



DOI:10.12404/j.issn.1671-1815.2405854

引用格式:杨正雄,张显云,任明亚,等.基于CatBoost的四川木里县森林火灾预测[J].科学技术与工程,2025,25(21):8823-8832.

Yang Zhengxiong, Zhang Xianyun, Ren Mingya, et al. Forest fire prediction in Muli County, Sichuan based on CatBoost[J]. Science Technology and Engineering, 2025, 25(21): 8823-8832.

农业科学

## 基于 CatBoost 的四川木里县森林火灾预测

杨正雄,张显云\*,任明亚,吴雪,龙安成

(贵州大学矿业学院,贵阳 550025)

**摘要** 森林火灾严重威胁着人类生命和财产安全,森林火灾风险的精确预测对于减灾防灾具有重要意义。受地形、气象、植被覆盖和人类活动等因素的影响,森林火灾诱发的原因存在区域差异性。以四川省木里县历史森林火灾事件为响应变量,以地形、气象、植被覆盖和人类活动数据为解释变量,充分发挥 CatBoost 在高维稀疏数据和分类问题方面的优势,构建了一种基于 CatBoost 的高精度森林火灾预测模型。实验结果表明,相较于随机森林(random forest, RF)、极端梯度提升(extreme gradient boosting, XGBoost)和梯度提升决策树(gradient boosting decision trees, GBDT)模型, CatBoost 模型不仅可获得更高的建模精度,而且森林火灾的预测精度也得到了显著改善,预测准确率达 91.36%,曲线下的面积(area under curve, AUC)为 0.970。基于所构建模型进行火灾预测,可为木里县森林火灾的早期防范提供参考依据。

**关键词** 森林火灾预测模型;木里县;森林火灾;CatBoost;准确率

中图法分类号 S762.2;

文献标志码 A

### Forest Fire Prediction in Muli County, Sichuan Based on CatBoost

YANG Zheng-xiong, ZHANG Xian-yun\*, REN Ming-ya, WU Xue, LONG An-cheng

(Mining College, Guizhou University, Guiyang 550025, China)

**[Abstract]** Forest fires pose a significant threat to human lives and property. Accurate prediction of forest fire risk is crucial for disaster mitigation and prevention. Influenced by factors such as terrain, meteorology, vegetation cover, and human activities, the causes of forest fires exhibit regional differences. This study uses historical forest fire events in Muli County, Sichuan Province as the response variable, with terrain, meteorological data, vegetation cover, and human activity data as explanatory variables. Leveraging CatBoost's strengths in handling high-dimensional sparse data and classification problems, a high-precision forest fire prediction model was constructed based on CatBoost. The experimental results indicate that, compared to random forest (RF), extreme gradient boosting (XGBoost), and gradient boosting decision trees (GBDT) models, the CatBoost model achieves higher modeling accuracy and significantly improves forest fire prediction accuracy, with a prediction accuracy rate of 91.36% and an area under curve (AUC) value of 0.970. Predictions made using this model can provide valuable references for the early prevention of forest fires in Muli County.

**[Keywords]** forest fire prediction model; Muli County; forest fire; CatBoost; accuracy

森林火灾作为一种严重的自然灾害,对生态环境和人类社会均有着巨大的影响。随着全球气候变化和人类活动的增加,森林火灾频发,森林防火工作异常严峻<sup>[1-2]</sup>。森林火灾不仅会造成巨大的经济损失,而且还威胁着人类生命安全和生态系统的稳定。火灾引发的烟雾和有害气体,将会严重污染区域空气质量,并危及动植物、土壤甚至于微生物,给人类健康与生态安全带来不可忽视的影响<sup>[3-4]</sup>。因此,森林火灾的预防对于维护社会稳定和生态平

衡至关重要。

对森林火灾风险进行高精度预测,是预防森林火灾的强有力措施。近年来,机器学习在森林火灾预测中的应用研究已取得众多丰硕成果。Milanovic等<sup>[5]</sup>的研究发现,干旱指数是东塞尔维亚林火发生的最重要变量,其次是各种人为因素,且随机森林模型的预测能力优于逻辑回归模型。Prapas等<sup>[6]</sup>提出了一种捕捉时空数据的深度学习模型,成功运用于希腊单日野火风险预报。汪祖民等<sup>[7]</sup>以云南省

收稿日期:2024-08-04 修订日期:2025-04-11

基金项目:贵州省省级科技计划(黔科合支撑[2022]一般204);贵州省省级科技计划(黔科合基础-ZK[2024]一般093)

第一作者:杨正雄(2000—),男,侗族,贵州黎平人,硕士研究生。研究方向:资源环境遥感与算法。E-mail:1037899052@qq.com。

\*通信作者:张显云(1974—),男,汉族,贵州遵义人,硕士,副教授。研究方向:高分辨率信息影响提取。E-mail:mec.xy Zhang@gzu.edu.cn。

投稿网址:www.stae.com.cn

为研究区,采用 LightGBM 进行火灾预测,准确率达到 90.5%。张运林等<sup>[8]</sup>在分析贵州省 10 年森林火灾时空格局中,在确定出森林火灾驱动因子的基础上,构建了一种森林火灾概率预测模型,实验结果表明模型预测准确率为 81.9%。李史欣等<sup>[9]</sup>开展了安徽省滁州韭山森林火灾风险预测模型研究,结果表明随机森林较逻辑回归模型预测效果更好。张全文等<sup>[10]</sup>基于元胞自动机模拟火灾蔓延,将火灾蔓延过程转换为三维可视化,在林火动态监测方面具有指导意义。苗新等<sup>[11]</sup>以阿尔及利亚 Bejaia 地区和 Sidi Bel-abbesl 地区森林火灾为研究区,开展基于机器学习的森林火灾风险预测研究,结果表明相较于神经网络和逻辑回归,随机森林可获得更高的预测精度。郝婕等<sup>[12]</sup>以嘉陵江流域重庆段为研究区,分别基于随机森林、支持向量机、人工神经网络和梯度提升决策树 (gradient boosting decision trees, GBDT) 4 种机器学习模型对火灾进行评估,发现模型性能梯度提升决策树最优。罗永明等<sup>[13]</sup>以历史数据建立森林火险气象指数模型,利用加权算法得到预报指数,并对火灾样本进行评估,得到有效的预报提升。总的来说,机器学习算法因其卓越的性能,在森林火灾预测中得到了越来越多的应用<sup>[14-16]</sup>。

