

基于 XGBoost 的输电线路火灾风险 等级评估方法

罗昊^{1,2}, 杨琪^{1,2}, 王炜宇³, 刘勇辉^{1,2}, 肖珍珍^{1,2}, 杨东济^{1,2}

(1. 国网湖南省电力有限公司智能带电作业技术及装备(机器人)湖南省重点实验室, 湖南 衡阳 420100; 2. 国网湖南省电力有限公司带电
巡检与智能作业技术国家电网公司实验室, 湖南 衡阳 420100; 3. 长沙理工大学电气与信息工程学院, 湖南 长沙 410114)

摘要: 山林火灾是威胁输电线路安全运行的威胁之一。首先, 回顾湖南省 2018—2020 年的历史山火事件发生规律, 在此基础上构建山火事件样本库, 引入独热编码技术对线路植被类型等文本类特征进行数值化处理; 然后, 利用 XGBoost 技术构建输电线路山火风险评估模型, 针对山火样本和正常运行样本的不平衡问题, 基于代价敏感机制提出加权目标函数, 减小样本不平衡导致的山火漏判问题; 最后, 通过湖南省永州市 2020—2021 年的山火事件进行测试, 验证本文所提出的输电线路山火风险评估模型的有效性。

关键词: 山火风险评估; 样本不平衡; 代价敏感机制; 独热编码; 极端梯度提升树

DOI: 10.19781/j.issn.1673-9140.2023.06.014 中图分类号: TM863 文章编号: 1673-9140(2023)06-0132-10

XGBoost-based assessment method for fire risk levels of transmission lines

LUO Hao^{1,2}, YANG Qi^{1,2}, WANG Weiyu³, LIU Yonghui^{1,2}, XIAO Zhenzhen^{1,2}, YANG Dongchai^{1,2}

(1. Intelligent Live Operation Technology and Equipment (Robot) Hunan Provincial Key Laboratory, State Grid Hunan Electric Power Co., Ltd., Hengyang 420100, China; 2. Live Inspection and Intelligent Operation Technology State Grid Corporation Laboratory, State Grid Hunan Electric Power Co., Ltd., Hengyang 420100, China; 3. School of Electrical & Information Engineering, Changsha University of Science & Technology, Changsha 410114, China)

Abstract: Mountain fires pose a threat to the safe operation of transmission lines. Firstly, the historical pattern of mountain fire incidents in Hunan province from 2018 to 2020 is reviewed. On this basis, a database of mountain fire incidents is constructed. One-hot encoding technique is introduced to numerically process textual features such as the vegetation type along the transmission lines. Then, XGBoost technology is utilized to build a mountain fire risk assessment model for transmission lines. Aiming at imbalanced samples between forest fire incidents and normal operations, based on a cost-sensitive mechanism, a weighted objective function is presented to mitigate the problem of overlooking mountain fires caused by sample imbalances. Finally, the proposed mountain fire risk assessment model for transmission lines is tested using the forest fire incidents in Yongzhou City, Hunan Province, from 2020 to 2021 to validate its effectiveness.

Key words: forest fire risk assessment; sample imbalance; cost-sensitive mechanism; one-hot encoding; XGBoost

收稿日期: 2021-07-30; 修回日期: 2021-11-05

基金项目: 湖南省电力公司科技项目(B116AJ210015)

通信作者: 罗昊(1992—), 男, 硕士, 工程师, 主要从事电力设备状态评价与检测方面的研究; E-mail: lhlh33@qq.com

2020年中国提出了“2030碳达峰、2060碳中和”的伟大目标,安全可靠的电力系统成为实现这一目标的重要保证^[1]。中国输电线路通常输电距离较长、跨越地形复杂,各种自然、人为因素对输电线路的安全运行造成威胁,其中山林火灾是最为常见的威胁之一^[2-3]。由于山林地区植被覆盖率高、地形复杂,在每年的重要节日和农耕时节,如春节、春耕和清明祭祀等,容易出现人为引发的山林火灾,火灾一旦蔓延至输电线路附近,可能会降低输电线路的绝缘水平,引发输电线路跳闸等问题^[3],而随着全球气候的不断变暖,中国大部分地区山火发生频率显著增加。因此,对输电线路进行山火风险评估,加大对山火活跃地区输电线路的防护对电力系统的安全可靠运行至关重要。

目前,对于输电线路山火防治工作主要可分为2个方面:山火实时监控和山火风险评估。山火实时监控方面的研究目前较多,这类研究通常利用视频监控、气象雷达等手段,在山火发生刚起燃时快速定位火源位置。其中,视频监控主要利用视频中火焰图像的色彩信息、山火面积等多种特征因素,实现山火辨识^[4-6]。卫星遥感能够监控更为广阔的地表火源情况,电网输变电设备防灾减灾国家重点实验室提出了一系列报警方法,如:火点识别算法^[7]、火点定位算法^[8]及山火预报告警方法^[9]等,有效提高了山火的预测和实时辨识。然而,以上方法由于数据传输的延迟、卫星过境时间间隔长以及扫描覆盖有盲区等问题^[10],可能造成山火预警的滞后甚至漏报。

由于山林地区地形复杂,山火发生时即使能够准确定位火源位置,运维消防人员也难以及时达到并进行灭火处理,因此,对各输电线路的山火风险等级进行评估,从而有针对性地增强高风险线路的山火防护工作,能够有效降低输电线路山火发生的概率和跳闸风险。为了准确评估输电线路的山火风险等级,基于已有的历史数据、专家经验等构建评估模型,对各输电线路进行火灾风险打分。文献^[11-12]基于偏最小二乘(partial least squares, PLS)技术和图模型的风险预警模型,通过对历史气象、植被和地形数据进行分析,筛选出关键的山火特征并构建火灾风险预警模型;文献^[13]提出了基于层

