

中文引用格式:高晓旭,田佳可,高璐,等.基于近似马尔可夫毯与GBDT的煤矿职业健康噪声风险分类[J].中国安全科学学报,2025,35(9):253-262.

英文引用格式:GAO Xiaoxu, TIAN Jiake, GAO Lu, et al. Noise risk classification of coal mine occupational health based on approximate Markov blanket and GBDT[J]. China Safety Science Journal, 2025, 35(9): 253-262.

基于近似马尔可夫毯与GBDT的煤矿职业健康噪声风险分类*

高晓旭^{1,2}教授,田佳可^{**1},高璐³,杜芦¹,范萌杰¹

(1 西安科技大学 能源与矿业工程学院,陕西 西安 710054;

2 西安科技大学 西部矿井开采及灾害防治教育部重点实验室,陕西 西安 710054;

3 陕西靖神铁路有限责任公司,陕西 榆林 719000)

中图分类号:X966

文献标志码:A

DOI: 10.16265/j.cnki.issn1003-3033.2025.09.1075

基金项目:国家自然科学基金资助(52074208)。

【摘要】为精准判断综采工作面噪声对作业人员的健康损害程度,采用人-机-环-管系统理论结合 Fisher Score、最大信息系数及近似马尔可夫毯方法,确定噪声职业健康的关键影响因素;构建基于梯度提升决策树(GBDT)算法的煤矿噪声风险分类预测模型,并以 Kappa 系数及其准确率作为模型效率的指标,来对比验证本算法模型的准确性。结果表明:综采工作面噪声职业健康损害情况与个体状况、设备配置、环境因素及职业健康管理等因素密切相关,其中,岗位类别、个体年龄、工龄时长、防护意识、设备自动化程度、噪声监测点合格率、噪声暴露量、混响时间和管理机构及人员为职业健康风险分类预测的关键指标;基于 GBDT 构建的煤矿噪声职业健康风险分类预测模型准确率最大达 99.6%,平均准确率和 Kappa 系数分别为 98.3%和 0.958;计算确定 6 种综采工作面噪声职业健康风险分类预测模型评估准确率次序为:GBDT > 遗传算法优化随机森林算法(GA-RF) > 粒子群算法优化最小二乘支持向量机算法(PSO-LSSVM) > 随机森林算法(RF) > 支持向量机算法(SVM) > 决策树。

【关键词】近似马尔可夫毯; 梯度提升决策树(GBDT); 煤矿噪声; 噪声健康风险分类; 职业健康; Fisher Score

Noise risk classification of coal mine occupational health based on approximate Markov blanket and GBDT

GAO Xiaoxu^{1,2}, TIAN Jiake¹, GAO Lu³, DU Lu¹, FAN Mengjie¹

(1 College of Energy and Mining Engineering, Xi'an University of Science and Technology, Xi'an Shaanxi 710054, China; 2 Key Laboratory of Western Mine Exploration and Hazard Prevention Under Ministry of Education, Xi'an University of Science and Technology, Xi'an Shaanxi 710054, China; 3 Shaanxi Jingshen Railway Co., Ltd., Yulin Shaanxi 719000, China)

Abstract: To accurately assess the impact of noise on the health of workers in fully mechanized mining face, the key influencing factors of noise occupational health were determined by using the theory of man-

* 文章编号:1003-3033(2025)09-0253-10; 收稿日期:2025-04-26; 修稿日期:2025-07-12

** 通信作者:田佳可(2001—),女,陕西西安人,硕士研究生,研究方向为职业健康与信息化。E-mail:1319650243@qq.com。

machine-environment-management system, combined with Fisher Score, maximum information coefficient and approximate Markov blanket method. The prediction model of coal mine noise risk classification based on the GBDT algorithm was constructed, and the Kappa coefficient and its accuracy were used as the index of model efficiency to compare and verify the accuracy of the model. The results show that the occupational health damage of noise in fully mechanized mining face is closely related to individual status, equipment configuration, environmental factors and occupational health management. Among these, job category, individual age, length of service, protection awareness, degree of equipment automation, pass rate of noise monitoring points, noise exposure, reverberation time and management institutions and personnel are the key indicators of occupational health risk classification and prediction. The accuracy of the occupation health risk classification prediction model of coal mine noise based on GBDT is up to 99.6%, and the average accuracy and Kappa coefficient are 98.3% and 0.958, respectively. The evaluation accuracy of six prediction models for noise occupational health risk classification in fully mechanized mining face is ranked as follows: GBDT > Genetic Algorithm optimization Random Forest (GA-RF) > Particle Swarm optimization Least Squares Support Vector Machine (PSO-LSSVM) > Random Forest (RF) > Support Vector Machine (SVM) > Decision tree.