CatBoost (categorical boosting) 作为一种基于对称决策树的机器学习算法,具有参数少、不易发生过拟合或欠拟合,以及在处理高维稀疏数据和分类问题方面更具优势等优点,已被广泛应用于医疗、气象、金融、建筑、能源等领域<sup>[17-21]</sup>,但迄今为止其在森林火灾风险预测中的应用研究却相对较少。此外,受到当地气象、植被覆盖、地形、社会、经济以及人文等因素的共同作用,森林火灾发生的原因具有空间差异性<sup>[22-23]</sup>。基于此,现以四川省木里县森林火灾事件为研究对象,顾及地形、气象、植被覆盖和人类活动等的影响,充分发挥 CatBoost 的优势,开展木里县森林火灾风险预测模型和火灾风险等级划分研究,以期对相关决策和火灾管理提供科学依据,并为具有类似地理环境和气候条件地区的森林火灾预防提供参考。

## 1 研究区与数据

### 1.1 研究区

研究区为四川省木里藏族自治县,地处青藏高原东南缘,横断山脉中段东侧,介于东经 100°03′~101°40′,北纬 27°40′~29°10′。研究区东西宽约 160 km,南北长约 170 km,总面积 13 252.7 km<sup>2</sup>。木里县林业资源丰富,森林覆盖率达 69.86%。受地理位置、气候条件和地形地貌等因素的共同影响,木

里县森林火灾频发,2005—2009 年 5 年间共发生森林火灾 42 次,其中一般森林火灾 30 次,较大森林火灾 9 次,重大森林火灾 3 次。其中,2019 年 3 月 30 日—4 月 10 日的森林火灾最为严重,造成 31 人遇难。火灾起因是雷击,火场面积达到 69.24 km<sup>2</sup>,受损森林面积约 17.56 km<sup>2</sup>。图 1 为研究区的地理位置及 2010—2020 年所发生的森林火灾火点。

### 1.2 研究数据及时空匹配

以四川省木里县历史森林火灾事件为响应变量(因变量),以地形、气象、植被覆盖和人类活动数据为解释变量(自变量),各类数据的来源、空间分辨率和格式如表 1 所示。火点主要集中在县域的东南部和西南部。这些区域可能由于特定的植被类型、地形条件和人类活动而成为火灾高发区。除此之外,火点在中部地区及其他区域呈零散分布,显示出较低的火灾发生率或不同的致灾因素。

#### 1.2.1 研究数据

##### 1) 火点及非火点数据

火点数据为 2010—2020 年的 MODIS 1 km 分辨率火点数据集,该数据集由 NASA 提供,并经过了地理位置和时间精度的校正,主要包含火灾发生的地理位置、火灾发生时间、火灾等级等信息。火点数据分布如图 1 所示。非火点数据由随机采样生成,数量与火点数量相当,遵循时间和空间双重随机性原则。建模时对火点赋值 1,非火点赋值 0。

##### 2) 地形数据

地形数据包括海拔、坡度、坡向及表面曲率。首先,海拔高度影响着植被的类型和含水量,从而间接影响着植被的可燃性<sup>[24]</sup>。其次,坡度增加了火焰向上蔓延的速度,因为热量的上升趋势预热了上方的植被,降低了其点燃温度,从而加快了火焰的蔓延。再次,坡向会影响植被接受太阳辐射能量的大小,太阳辐射能量决定植被含水量<sup>[25]</sup>。最后,平面曲率决定着火灾传播速度,平坦地区传播更快。海拔源于 ASTER 30 m DEM 数据,而坡度、坡向和表面曲率可由海拔间接计算得到。坡度数据由 Arcgis10.3 软件基于式(1)计算得到,公式为

$$s = 57.3 \tan^{-1} \left[ \sqrt{\left(\frac{dz}{dy}\right)^2 + \left(\frac{dz}{dx}\right)^2} \right] \quad (1)$$

式(1)中:  $s$  为坡度;  $dz/dx$  为像元中心水平方向上的变化率;  $dz/dy$  为像元中心垂直方向上的变化率<sup>[26]</sup>。

坡向数据由 Arcgis10.3 软件基于式(2)计算得到,公式为

$$A = \frac{180}{\pi} \tan^{-1} \left( \frac{dz/dy}{dz/dx} \right) + 180 \quad (2)$$