次分析法的输电线路评估模型,根据输电线路历史山火记录、实时气象信息和电网专家经验,构建了层次分析模型,对输电线路山火风险进行打分;文献^[14]构建了输电线路防灾减灾预警分析系统,该系统能够将输电业务、气象监测以及地理信息的交叉分析等多维信息进行收集和分析,对多种安全隐患和自然灾害提前预警,通过灾情分析模块对各种安全隐患和灾害进行提前预警,确保输电线路的安全运行;文献^[15]则针对输电线路在极端灾害天气下的故障作用机理、故障概率模型进行了更为详细的研究,并基于支持向量机和AdaBoost技术构建了双层分类器,从而对极端天气下的输电线路故障进行预警;文献^[16]基于朴素贝叶斯网络的输电走廊山火风险评估模型,首先通过Relief算法对重要致火因子进行筛选,然后根据筛选后的特征构建朴素贝叶斯网络进行山火风险计算;文献^[17]则考虑了发生山火时空气间隙被击穿的概率计算模型,结合火险气象等级构建了山火输电通道风险评估模型。然而,以上这些算法仍存在一些缺陷,如:依赖于专家经验等主观判断、无法处理样本不平衡问题等,导致预警模型的可靠性受到影响。

针对以上问题,本文基于极端梯度提升树(extreme gradient boosting tree, XGBoost)技术构建一套输电线路山火风险识别算法。首先,分析湖南省2018—2020年的历史山火事件发生规律,构建山火事件样本库,引入独热编码(one-hot encoding)技术对线路植被类型等文本类数据进行数值离散化处理;接着,利用XGBoost技术训练输电线路山火风险评估模型,利用代价敏感机制减小样本不平衡导致的山火漏判问题;最后,通过湖南省永州市2020—2021年的山火事件进行测试,验证本文所提出的输电线路山火风险评估模型的有效性。

1 湖南省历史输电线路山火事件

湖南省地形地貌主要为山地、丘陵、盆地以及平原,其中山地占比达51.22%,因此,省输电线路穿过的大部分地区为树木繁茂的山地地形。据统计,湖南省内自2018年1月至2021年2月期间,500 kV线路山火共发生776次,成为继雷击之后导致跳闸

的第二大原因。当雷击造成输电线路跳闸后,通过线路自动重合闸装置能够恢复送电。然而山火的持续燃烧会造成空气分子热游离并使带电粒子上升,导致输电线路绝缘强度降低,重合闸动作后输电线路会再次跳闸。

本文首先从地区、时间和线路植被类型3个不同角度对2018至2021年2月累计的山火事件进行初步分析,历史山火统计数据如图1~3所示。

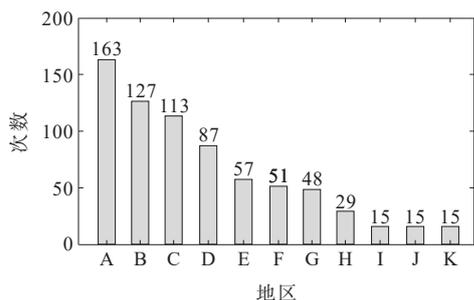


图1 各地区历史山火情况

Figure 1 Historical mountain fire incidents in different areas

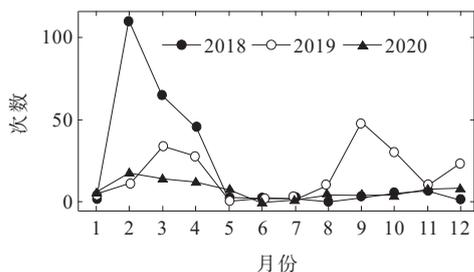


图2 各月份历史山火情况

Figure 2 Historical mountain fire incidents in different months

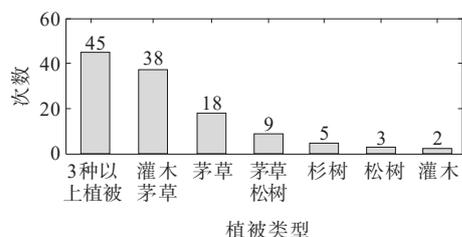


图3 输电线路山火植被类型

Figure 3 Types of plants in mountain fire region under the transmission lines

图1中,山火频发的A~E地区为植被茂盛的山区,从2018年至今,这些地区由于山火导致输电线路跳闸的事件频繁发生,其中A、B、C地区分别跳闸27、17、16条次,这3个地区跳闸总和占山火跳闸总数的47.61%。

由图2可知,湖南电网的山火发生时间具有较强的季节性特点,1—4月、9—12月湖南气候干燥,植被干枯易燃,且此段时间正值烧荒、春节庆典及祭祖等时期,由于人们不当用火行为,极易引发山火,导致输电线路跳闸。

此外,山火发生的难易也与输电线路通道内的植被相关。由图3可知,最容易导致山火跳闸的4种植被类型为3种以上植被混合、灌木与茅草混合、茅草以及茅草与松树混合。其中,茅草与山火有着较为紧密的关系,在冬季和春季为干枯状态,极易引燃,当通道内具有其他树木植被,可能进一步引发树冠火而导致输电线路跳闸。

