 等. 火灾风险评估相关概念辨析[J]. 中国安全科学学报, 2008, 18(6): 75–79.

## Research on Urban Fire Risk Evaluation Based on The Random Forest Algorithm

WU Lizhi<sup>1</sup>, CHEN Zhennan<sup>2</sup> and ZHANG Peng<sup>3</sup>

(1. China People's Police University, Langfang 065000, China;

2. School of Graduate, China People's Police University, Langfang 065000, China;

3. School of Intelligent Police, China People's Police University, Langfang 065000, China)

**Abstract:** In order to quantitatively analyze urban fire risk, random forest algorithm is introduced. Firstly, a recursive feature elimination method based on random forest algorithm is designed to eliminate irrelevant or redundant variables from 14 indicators selected initially; secondly, feature importance is calculated through random forest with average accuracy reduction method, whose normalized result is then taken as indicator weight; finally, combined with index data, urban fire risk is evaluated by linear weighting method. Taking Jinan urban area as an example, the case study showed that: after selecting features, random forest model only preserved a set containing 10 factors with minimum  $MSE = 7.43 \times 10^{-4}$ ; compared with fire density, the determination coefficient ( $R^2$ ) of random forest model used to calculate index weight is greater than 0.85, proving that objective weight is feasible; the evaluation result not only revealed that high fire density districts were high risk areas, also can get fire risk value of those low fire density areas.

**Key words:** urban fire; risk evaluation; random forest algorithm; feature selection; index weight