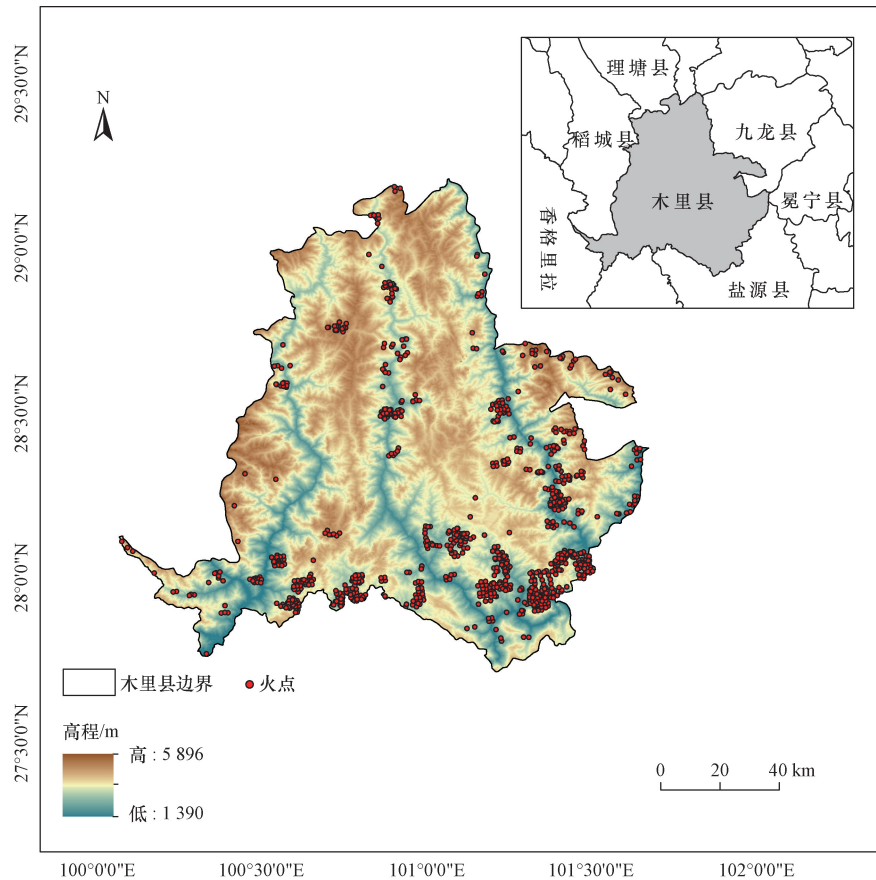


图1 研究区以及火点

Fig. 1 Research area and fire spots

表1 数据相关信息

Table 1 Data related information

数据	来源	网址	精度	格式
高程	地理空间数据云	<a href="https://www.gscloud.cn/">https://www.gscloud.cn/</a>	30 m	栅格
归一化植被指数	国家地球数据系统科学数据中心	<a href="https://www.geodata.cn/">https://www.geodata.cn/</a>	1 km	栅格
气象数据	国家地球数据系统科学数据中心	<a href="https://www.geodata.cn/">https://www.geodata.cn/</a>	1 km	栅格
道路数据	全国地理信息资源目录服务系统	<a href="http://www.webmap.cn">www.webmap.cn</a>	—	矢量
水系数据	全国地理信息资源目录服务系统	<a href="http://www.webmap.cn">www.webmap.cn</a>	—	矢量
居民点数据	全国地理信息资源目录服务系统	<a href="http://www.webmap.cn">www.webmap.cn</a>	—	矢量
火点数据	NASA FIRMS	<a href="https://firmsmodaps.eosdis.nasa.gov">https://firmsmodaps.eosdis.nasa.gov</a>	—	矢量

式(2)中: $A$ 为坡向; $dz/dx$ 为像元中心水平方向上的变化率; $dz/dy$ 为像元中心垂直方向上的变化率。结果范围: $0^\circ \sim 360^\circ$ ,表示从北方顺时针的角度。

表面曲率数据由 Arcgis10.3 软件基于式(3)计算得到,公式为

$$C = \frac{d^2z}{dx^2} + \frac{d^2z}{dy^2} \quad (3)$$

式(3)中: $C$ 为表面曲率; $d^2z/dx^2$ 为像元中心水平方向上的变化率二阶偏导数; $d^2z/dy^2$ 为像元中心垂直方向上的变化率二阶偏导数。研究区坡度、

坡向、高程和表面曲率如图 2(a) ~ 图 2(d) 所示。

### 3) 气象数据

建模所用气象数据主要包括气温、降水、相对湿度和风速。首先,气温升高会导致植被干燥,火灾易发,且火焰蔓延速度更快。其次,降水影响植被含水量,从而影响火灾的发生概率和蔓延速度。再次,高相对湿度使植被湿润,减少易燃性,低湿度则增加火灾风险。最后,强风加速火势蔓延,提供更多氧气,使火焰扩散迅速,并可能形成火旋风,增加火灾的不稳定性和破坏力。研究区 2018 年 6 月气象数据月均值如图 2(e) ~ 图 2(h) 所示。