通过对以上初步分析发现,湖南省输电线路山火发生的条件主要有火险天气、可燃物和火源。湖南特殊的气候特征和地理环境在一定条件下为冬春天气干燥时节提供了火险天气和干燥的可燃物,而输电线路附近居民的生产、生活习惯和民风民俗为山火发生提供了火源。

2 输电线路山火样本库构建

基于湖南省山火初步分析,本文将以历史天气、输电线路通道内植被类型以及人为因素为核心构建山火样本库,为下一步输电线路山火风险评估模型奠定基础。

2.1 关键天气特征筛选

山火样本主要由历史天气、输电线路通道内的植被类型、人为因素以及是否发生山火构成,其中历史天气数据包含特征较多,如:最高温度、最低温度、平均温度、湿度、风速、风级、风向角度、气压、能见度以及日降雨量等特征。这些特征属于一次特征,即原始的气象数据,在此基础上,进一步构建二次特征。本文增加持续干旱天数作为二次特征,即上一次雨天到当天所持续的天数,即

$$N_{dry,i} = N_{date,i} - N_{rain,date,i} \quad (1)$$

式中, $N_{dry,i}$ 为第*i*天持续干旱天数; $N_{date,i}$ 为第*i*天具体日期; $N_{rain,date,i}$ 为上一次雨天的日期。

值得注意的是,山火数据特征与其他灾害量测数据相比,其特征数据种类更为广泛,如输电线路冰灾预警中主要考虑气象因素和地理因素,对植被

和人为因素考虑较少。由于山火样本中含有多种特征数据,各特征数据之间可能存在内部的联系,导致评估模型在训练时出现过拟合现象,因此有必要对关键样本特征进行筛选降维。

特征降维的常用方式是通过计算2个特征之间的相关性判断其关联程度,若相关度的绝对值大于某一阈值,则认为这2个特征强相关,可以任意去除1个特征,从而达到特征降维的目的。本文利用spearman相关系数计算特征之间的相关程度,其核心思想^[18]是,将2个特征*a*、*b*的所有样本数据按照大小重新排序,记*a*'_{*i*}、*b*'_{*i*}为排序后的位置,定义秩次差*d*_{*i*}=*a*'_{*i*}-*b*'_{*i*},则特征相关性计算式为

$$\rho_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{n(n^2 - 1)} \quad (2)$$

式中, ρ_s 为特征之间的相关系数;*n*为样本数量。

2.2 基于独热编码的植被类型表征

输电线路通道内的植被类型属于字符类信息,需要将其转化为数值类信息,从而与天气数据一起构成训练样本的特征。然而如果简单地用数字编号代替不同植被类型,由于数字编号之间天然的大小关系,可能会使机器学习算法错误地利用这种序列关系进行风险评估模型训练,人为地影响各特征之间重要程度关系,降低风险评估模型的准确性。

本文利用独热编码技术对植被类型进行表征。独热编码使用*m*个二元状态位对某特征的所有类型进行表征^[19],每个状态位只有0或1这2种选择,若某特征含有某类型,则对应状态位置1,否则置0。举例而言,根据图3中的重要植被类型,可将其归类为茅草、灌木及树木3类,若某线路通道内主要为茅草和灌木混合的情况,则用“[110]”表示,其他植被类型可以此类推。独热编码特别适用于通道内多种植被混合的情况。

2.3 输电线路山火样本库构建

人为原因是诱发山火的因素之一,虽然居民的野外用火行为往往发生在烧荒、春节庆典及祭祖等时期,但是其发生还是具有较强的随机性,难以定量描述,因此,本文构建样本库时主要考虑天气和线路通道内植被因素。样本的时间段则选择为居民野外用火的高发时期,即1—4月、8—12月。气象

数据以天为单位,因此,每个样本为每天天气数据和每段线路植被类型构建。其中,将山火样本作为正样本,未发生山火的各天作为负样本,则样本集为

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (3)$$

式中, x_i 为第*i*个样本的特征向量,具体为 $x_i = [x_{1i}, x_{2i}, \dots, x_{Ni}, x_{(N+1)i}, \dots, x_{Mi}]$,每个样本含有*M*个特征,前*N*个特征为天气特征,第*N*+1到第*M*个特征为植被特征和线路对地距离, $x_i \in \mathbf{R}^M$; y_i 代表第*i*个样本是否发生山火,发生则 $y_i = 1$,反之则 $y_i = 0$, $y_i \in \mathbf{R}$ 。

3 基于XGBoost的评估模型训练

3.1 XGBoost算法简介

本文将基于XGBoost技术提出输电线路山火风险评估模型。XGBoost属于集成学习方法^[20],通过将串行生成的多个弱决策树集成为一个强决策树,实现预测模型的构建。令第*i*个样本 x_i 的预测值为

$$y_i' = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (4)$$

式中, y_i' 为XGBoost输出的预测值; f_k 为第*k*棵决策树;*K*为决策树总数;*F*为所有决策树的集合。

XGBoost对样本 x_i 的最终预测结果是将所有*K*棵树的叶子节点权重相加而得到的。令XGBoost的目标函数为

$$\text{Obj. } f = \sum_{i=1}^n l(y_i', y_i) + \sum_{k=1}^K \Omega(f_k) \quad (5)$$

式中,*f*为目标函数, $l(y_i', y_i)$ 为预测值与真实值之间的损失函数; Ω 为第*k*棵决策树的正则化项,该项可避免训练出的模型过拟合,具体定义为

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (6)$$

式中,*T*为叶子节点数目; γ 、 λ 为超参数,调节 γ 和 λ 可改变决策树的结构复杂度和各叶子权重系数; w_j 为叶子节点*j*的权重。结构复杂的决策树模型往往泛化能力较差。