Keywords: approximate Markov blanket; gradient boosting decision tree (GBDT); coal mine noise; noise health risk classification; occupational health; Fisher Score

0 引言

井下煤炭开采过程中,综采工作面作业环境恶劣,粉尘、噪声等职业病危害严重,作业人员长期接触高噪声、粉尘等而引发的职业健康问题逐渐成为限制煤炭领域绿色高质量发展的短板^[1]。煤矿噪声具有强度大、频谱较宽及持续时间长等特点,长期暴露于这种环境中易引起听力下降、耳鸣等听觉系统疾病,同时,还会引起神经衰弱、血压升高等非听觉系统疾病。煤矿噪声健康风险分类是目前降低工作面噪声危害的前提,开展噪声健康风险分类对煤矿职业人员的健康和安全生产具有指导意义和应用价值。

关于煤矿职业噪声损害的研究,目前主要包括噪声职业危害影响因素分析及健康风险损伤模型构建2个方面。SUN Kan等^[2-3]利用公开的健康安全检查数据库评估,证明了听力损失风险在采矿部门、子单位运营和职业类别中并不统一。常波峰等^[4]发现,噪声水平超过噪声职业接触限值的煤矿作业岗位数量占比达到50%,并提出了针对性的降噪措施。王海椒等^[5]利用模糊层次分析法建立了煤矿噪声风险评价体系,并准确评估了煤矿噪声风险级别。MILAD等^[6]使用结构方程建模分析,发现噪声暴露量对工人听力损失影响最大。ASGHARI等^[7]通过建立 Logistic 回归模型,发现噪声暴露、职业类型、年龄及吸烟均会提高工人听力损失风险。QI

Haihui等^[8]使用广义估计方程建立了基于个体的听力损失预警模型,实现了工人健康状况提前管理。王坤^[9]利用反向传播神经网络、健康风险和损害评价相结合的方法评价了噪声职业健康损害,研究了不同指标对工人的健康损害情况。陈春艳等^[10]使用 Logistic 回归分析计算出累计噪声暴露与疾病之间的关联指标,并基于累计噪声暴露量建立噪声健康损害评估模型。上述研究极大地提高了我国煤矿职业噪声损害分类研究水平,但客观来讲,目前神经网络及机器学习等理论并未将煤矿内外部作业环境、生产设备和企业管理等因素纳入评估预警模型并进行综合性分析,使得所构建的模型未能充分考量实际噪声风险分类预测的复杂程度。

鉴于此,笔者将从人-机-环-管系统理论出发,分析煤矿噪声职业健康损害评估指标,并以 Fisher Score 法为基础,结合最大信息系数和近似马尔可夫毯特征选择方法,筛选出高相关、低冗余的健康损害关键影响因素,基于梯度提升决策树 (Gradient Boosting Decision Tree, GBDT) 算法构建煤矿噪声风险分类预测模型,并通过 Kappa 系数及总体分类精度验证 GBDT 算法的准确性与优越性,以期为综采工作面噪声职业健康损害风险分类提供一种新方法,保护作业人员身心健康,提高煤矿职业健康管理水平和工作效率。

1 煤矿噪声健康风险分类体系

1.1 噪声健康损害评估模型

利用人-机-环-管系统理论,从作业人员自身(人)、生产工艺及相关设备(机)、内外部作业环境(环)和企业职业健康管理(管)出发,初步确定

综采工作面噪声职业健康损害评估体系个体状况等 4 个一级指标。根据各一级指标的不同性质及其条件,确定包含岗位类别、设备自动化程度、距噪声源距离和管理机构及人员等 22 个二级指标。综采工作面噪声职业健康损害评估指标体系如图 1 所示。