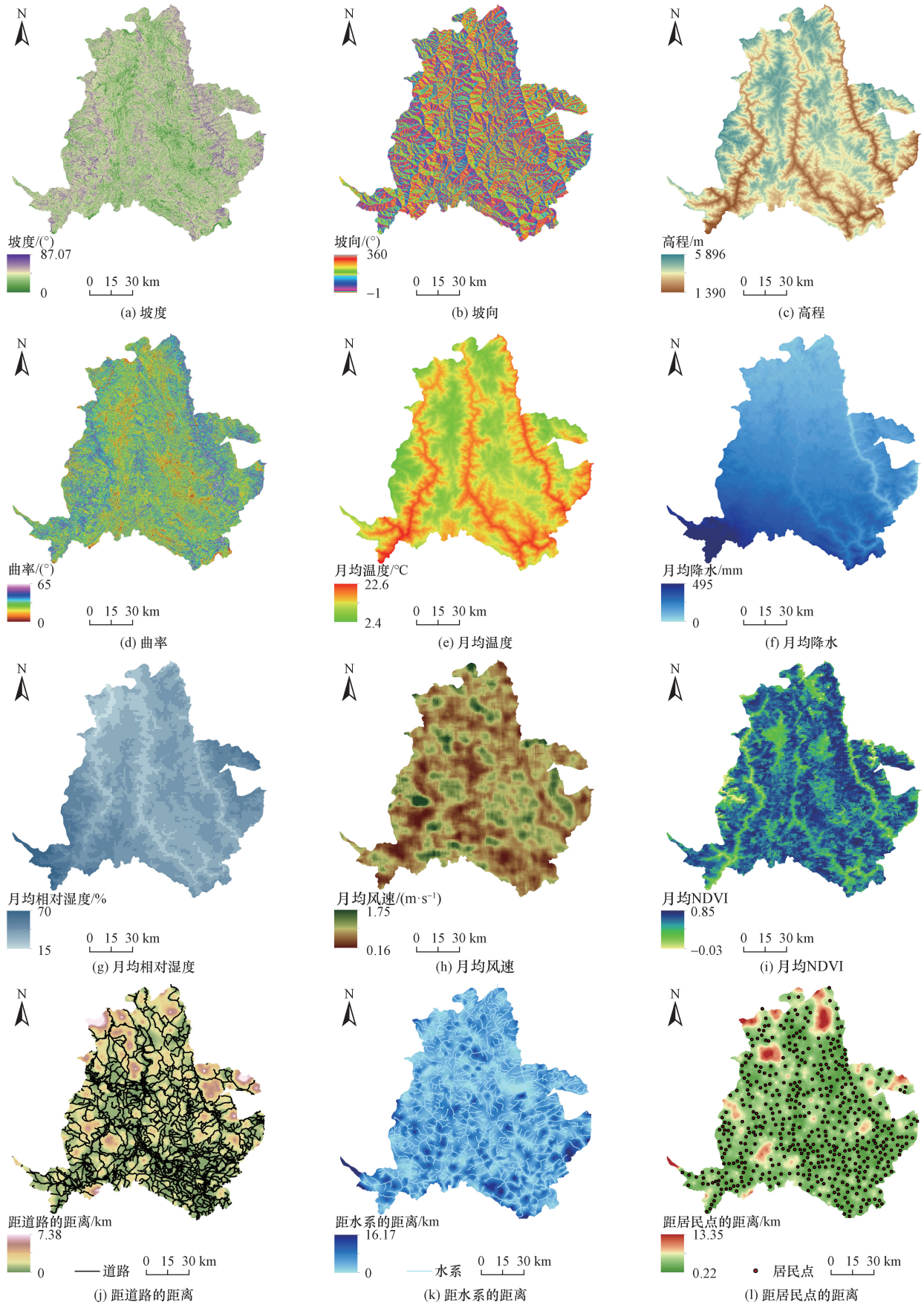


图2 影响因子制图

Fig. 2 Impact factor mapping

#### 4) 植被覆盖

归一化植被指数(normalized difference vegetation index, NDVI)是表征植被生长态势的重要指标,可以反映出植被的健康状态。一般而言,较高的 NDVI 数值通常表示植被生长状况良好,而在植被生长状况较差或受到干旱等因素影响的地区,NDVI 的数值则往往较低。因此,较低的 NDVI 值可能暗示着植被的生长状况不佳,从而增加森林火灾发生的风险。研究区 2018 年 6 月归一化植被指数月均值如图 2(i) 所示。

#### 5) 人类活动数据

人类活动是诱发森林火灾的另一重要原因。首先,人类活动或车辆排放可能诱发森林火灾,导致道路附近森林火灾的发生概率更高<sup>[27]</sup>。其次,远离水源的区域由于植被湿度低,更易发生火灾。另外,靠近居民点的区域人类火点密集,火灾可能性更高。由国家基础地理信息中心获得地图矢量数据,利用 Arcgis10.3 软件的“近邻分析”功能计算与火点和非火点的距离,再利用反距离加权插值生成栅格数据得到,结果如图 2(j)~图 2(l) 所示。

#### 1.2.2 数据的时空匹配

建模响应变量由 2010—2020 年的火点数据集及非火点构成。非火点数据由随机采样生成,数量与火点数量相当,遵循时间和空间双重随机性原则。建模时对火点赋值 1,非火点赋值 0。建模解释变量包括高程、坡度、坡向、平均曲率、归一化植被指数、降水、温度、风向、相对湿度、与道路的距离、与水系的距离、与居民点的距离。

为了确保模型中使用的火点数据与各类因子(地形、气象、人类活动、植被覆盖等)的一致性,对数据的时间分辨率和空间分辨率进行了统一处理。火点数据被与相关因子数据在相同的时间和空间尺度上进行匹配,让每一个火点事件都能与相应时间和空间的环境条件相联系。利用 Arcgis 10.3 软件中的“多值提取至点”,根据火点与非火点的时间与空间位置,将自变量数据提取至对应的点数据上。

## 2 研究方法及精度评价指标

### 2.1 研究方法

CatBoost 是一种基于 GBDT 的集成学习的改进算法,专门设计用于处理包含类别型特征的数据集。CatBoost 在梯度提升树的基础上,引入了对类别型特征的自动处理、基于加权样本的排序方法等,从而提高了模型的性能和泛化能力<sup>[28]</sup>。森林火灾预测通常涉及多种类型的特征,包括数值型和类别型特征,而 CatBoost 能够直接处理类别型特征,

减少了特征工程的复杂度,提高了模型的训练效率。此外,CatBoost 在处理缺失值和异常情况方面具有较强的鲁棒性,能够处理实际数据中存在的各种异常情况,并通过对称二叉树(如图 3 所示)和排序学习技术来减少模型发生过拟合的风险,有助于提高模型的稳定性和可靠性。训练一个单独的模型  $M_i$ ,训练模型  $M_i$  的数据是不包含  $x_i$  的训练集,然后使用模型  $M_i$  对样本的梯度进行估计,最后使用此梯度训练基学习器得到最终模型。转化公式为

$$x_{i,k} = \frac{\sum_{x_{i,j} \in D_k} [x_{i,k} = x_{i,j}] y_j + ap}{\sum_{x_{i,j} \in D_k} [x_{i,k} = x_{i,j}] + a} \quad (4)$$

式(4)中: $x_{i,k}$  为第  $k$  个样本的第  $i$  个样本特征; $x_{i,j}$  为第  $k$  个样本之前第  $j$  个样本的第  $i$  个类别特征; $y_j$  为第  $j$  个样本的标签值; $D_k$  为随机序列中在第  $k$  个样本之前的数据集; $p$  为添加的先验项; $a$  为通常大于 0 的权重系数;“=”表示逻辑判断符号,不是赋值或代数运算,它的含义是:样本  $j$  的第  $i$  个特征值是否等于样本  $k$  的第  $i$  个特征值,运算的作用是它的作用是判断括号内的条件是否成立,如果成立,输出 1;否则输出 0。

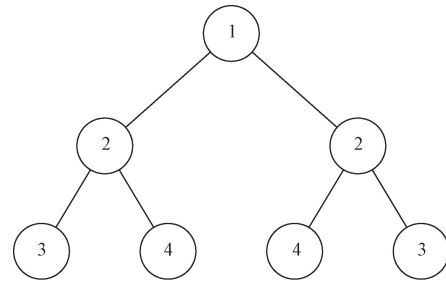


图3 对称二叉树示意图

Fig. 3 Symmetric binary tree diagram

### 2.2 模型性能评价指标

#### 2.2.1 混淆矩阵

混淆矩阵是一种用于评估模型分类精度的误差矩阵,大小为分类类别阶方阵,其中行表示实际类别,列表示预测类别<sup>[29]</sup>。如表 2 所示,模型分类结果通常包含以下 4 种情况。

真正例(true positives, TP):表示模型将正类样本正确地预测为正类的数量。即实际为正例的样本被模型正确预测为正例的数量。

假正例(false positives, FP):表示模型将负类样本错误地预测为正类的数量。即实际为负例的样本被模型错误地预测为正例的数量。

真负例(true negatives, TN):表示模型将负类样本正确地预测为负类的数量。即实际为负例的样本被模型正确预测为负例的数量。

表2 混淆矩阵  
Table 2 Confusion matrix

真实情况	预测为正例	预测为负例
实际为正例	TP (真正例)	FN (假负例)
实际为负例	FP (假正例)	TN (真负例)

假负例 (false negatives, FN): 表示模型将正类样本错误地预测为负类的数量。即实际为正例的样本被模型错误地预测为负例的数量。

通过混淆矩阵, 可以计算出准确率、精确率、召回率和  $F_1$  等精度指标, 其中: 准确率为预测正确样本占总样本的比例; 精确率为预测出来为正样本的结果中, 实际为正样本的比例; 召回率: 实际为正样本的结果中, 预测为正样本的比例;  $F_1$  为精确率和召回率的调和平均。