XGBoost是通过叠加式学习进行训练,从而使式(5)的目标函数达到最小,即第*k*棵树的训练是在第*k*-1棵树的基础上进行的。由式(4)可得第*k*棵

树的预测值为

$$y_i^{(k)} = y_i^{(k-1)} + f_k(x_i) \quad (7)$$

将第 k 棵树的目标函数改写为

$$f^{(k)} = \sum_{i=1}^n l(y_i^{(k-1)} + f_k(x_i), y_i) + \Omega_c + \Omega(f_k) \quad (8)$$

式中, Ω_c 为前 $k-1$ 棵树的正则化项之和, 由于前 $k-1$ 棵树已经训练完毕, 因此 Ω_c 为常数。

为了使目标函数快速达到最小值, 对式(8)进行泰勒展开, 这里认为 f_k 是偏差量, 在 $y_i^{(k-1)}$ 处的展开结果为

$$f^{(k)} \approx \sum_{i=1}^n [l(y_i^{(k-1)}, y_i) + g_i \cdot f_k(x_i) + \frac{1}{2} h_i \cdot f_k^2(x_i)] + \Omega_c + \Omega(f_k) \quad (9)$$

式中, g_i 和 h_i 为 $l(y_i^{(k-1)}, y_i)$ 对 $y_i^{(k-1)}$ 的一阶和二阶导数, 即

$$\begin{cases} g_i = \partial_{y_i^{(k-1)}} l(y_i^{(k-1)}, y_i) \\ h_i = \partial_{y_i^{(k-1)}}^2 l(y_i^{(k-1)}, y_i) \end{cases} \quad (10)$$

式中, $y_i^{(k-1)}$ 为第 $k-1$ 棵树的预测值, 忽略式(9)中的所有常数项后, 简化的目标函数可写为

$$f^{(k)} = \sum_{i=1}^n [g_i \cdot f_k(x_i) + \frac{1}{2} h_i \cdot f_k^2(x_i)] + \Omega(f_k) \quad (11)$$

为简化计算, 定义叶子 j 上的样本集合为 $I_j = \{i | q(x_i) = j\}$, q 为决策树结构, $q(x_i)$ 为样本 x_i 的叶子节点编号。式(11)可进一步被转换为

$$f^{(k)} = \sum_{j=1}^T [\omega_j \sum_{i \in I_j} g_i + \frac{1}{2} \omega_j^2 (\sum_{i \in I_j} h_i + \lambda)] + \gamma T \quad (12)$$

对于特定的树结构 q , 其目标函数最小值可由式(12)对 ω_j 求导得到, 对应的解为

$$\omega_j = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} (h_i + \lambda)} \quad (13)$$

则目标函数最小值为

$$f^{(k)} = - \frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (14)$$

在构建树的结构时, 常采用贪心算法计算树的最佳结构。从根节点开始, 令 I_L 和 I_R 分别为分裂之后的左、右节点集, I 为左、右节点集的并集, 选定某一特征作为测试节点, 分裂之后的增益为

$$G = \frac{1}{2} \left[\frac{(\sum_{i \in I_k} g_i)^2}{\sum_{i \in I_k} h_i + \lambda} + \frac{(\sum_{i \in I_c} g_i)^2}{\sum_{i \in I_c} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (15)$$

当某特征对应的增益 G 最大时, 说明用该特征进行样本划分后目标函数的降低程度最大, 因此选该特征作为划分节点, 进一步可计算下一层最佳划分节点, 以此类推。在此基础上, 各特征的重要性可通过其在所有树中的增益 G 均值来评估:

$$F_m = \frac{1}{KZ} \sum_{k=1}^K \sum_{z=1}^Z G_{mk}^z \quad (16)$$

式中, F_m 为特征 m 的重要性; G_{mk}^z 为特征 m 在第 k 棵树中第 z 次作为分裂节点的增益值; Z 为该特征在第 k 棵树中选作分裂节点的次数。

得到各特征的绝对重要性后可进一步计算相对重要性:

$$f_m = \frac{F_m}{\sum_{m=1}^M F_m} \quad (17)$$

3.2 代价敏感机制

实际电网在运行过程中, 由于可靠性的要求电网绝大多数时间都是正常运行的, 发生火灾的情况极少, 因此在所构建的样本库中, 山火样本相对正常样本而言较少, 这种样本的不平衡性对于风险评估模型的训练是有较大影响的。目前解决样本分类失衡的主要方法有欠采样、过采样和代价敏感机制, 前两者通过调整样本的分布和抽样方法解决样本的不平衡问题, 但是会引入额外的噪声数据, 而本文选择采用代价敏感机制^[21]。

对于输电线路山火风险评估, 不仅要提高评估的准确性, 还需要尽可能地降低错误评估的代价, 本文提出“代价矩阵”, 如表1所示。

表1 代价矩阵

Table 1 Cost matrix

实际类别	预测类别	
	山火	正常
山火	$\cos t_{++}$	$\cos t_{+-}$
正常	$\cos t_{-+}$	$\cos t_{--}$

将山火样本判作正常样本属于漏判, 其代价为 $\cos t_{+-}$; 将正常样本判作山火样本为误判, 其代价为

$\cos t_{-+}$; 定义 $\cos t_{++}$ 和 $\cos t_{--}$ 为正确预测, 显然预测正确不需要设置惩罚代价, 即 $\cos t_{++} = \cos t_{--} = 0$ 。在错误的预测结果中, 误判仅会增加输电线路检修人员排查的工作量, 而漏判则会使高风险山火线路被忽视, 一旦山火发生, 可能造成较为严重线路跳闸事故, 因此漏判的严重性要远远大于误判。所以需设置代价 $\cos t_{+-} \gg \cos t_{-+}$ (一般取 $\cos t_{-+} = 1$)。

用 N_- 、 N_+ 分别表示正常样本和山火样本的数量, N 为样本总数, 各样本初始权重值为

$$\beta_- = N \frac{\cos t_{-+}}{\cos t_{-+} N_- + \cos t_{+-} N_+} \quad (18)$$

$$\beta_+ = N \frac{\cos t_{+-}}{\cos t_{-+} N_- + \cos t_{+-} N_+} \quad (19)$$

在此基础上, 将式(5)定义的目标函数修改为

$$f = \sum_{i=1}^n \beta_i \ell(y'_i, y_i) + \sum_{k=1}^K \Omega(f_k) \quad (20)$$

其中, β_i 根据样本 i 的类别选择 β_- 或 β_+ 。因此, 最后本文使用式(20)所示的带有样本不平衡权重的目标函数对模型进行训练。

虽然本文使用代价敏感机制以降低样本不平衡带来的训练弊端, 但是山火样本较少仍是一个现实存在的问题, 因此, 在样本数据收集时, 应尽可能收集较长时间尺度的样本。为了对山火风险进行准确评估, 收集的样本数据至少应涵盖一年(四季)发生的山火事件, 此外, 还应该考虑样本比例的问题, 如果比例实在过于悬殊, 则应增加样本收集的时间范围, 以增加山火样本数量。对于湖南电网而言, 由于山林地区较多, 山火发生较多, 故其采样时间至少为一年。

3.3 输电线路风险评估打分方法

训练好的 XGBoost 模型属于分类模型, 只能判断火灾是否发生, 而电网运维人员更加关心的是各条输电线路的山火发生风险, 无需精确判断出山火是否会发生。本文将各特征的相对重要性作为对应分数, 当样本 x_i 的特征 m 满足某阈值时, 将该特征的相对重要性进行累加, 即

$$\text{score}(x_i) = \sum_{m=1}^M f_m \cdot t_m \quad (21)$$

其中, t_m 为特征 m 是否满足阈值的标志。 t_m 的计算取决于具体的特征类型, 如对气温而言, 当实时气温超过气温阈值时, $t_m = 1$; 而当实时湿度低于湿度

阈值时, $t_m = 0$ 。本文选取特征的绝对重要性作为特征阈值。

3.4 输电线路山火风险评估模型构建流程

以 XGBoost 技术、独热编码和代价敏感机制为基础, 构建山火风险评估模型, 其流程如图 4 所示。

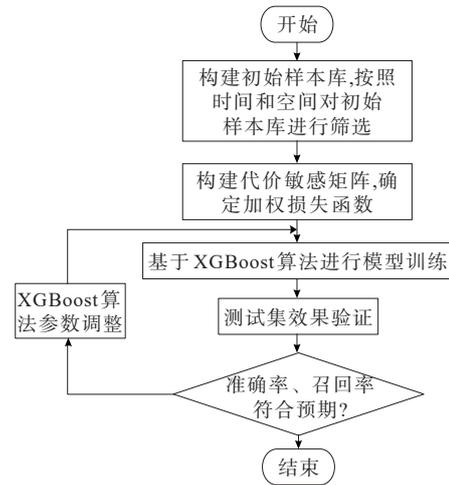


图 4 输电线路山火风险评估模型构建流程

Figure 4 Flow chart of mountain fire risk assessment model construction for transmission lines

1) 样本库构成。

获取所研究时段的历史天气数据和目标区域的线路通道植被数据, 以天为单位构建初始样本库, 样本总数 $n = \text{天数} \times \text{线路档段数}$ 。由于初始样本库中火灾样本数量极少而正常运行样本较多, 因此需要进行样本的筛选。根据山火的历史发生时段可知, 湖南电网山火主要发生在 1—4 月和 9—12 月, 因此初始样本库中剔除 5—8 月的所有样本。同时, 由于输电线路档段数很多, 因此, 选择历史发生过山火的档段以及其相邻的档段作为对象。通过从时间和空间两方面进行筛选, 可在一定程度上降低样本的不平衡情况。

2) 模型训练。

根据样本库中山火样本和正常运行样本的比例, 构建代价敏感矩阵及样本权重 β_i 。将样本库按照 7:3 的比例划分为训练和测试数据, 选取合适的损失函数和目标函数, 基于 XGBoost 算法对训练数据进行学习, 构建山火风险评估模型。

3) 参数调整和风险评估。

利用测试数据对评估模型的性能进行测试, 根据测试结果对模型参数进行调整, 如机器学习器数

量、最大层数、结构复杂度权重以及叶子权重等。对于最终训练好的评估模型,根据实时天气数据和线路通道制备数据,计算各基学习器的输出结果,对所有基学习器输出结果进行累加后即可得出输电线路的山火风险。

4 算例分析

本文以 Python 作为开发语言编写程序,选定湖南省永州市境内 500 kV 输电线路为研究对象,根据当地 2018—2020 年的历史天气数据、线路通道内的植被数据以及历史山火发生数据构建样本库。如前所述,为降低样本的不平衡程度,选取每年的 1—4 月和 9—12 月进行研究。为验证所提出的风险评估模型的泛化能力,选择 2021 年 1 月 1 日—2 月 22 日的山火事件进行验证。