### 2.2.2 受试者工作曲线。

受试者工作曲线 (receiver operating characteristic, ROC) 是一种用于评估二分类模型性能的工具, 通常用于比较模型在不同阈值下的分类能力。在模型训练过程中, 阈值的调整可以改变模型的输出类别。通过分析 ROC 曲线下的面积 (area under curve, AUC), 可以选择最优阈值, 使模型在正例和负例之间取得平衡, 从而提高预测准确性。ROC 曲线以真正例率 (true positive rate), 即召回率为纵轴, 以假正例率 (false positive rate, FPR) 为横轴, 展示了在不同阈值下模型的分类表现<sup>[30]</sup>。

在 ROC 曲线中, 横轴表示假正例率 (FPR), 计算公式为

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

式(5)中: FP 为假正例数量; TN 为真负例数量; FPR 为实际为负例但被错误地预测为正例的样本比例。

纵轴表示真正例率 (true positive rate, TPR), 也称为召回率 (recall, R), 计算公式为

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

式(6)中: TP 为真正例数量; FN 为假负例数量; TPR 为实际为正例且被正确地预测为正例的样本比例。

ROC 曲线为由不同分类阈值下的 TPR 和 FPR 连成的曲线。ROC 曲线下方的面积被称为 AUC, 通常用来评估分类模型的整体性能, AUC 越大, 模型性能越好。

### 2.3 基于 CatBoost 的四川木里县森林火灾预测方法

CatBoost 模型预测林火流程的主要步骤包括数据准备、输入数据、数据预处理、模型训练与测试、模型的评估。具体流程如图4所示。

模型训练完成后, 通过测试评估模型性能, 并将测试集的预测概率用于火灾发生概率的制图。模型能够预测每个点的火灾和非火灾类别的概率, 选择火灾类别的概率作为最终的预测值。最后利用反距离加权插值法生成森林火灾发生概率图和火险分级图。

## 3 结果与分析

### 3.1 林火预测模型的构建

为提升模型泛化能力, 建模时将数据按 8:2 的比例划分训练集和验证集, 基于 Python 进行林火预测模型的构建。使用 2010—2020 年未参与建模的 406 个点数据, 用于模型验证。这些数据点覆盖了不同时期和不同季节的火灾发生情况, 确保了数据的多样性和代表性, 验证集能够评估模型在不同年份和环境条件下的稳定性和准确性。该验证集不仅包含了火灾高发期的数据, 也涵盖了火灾较少发生的时间段, 从而使得验证结果更加全面和客观。建模结果表明, 模型在训练集和验证集上准确率为 99.10% 和 91.36%, 表明所构建模型具有合理性。

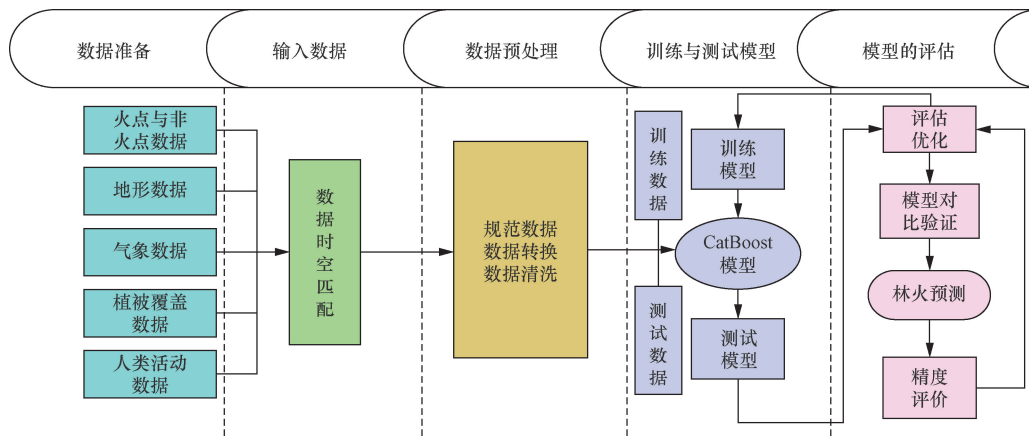


图4 预测模型构建流程

Fig. 4 Construction process of the prediction model

### 3.2 林火预测结果与分析

#### 3.2.1 混淆矩阵精度评价结果

混淆矩阵是评估机器学习模型分类精度的常用方法。所构建 CatBoost 模型的混淆矩阵计算结果如图 5 所示。CatBoost 模型的正确例率和假正例率分别为 91.46% 和 8.74%, 优于 RF 模型(正确例率为 91.41%,假正例率为 13.53%)、XGBoost 模型(正确例率为 89.06%,假正例率为 12.26%)和 GBDT 模型(正确例率为 87.96%,假正例率为 11.68%)。

表 3 展示了混淆矩阵计算得到的精度指标, CatBoost 模型在研究中的表现优于其他对比模型,其准确率达到 91.36%,精确率为 91.00%,召回率为 91.46%, $F_1$  为 91.23%,表明 CatBoost 能够更精确地预测火灾事件并有效减少误报和漏报,体现了其在森林火灾预测中的显著优势。

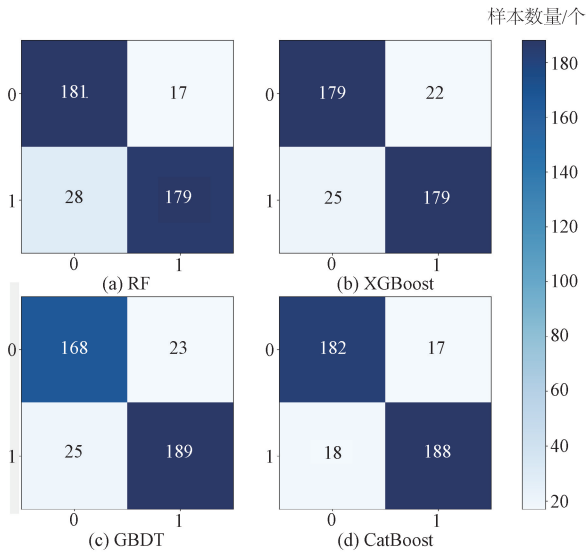


图 5 4 种模型的混淆矩阵结果图

Fig. 5 Confusion matrix results for the four models

表 3 混淆矩阵结果

Table 3 Confusion matrix results

模型	准确率/%	精确率/%	召回率/%	$F_1$ /%
RF	88.89	86.60	91.41	88.94
XGBoost	88.40	87.75	89.05	88.40
GBDT	88.15	87.05	87.96	87.50
CatBoost	91.36	91.00	91.46	91.23

#### 3.2.2 ROC 曲线评价结果

表 4 展示了不同模型的性能指标,可以看出,模型 CatBoost 在准确率方面表现最佳,达到了 91.4% 的准确率,AUC 达到了 0.970,表明其在正负样本的区分能力最强,预测性能最优。其次是随机森林模型,准确率为 88.9%,AUC 为 0.954。

各模型在测试数据集上的 ROC 曲线如图 6 所示。可知,CatBoost 模型在森林火点预测中优于 RF、

表 4 4 种机器学习模型性能对比

Table 4 Performance comparison of 4 machine learning models

模型	TP	TN	FP	FN	FPR/%	TPR/%	AUC
RF	181	179	28	17	13.53	91.41	0.954
XGBoost	179	179	25	22	12.25	89.05	0.950
GBDT	168	189	25	23	11.68	87.96	0.938
CatBoost	182	188	17	18	8.74	91.46	0.970

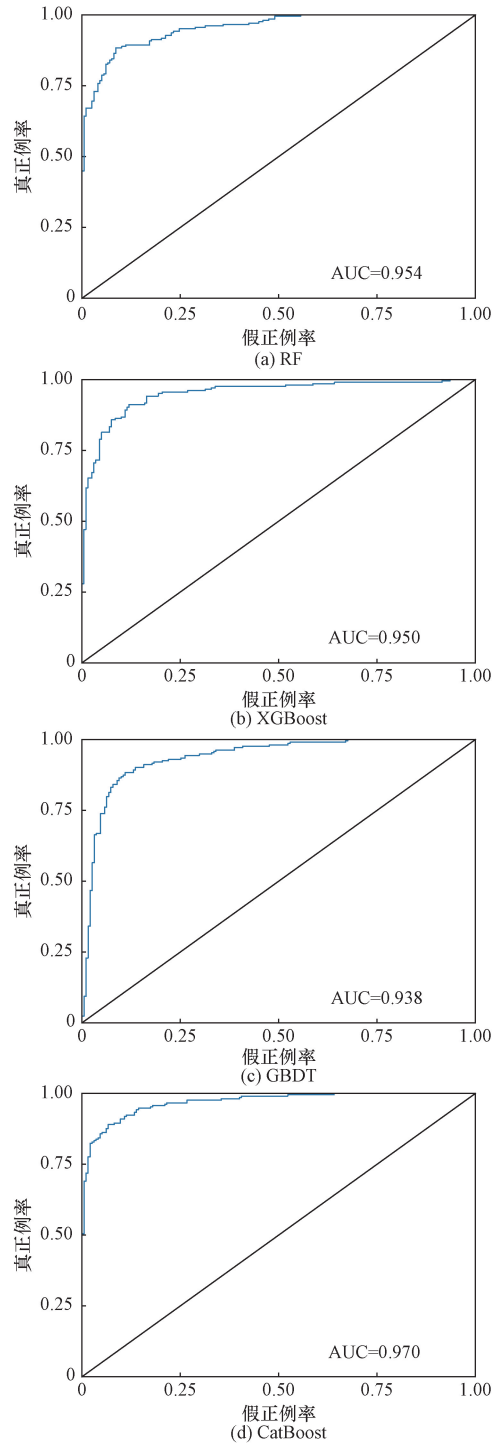


图 6 4 种模型的 ROC 曲线

Fig. 6 ROC curves of four models

XGBoost、GBDT 等模型,4 种模型平均精度为 89.15%,且 AUC 均大于 0.90,说明 4 种预测模型研究得到的结果是可靠的。

### 3.2.3 模型特征重要性

在森林火灾预测模型中,特征重要性是衡量每个输入变量对模型预测结果贡献的指标。通过分析特征重要性,可以识别出对火灾预测影响最大的因素,从而制定相应的防灾减灾策略。如图 7 所示,月均降水对森林火灾预测的影响最大,占特征重要性的 29.70%。降水量直接影响森林的湿度和可燃性,是火灾发生的关键因子。其次是相对湿度和高程,分别为 14.5% 和 10%。

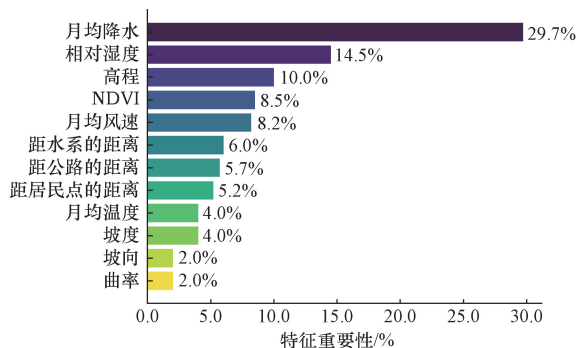


图 7 特征重要性

Fig. 7 Feature importance

### 3.3 森林火灾灾害制图

使用验证集数据并通过克里金插值来制作森林火灾灾害制图(图 8)。根据前人经验,将风险范围区域分为五类,利用发生概率分类,分别为风险指数非常低 $[0, 0.078)$ 、低 $[0.078, 0.234)$ 、中等 $[0.234, 0.441)$ 、高 $[0.441, 0.675)$ 和非常高 $[0.675, 1)$ <sup>[31]</sup>。研究区内东南部发生火灾的概率最大,此处处于峡谷地带,地形复杂,月均降水较少,植被覆盖率高,山体坡度和坡向变化大。其次是西南部和中部,该区域植被覆盖相对较少,距离道路相对较远。

## 4 结论

以木里县 2010—2020 年卫星火点数据、地形因子、气象因子、植被覆盖因子和人类活动因子等为数据源,基于 CatBoost 机器学习算法构建了一种适宜于木里县森林火灾的预测模型。由森林火灾预测结果及火灾灾害制图得出如下结论。

(1) 相较于 RF、XGBoost 和 GBDT 模型,基于 CatBoost 的林火发生概率预测模型取得了较高的拟合度和精度,火灾预测准确率达到 91.36%。这是由于 CatBoost 能够直接处理类别型特征,无需进行独热编码或者其他特征转换操作,从而减少了特征