首先构建样本库。永州 2018—2020 年的原始天气数据包含最高温度、最低温度、平均温度、湿度、风速、风级、风向角度、气压、能见度以及日降雨量等 10 个特征,计算各天气特征之间的相关性,其对应的热力图如图 5 所示。本文假设 2 个特征的相关系数绝对值大于 0.9 时为强相关,由图 5 可知,在各天气特征中,最高温度、最低温度、平均温度和气压这四者强相关,风速和风级强相关,本文选择保留最高温度和风速,最终天气特征保留 6 个:最高温度、湿度、风速、降雨量、能见度以及持续干旱天数。

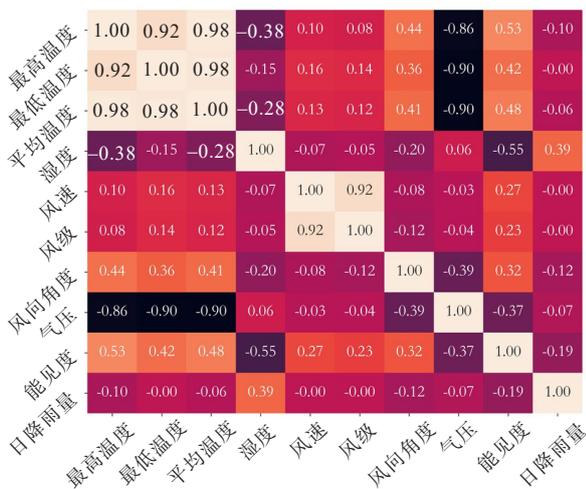


图 5 天气特征相关性热力图

Figure 5 Spearman coefficient heatmap between weather features

接着考虑输电线路通道内的植被类型。永州境内输电线路通道内的植被类型主要有田土、茅草、灌木、竹子、杉树、松树、杨树以及玉米地等,由于植被类型较多,如果令每一个类型都为一个特征,则会大大增加特征数量,导致训练出的模型泛化能力较差。本文将所有植被类型归纳为 4 类:田地、树木、茅草及坟区。根据独热编码将各线路通道内的植被类型转换为二进制数据。最后将历史天气特征、植被类型以及山火标签组合,形成样本库。

在 2018—2020 年中,1—4 月和 9—12 月间样本总数为 76 954 个,其中山火样本共 78 个,正常运行样本达 76 876 个。由于山火样本和正常运行样本之间极不平衡,因此构建代价敏感矩阵,其中,令 $\cos t_{++} = \cos t_{--} = \cos t_{+-} = 1$, 逐次改变 $\cos t_{+-}$, 观察模型的评分,具体如表 2 所示,可知选取 $\cos t_{+-}$ 为 400 和 800 时效果较好。

表 2 不同代价敏感矩阵下模型性能

Table 2 Model performance with different cost sensitive matrices

$\cos t_{+-}$	β_+	β_-	准确率/%	召回率/%
200	166.43	0.83	99.96	66.67
400	284.81	0.71	99.96	70.83
600	373.33	0.62	99.96	66.67
800	442.02	0.55	99.96	70.83
1 000	496.87	0.49	99.96	66.67

基于 XGBoost 技术训练评估模型,本文预设训练 1 000 棵基学习器,最终所得的评估模型为 1 000 棵基学习器的总和。由于基学习器数目较多,这里仅选取第 1 棵展示,如图 6 所示。根据这 1 000 棵基学习器构成的风险评估模型,可计算得到样本特征重要度,如图 7 所示,在所有 10 个特征值中,降雨量具有最大的重要度,其次是草地、坟区等。

根据训练出的风险评估模型,对永州地区 2021 年 1 月 1 日—2 月 22 日的输电线路山火事件进行预测,此段时间永州境内共发生山火 29 起,正常运行样本 8 200 个,所有预测结果的准确性统计如表 3 所示,在 2021 年发生的 29 起山火中,26 起被成功预测,3 起山火事件被漏判,8 200 个正常运行状态全部预测正确,可见所训练的山火风险评估模型能够较为准确地评估输电线路通道内发生山火的可能性。

为了进一步对比本文所提模型的正确性,将本文方法与逻辑回归、C4.5 决策树、支持向量机以及神经网络等算法进行对比,如表 4 所示。

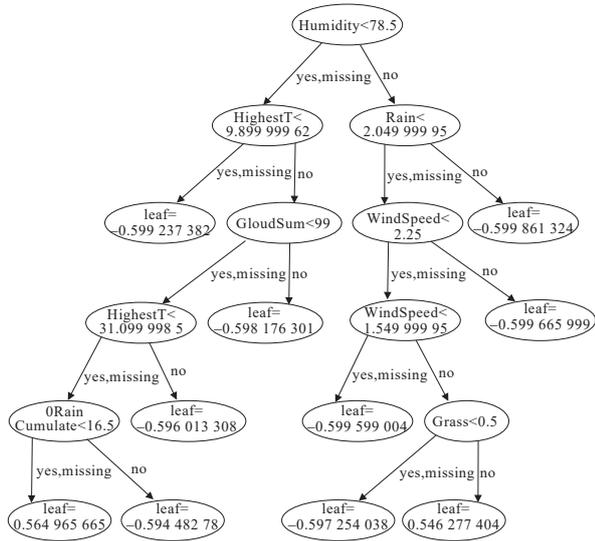


图 6 决策树示例

Figure 6 Example of decision tree

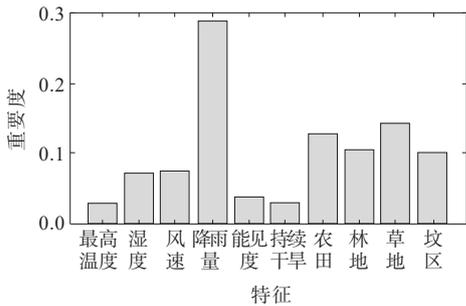


图 7 重要特征辨识结果

Figure 7 Identification results of important features

表 3 统计结果

Table 3 Statistical results

实际类别	预测类别	
	山火	正常
山火	26	3
正常	0	8 200

表 4 不同算法性能对比

Table 4 Performance comparison of different algorithms

算法	漏判	误判	准确率/%	召回率/%
XGBoost	3	1	99.95	89.65
逻辑回归	16	5	99.74	44.83
SVM	13	3	99.81	55.17
BP 神经网络	10	5	99.82	65.52
KNN	25	12	99.50	13.79