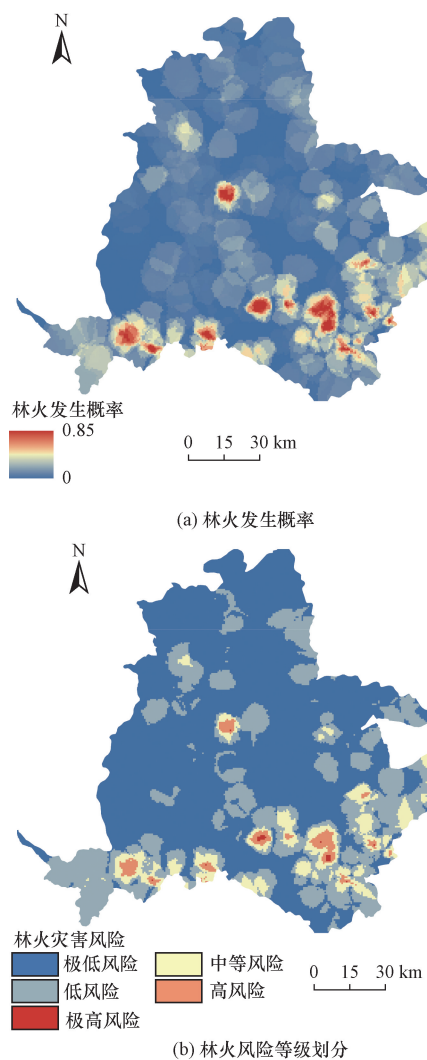


图 8 基于 CatBoost 模型的森林火灾灾害制图

Fig. 8 Forest fire hazard mapping based on CatBoost model

工程的复杂度,并且模型对于异常值和噪声数据具有较强的鲁棒性,能够有效地处理实际数据中的异常情况,提高了模型的可靠性和泛化能力。

(2) 特征重要性分析结果表明降水、相对湿度和高程是研究区森林火灾最重要的 3 个的影响因素,其与 NDVI、风速、距水系和公路的距离等共同作用,形成木里县森林火灾的发生。

(3) 森林火灾灾害制图结果表明,木里县东南部发生森林火灾的概率较大,其次依次是西南部和中部。

### 参 考 文 献

[1] 吴月圆,舒立福,王明玉,等. 近年世界森林火灾综述[J]. 温带林业研究, 2022, 5(4): 49-54.  
Wu Yueyuan, Shu Lifu, Wang Mingyu, et al. A review of forest fires worldwide in recent years[J]. Journal of Temperate Forestry Research, 2022, 5(4): 49-54.

[2] 翟杰休,李勇,张博,等. 世界主要林火多发国家的森林火灾与雷击火概况分析[J]. 亚热带资源与环境学报, 2022, 17

- (4): 72-79.
- Zhai Jiexiu, Li Yong, Zhang Bo, et al. Analysis of forest fires and lightning fires in representative fire-prone countries over the world [J]. *Journal of Subtropical Resources and Environment*, 2022, 17(4): 72-79.
- [3] Sungmin O, Hou X, Orth R. Observational evidence of wildfire-promoting soil moisture anomalies[J]. *Scientific Reports*, 2020, 10(1): 11008.
- [4] Yang J, He H S, Shifley S R, et al. Spatial patterns of modern period human-caused fire occurrence in the Missouri Ozark Highlands [J]. *Forest Science*, 2007, 53(1): 1-15.
- [5] Milanovi S, Milanovi S D, Markovi N, et al. Forest fire probability mapping in eastern Serbia: logistic regression versus random forest method[J]. *Forests*, 2021, 12(1): 1-17.
- [6] Prapas I, Kondylatos S, Papoutsis I, et al. Deep learning methods for daily wildfire danger forecasting[J]. *ArXiv*, 2021: 2111.02736.
- [7] 汪祖民, 王恺锋, 李艳志, 等. 基于 LightGBM 和 SHAP 的云南省森林火灾预测研究[J]. *消防科学与技术*, 2023, 42(11): 1567-1571.
- Wang Zumin, Wang Kaifeng, Li Yanzhi, et al. Research on forest fire prediction in Yunnan Province based on LightGBM and SHAP [J]. *Fire Science and Technology*, 2023, 42(11): 1567-1571.
- [8] 张运林, 田玲玲, 丁波, 等. 贵州省林火发生驱动因子及预测模型[J]. *生态学杂志*, 2024, 43(1): 282-289.
- Zhang Yunlin, Tian Lingling, Ding Bo, et al. Driving factors and prediction model of forest fire in Guizhou Province [J]. *Chinese Journal of Ecology*, 2024, 43(1): 282-289.
- [9] 李史欣, 张福全, 林海峰. 基于机器学习算法的森林火灾风险评估研究[J]. *南京林业大学学报(自然科学版)*, 2023, 47(5): 49-56.
- Li Shixin, Zhang Fuquan, Lin Haifeng. Research on forest fire risk evaluation based on machine learning algorithm[J]. *Journal of Nanjing Forestry University (Natural Science Edition)*, 2023, 47(5): 49-56.
- [10] 张全文, 杨永崇, 王涛, 等. 基于元胞自动机的高原林火蔓延三维可视化模拟[J]. *科学技术与工程*. 2021, 21(4): 1295-1299.
- Zhang Quanwen, Yang Yongchong, Wang Tao, et al. Three-dimensional visual simulation of forest fire spread based on cellular automata [J]. *Science Technology and Engineering*, 2021, 21(4): 1295-1299.
- [11] 苗新, 王倚天, 刘爽. 机器学习在森林火灾预测方面的应用[J]. *信息与电脑(理论版)*, 2022, 34(7): 123-125.
- Miao Xin, Wang Yitian, Liu Shuang. Application of machine learning in forest fire prediction [J]. *Information & Computer*, 2022, 34(7): 123-125.
- [12] 郗婕, 傅微. 基于机器学习的流域尺度森林火灾灾害风险预测[J]. *自然灾害学报*, 2024, 33(1): 89-98.
- Xi Jie, Fu Wei. Watershed-scale forest fire risk prediction based on machine learning [J]. *Journal of Natural Disasters*, 2024, 33(1): 89-98.
- [13] 罗永明, 曾行吉, 谢映, 等. 基于观测与预报数据融合的森林火灾险预报[J]. *科学技术与工程*, 2024, 24(23): 9804-9810.
- Luo Yongming, Zeng Xingji, Xie Ying, et al. Forest fire risk forecast based on the fusion of observation forecast data [J]. *Science Technology and Engineering*, 2024, 24(23): 9804-9810.
- [14] Preethi T. K S B A. Forest fire prediction using machine learning techniques [C]//IEEE International Conference on Intelligent Technologies (CONIT). Hubli, India: IEEE, 2021: 1-6.
- [15] Ma W, Feng Z, Cheng Z, et al. Identifying forest fire driving factors and related impacts in china using random forest algorithm [J]. *Forests*, 2020, 11(5): 507.
- [16] Wu Z, Li M, Wang B, et al. Using artificial intelligence to estimate the probability of forest fires in heilongjiang, northeast China [J]. *Remote Sensing*, 2021, 13(9): 1813.
- [17] 符鑫隆, 林姗, 牛辉, 等. 基于 CatBoost 的患者住院优先级预测模型[J]. *信息化研究*, 2023, 49(1): 43-47.
- Fu Xinlong, Lin Shan, Niu Hui, et al. Hospitalization priority prediction model for patients based on CatBoost [J]. *Information Research*, 2023, 49(1): 43-47.
- [18] 谭勇, 陈记, 杨忠民, 等. 基于 CatBoost 集成学习的边坡稳定性预测方法 [J]. *科学技术与工程*, 2024, 24(30): 13153-13160.
- Tan Yong, Chen Ji, Yang Zhongmin, et al. Slope stability prediction method based on CatBoost ensemble learning [J]. *Science Technology and Engineering*, 2024, 24(30): 13153-13160.
- [19] 张洪飞, 杜宁, 王莉, 等. 基于 Catboost 模型的广东省近地面 NO<sub>2</sub> 浓度估算[J]. *环境科学*, 2024, 45(11): 6276-6285.
- Zhang Hongfei, Du Ning, Wang Li, et al. Estimation of near-surface NO<sub>2</sub> concentration in Guangdong Province based on the CatBoost model [J]. *Environmental Science*, 2024, 45(11): 6276-6285.
- [20] 王强, 陈浩, 刘炼. 基于多层 CatBoost 的电力系统暂态稳定评估[J]. *科学技术与工程*, 2022, 22(4): 1456-1464.
- Wang Qiang, Chen Hao, Liu Lian. Transient stability assessment of power system based on multi-layer CatBoost [J]. *Science Technology and Engineering*, 2022, 22(4): 1456-1464.
- [21] 程楠楠. 基于混合特征选择模型 CatBoost-LightGBM 的违约风险预测研究[J]. *现代信息科技*, 2021, 5(14): 116-120.
- Cheng Nannan. Default risk prediction research based on hybrid feature selection model Catboost-LightGBM [J]. *Modern Information Technology*, 2021, 5(14): 116-120.
- [22] 李顺, 吴志伟, 梁宇, 等. 大兴安岭林火发生的时空聚集性特征[J]. *生态学杂志*, 2017, 36(1): 198-204.
- Li Shun, Wu Zhiwei, Liang Yu, et al. The temporal and spatial clustering characteristics of forest fires in the Great Xing'an Mountains [J]. *Chinese Journal of Ecology*, 2017, 36(1): 198-204.
- [23] 朱贺, 张珍, 杨焱, 等. 中国南北方林火时空分布及火险期动态变化特征——以黑龙江省和江西省为例[J]. *生态学杂志*, 2023, 42(1): 198-207.
- Zhu He, Zhang Zhen, Yang Song, et al. Temporal and spatial distribution of forest fire and the dynamics of fire danger period in southern and northern China: a case study in Heilongjiang and Jiangxi Provinces [J]. *Chinese Journal of Ecology*, 2023, 42(1): 198-207.
- [24] Eugenio F C, Dos Santos A R, Fiedler N C, et al. Applying GIS to develop a model for forest fire risk: a case study in Espirito Santo, Brazil [J]. *Journal of Environmental Management*, 2016, 173: 65-71.
- [25] Setiawan I, Mahmud A R, Mansor S, et al. GIS-grid-based and multi-criteria analysis for identifying and mapping peat swamp for-

- est fire hazard in Pahang, Malaysia[J]. *Disaster Prevention and Management*, 2004, 13(5): 379-386.
- [26] Mcdonnell R A L C, Burrough P. *Principles of geographical information systems*[M]. London: Oxford University Press, 2015.
- [27] Abedi G H. Using GIS to develop a model for forest fire risk mapping[J]. *Journal of the Indian Society of Remote Sensing*, 2019, 47(7): 1173-1185.
- [28] 杨怀珍, 张静, 李雷. 基于多重相似度和 CatBoost 的个性化推荐[J]. *计算机工程与设计*, 2023, 44(9): 2687-2693.  
Yang Huaizhen, Zhang Jing, Li Lei. Personalized recommendation based on multiple similarity and CatBoost[J]. *Computer Engineering and Design*, 2023, 44(9): 2687-2693.
- [29] 豆红强, 黄思懿, 简文彬, 等. 基于遥感数据的闽东南山区公路滑坡快速识别技术研究[J]. *自然灾害学报*, 2023, 32(1): 217-227.  
Dou Hongqiang, Huang Siyi, Jian Wenbin, et al. Research on rapid identification technology of highway landslide in mountainous areas of southeast Fujian based on remote sensing data[J]. *Journal of Natural Disasters*, 2023, 32(1): 217-227.
- [30] 田述军, 张珊珊, 唐青松, 等. 基于不同评价单元的滑坡易发性评价对比研究[J]. *自然灾害学报*, 2019, 28(6): 137-145.  
Tian Shujun, Zhang Shanshan, Tang Qingsong, et al. Comparative study of landslide susceptibility assessment based on different evaluation units[J]. *Journal of Natural Disasters*, 2019, 28(6): 137-145.
- [31] 安佳怡, 冯仲科, 马天天, 等. 基于 GIS 格网的重庆合川区森林火险等级区划[J]. *中南林业科技大学学报*, 2022, 42(9): 91-101.  
An Jiayi, Feng Zhongke, Ma Tiantian, et al. Zoning of forest fire risk levels in the Hechuan District of Chongqing based on GIS grid [J]. *Journal of Central South University of Forestry & Technology*, 2022, 42(9): 91-101.