由表 4 可见,虽然不同训练算法的准确率都较高,但是除 XGBoost 外,其他算法的召回率皆不超过 70%,这意味着会存在较多的漏判情况,无法准确预判山火的发生与否,无法有效地为输电线路山火防护提供指导。本文所提出的基于 XGBoost 的风险评估模型,能够较为准确地判断出输电线路山火的风险,山火漏判的数量较少,召回率达 89.65%。

进一步,根据式(21)可计算输电线路的山火风险等级。为说明风险等级评估的有效性,选取实际发生山火的线路前一天和山火当天的数据进行计算,由于输电线路众多,此处选取部分结果进行展示(2021 年),具体如表 5 所示,可见在线路火灾前一天,风险等级评估结果不超过 0.6,即火灾风险较低,火灾当天风险指数均超过 0.6,与实际山火情况较为吻合,说明了本评估算法的有效性。

表 5 线路山火风险

Table 5 Mountain fire risk of lines

日期	线路	档段	风险指数	实际情况
0103	长宗线	#256—#257	0.432	无火灾
0114	长宗线	#256—#257	0.667	发生火灾
...
0119	苏紫 I 线	#206—#208	0.537	无火灾
0120	苏紫 I 线	#206—#208	0.773	发生火灾
...
0218	苏紫 I 线	#209—#210	0.494	无火灾
0219	苏紫 I 线	#209—#210	0.639	发生火灾

5 结语

本文基于 XGBoost 技术提出了输电线路山火风险评估算法。针对实际中存在的样本不平衡问题,分别从时间和空间 2 个维度对样本库进行筛选,同时引入代价敏感机制,调整了不同类型样本的惩罚系数,提出了加权目标函数,从样本构建和模型训练 2 个方面削弱了样本失衡对训练模型的不利影响,提高了评估模型的泛化能力;引入了独热编码对线路通道内植被类型进行定量描述,解决了字符类特征的数值化问题;基于 XGBoost 技术构建了山火风险评估模型,有效提高了预测结果的召回率和

准确性,保证对输电线路山火风险的准确评估。本文在如下两方面仍有继续改进的空间。

1) 居民野外用火的定量化描述。由于输电线路周围居民野外用火行为的随机性,以往山火事件只能在事故后追溯火源是否来自于居民,而线路正常运行时则难以得知附近是否有居民野外用火,因此,下一步工作将研究如何在火灾样本和正常运行样本中定量描述居民的野外用火行为。

2) 评估模型召回率仍有提升空间。虽然本文所提的评估模型在一定程度上弥补了样本不平衡导致的召回率较低问题,但是仍存在漏判的可能性,未来将进一步研究如何继续提高评估模型的召回率,尽可能消除漏判的情况。

参考文献:

- [1] 文明,胡资斌,龙乙林,等.考虑碳排放惩罚因子的综合能源系统优化规划[J].电力科学与技术学报,2021,36(3):11-18.
WEN Ming,HU Zibin,LONG Yilin,et al.Optimal planning of integrated energy system considering carbon emission penalty factor[J]. Journal of Electrical Power and Technology,2021,36(3):11-18.
- [2] 黄欢,雷加智,曾华荣,等.极端外部环境下输电线路的综合风险评估方法[J].电力科学与技术学报,2019,34(2):119-127.
HUANG Huan,LEI Jiazhi,ZENG Huarong,et al.Integrated risk assessment system of transmission line under extreme external environment[J]. Journal of Electrical Power and Technology,2019,34(2):119-127.
- [3] 吴田,刘凯,刘庭,等.山火对输电线路设备及外绝缘影响分析[J].电力科学与技术学报,2013,28(3):83-89.
WU Tian,LIU Kai,LIU Ting,et al.Effect of forest fires on transmission line equipments and external insulation[J]. Journal of Electrical Power and Technology, 2013, 28(3): 83-89.
- [4] 朱思思,丁德红,陈朝迎,等.基于图像处理的森林火灾识别方法研究[J].红外技术,2016,38(5):440-446.
ZHU Sisi,DING Dehong,CHEN Chaoying,et al.Research on forest fire recognition method based on image processing[J].Infrared Technology,2016,38(5):440-446.
- [5] 曾思通,吴海彬,沈培辉.基于多特征融合的视频火焰检测方法研究[J].图学学报,2017,38(4):549-557.
ZENG Sitong,WU Haibin,SHEN Peihui.Video fire detection based on fusion of multiple features[J].Journal of Graphics,2017,38(4):549-557.
- [6] 邢晓强,黄新波,纪超,等.基于多特征融合的输电线路山火识别预警系统设计与实现[J].广东电力,2018,31(6):107-113.
XIN Xiaoqiang,HUANG Xingbo,JI Chao,et al.Design and realization of mountain fire identification pre-warning system for transmission lines based on multi-feature fusion [J].Guangdong Electric Power,2018,31(6):107-113.
- [7] 陆佳政,吴传平,杨莉,等.输电线路山火监测预警系统的研究及应用[J].电力系统保护与控制,2014,42(16):89-95.
LU Jiazheng,WU Chuanping,YANG Li,et al.Research and application of forest fire monitor and early-warning system for transmission line[J]. Power System Protection and Control,2014,42(16):89-95.
- [8] 陆佳政,刘毓,徐勋建,等.架空输电线路山火预测预警技术[J].高电压技术,2017,43(1):314-320.
LU Jiazheng,LIU Yu,XU Xunjian,et al.Prediction and early warning technology of wildfire nearby overhead transmission lines[J]. High Voltage Engineering, 2017, 43 (1):314-320.
- [9] 刘毓,陆佳政,罗晶,等.架空输电线路山火同步卫星广域监测与杆塔定位[J].电网技术,2018,42(4):1322-1327.
LIU Yu,LU Jiazheng,LUO Jin,et al.Synchronous satellite wide area monitoring for overhead transmission line wildfire and tower location[J].Power System Technology, 2018,42(4):1322-1327.
- [10] 舒胜文,张深寿,许军,等.基于新一代天气雷达组网监测的输电线路山火自动识别算法研究[J].中国电机工程学报,2020,40(13):4200-4210.
SHU Shengwen,ZHANG Shenshou,XU Jun,et al.Study on automatic identification algorithm of wild fire near transmission lines based on CINRAD-net monitoring[J]. Proceedings of the CSEE,2020,40(13):4200-4210.
- [11] 周恩泽,胡思雨,张录军,等.电网山火灾害特征及风险预警技术[J].电力工程技术,2020,39(3):58-64.
ZHOU Enze,HU Siyu,ZHANG Lujun,et al.Characteristics and risk warning technology of wildfire disaster in power grid[J].Electric Power Engineering Technology,2020,39(3): 58-64.
- [12] 周恩泽,黄勇,陈洁,等.基于图模型的架空输电线路山

- 火风险等级预测模型[J].南方电网技术,2020,14(4):8-16.
- ZHOU Enze, HUANG Yong, CHENG Jie, et al. Graph theory based fire risk prediction level model of overhead transmission lines[J]. Southern Power System Technology, 2020,14(4):8-16.
- [13] 朱奇,郭江,曾兵,等.基于层次分析法的输电线路山火预警评估模型[J].电测与仪表,2018,55(6):71-75+88.
- ZHU Qi, GUO Jiang, ZENG Bing, et al. An early-warning evaluation model against mountain fire of transmission line based on analytic hierarchy process[J]. Electrical Measurement & Instrumentation, 2018,55(6):71-75+88.
- [14] 姜萌萌,佟永吉,陈明丰,等.基于灾害条件的主动配电网快速重构策略研究[J].智慧电力,2021,49(5):85-92.
- JIANG Mengmeng, TONG Yongji, CHEN Mingfeng, et al. Rapid reconfiguration strategy of active distribution network under disaster conditions[J]. Smart Power, 2021,49(5):85-92.
- [15] 黄亚飞,蒋兴良,任晓东,等.采用涡流自热环防止输电线路冰雪灾害的方法研究[J].电工技术学报,2021,36(10):2169-2177.
- HUANG Yafei, JIANG Xingliang, REN Xiaodong, et al. Study on preventing icing disasters of transmission lines by use of eddy self-heating ring[J]. Transactions of China Electrotechnical Society, 2021,36(10):2169-2177.
- [16] 周恩泽,黄勇,龚博,等.基于朴素贝叶斯网络的输电走廊山火风险评估模型[J].南方电网技术,2021,15(8):120-129.
- ZHOU Enze, HUANG Yong, GONG Bo, et al. Wildfire risk assessment model of power transmission line corridors based on naive bayes network[J]. Southern Power System Technology, 2021,15(8):120-129.
- [17] 黄新波,吴明松,朱永灿,等.基于模糊数学的电缆线路风险评估模型研究[J].高压电器,2021,57(9):19-25.
- HUANG Xinbo, WU Mingsong, ZHU Yongcan, et al. Research on risk assessment model of cable line based on fuzzy mathematics[J]. High Voltage Apparatus, 2021,57(9):19-25.
- [18] 单节杉,任敏,田鑫萃,等.基于故障电流回路特性的柔性直流架空线路纵联保护[J].电力系统自动化,2022,46(21):152-159.
- SHAN Jieshan, REN Min, TIAN Xincui, et al. Pilot protection of flexible DC overhead line based on characteristics of fault current loop[J]. Automation of Electric Power Systems, 2022,46(21):152-159.
- [19] 赵晨光,周次明,庞彦东,等.基于独热码有限状态机的斐索干涉解调相位补偿方法[J].光子学报,2020,49(5):7-15.
- ZHAO Chenguang, ZHOU Ciming, PANG Yandong, et al. Phase compensation method of Fizeau interference demodulation based on one-hot encode finite state machine[J]. Acta Photonica Sinica, 2020,49(5):7-15.
- [20] CHEN T, GUEST R IN C. XGBoost: a scalable tree boosting system[C]//22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 2016.
- [21] 朱利鹏,陆超,黄河,等.基于广域时序数据挖掘策略的暂态电压稳定评估[J].电网技术,2016,40(1):180-185.
- ZHU Lipeng, LU Chao, HUANG He, et al. Wide-area time series data mining based transient voltage stability assessment[J]. Power System Technology, 2016, 40(1):180-185.