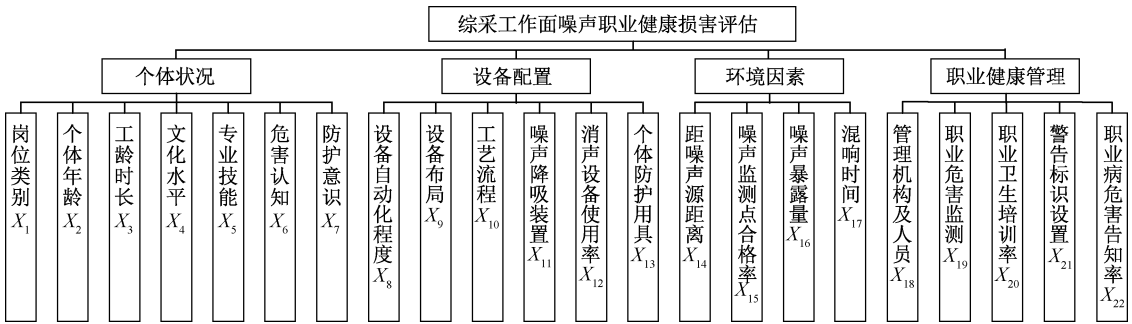


图 1 综采工作面噪声职业健康损害评估指标体系

Fig. 1 Evaluation index system for occupational health damage caused by noise in fully mechanized mining face

由于各评估指标的测度不同,将不同指标数据作标准化处理,并结合指标特点将各指标分为效益型、成本型和固定型。其中,作业人员岗位类别 X_1 为固定型。依据文献[11]中美国环境保护局健康风险评价方法风险分级标准,将矿工噪声职业健康

风险 R 分为 I 级(严重风险, $R>1.0\times10^{-4}a^{-1}$)、II 级(可接受风险, $1.0\times10^{-6}\leq R\leq 1.0\times10^{-4}a^{-1}$)、III 级(潜在风险, $R<1.0\times10^{-6}a^{-1}$),噪声职业健康损害评估指标测度见表 1。

表 1 综采工作面噪声职业健康损害评估指标风险分级标准

Table 1 Risk classification standard for assessment indicators of occupational health damage caused by noise in fully mechanized mining face					
指标类型		目标类型	等级界限(可能遭受的噪声职业健康风险)		
一级	二级		I (严重)	II (可接受)	III (潜在)
个体状况	个体年龄 X_2 /岁	效益型	[20,30)	[30,50]	(50,60]
	工龄时长 X_3 /a	成本型	(20,40]	[2,20]	(0,2)
	文化水平 X_4 /分	效益型	1	[2,4]	5
	专业技能 X_5 /分	效益型	1	[2,4]	5
	危害认知 X_6 /%	效益型	[0,80)	[80,95)	[95,100]
	防护意识 X_7 /%	效益型	[0,80)	[80,95)	[95,100]
设备配置	设备自动化程度 X_8 /%	效益型	[0,80)	[80,95)	[95,100]
	设备布局 X_9 /%	效益型	[0,80)	[80,95)	[95,100]
	工艺流程 X_{10} /%	效益型	[0,80)	[80,95)	[95,100]
	噪声降吸装置 X_{11} /%	效益型	[0,80)	[80,95)	[95,100]
	消声设备使用率 X_{12} /%	效益型	[0,80)	[80,95)	[95,100]
	个体防护用具 X_{13} /%	效益型	[0,80)	[80,95)	[95,100]
环境因素	距噪声源距离 X_{14} /m	效益型	[0,20)	[20,90]	(90,200]
	噪声监测点合格率 X_{15} /%	效益型	[0,80)	[80,95)	[95,100]
	噪声暴露量 X_{16} /dB	成本型	[100,200]	[85,100)	[0,85)
职业健康管理	管理机构及人员 X_{18} /%	效益型	[0,80)	[80,95)	[95,100]
	职业危害监测 X_{19} /%	效益型	[0,80)	[80,95)	[95,100]
	职业卫生培训率 X_{20} /%	效益型	[0,80)	[80,95)	[95,100]
	警告标识设置 X_{21} /%	效益型	[0,80)	[80,95)	[95,100]
	职业病危害告知率 X_{22} /%	效益型	[0,80)	[80,95)	[95,100]

1.2 基于 Fisher Score-近似马尔可夫毯的特征选择

提出以 Fisher Score 法为基础,结合最大信息系数,建立基于特征重要度排序及近似马尔可夫毯下的特征降维方法。2 阶段特征选择方法流程如图 2。

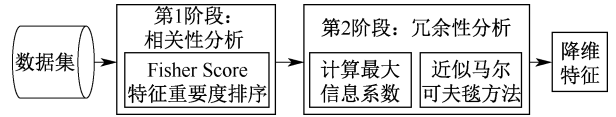


图 2 2 阶段特征选择方法流程

Fig. 2 Two-stage feature selection method flow

在第 1 阶段相关性分析中,通过计算影响因素 Fisher Score 值,选择类间差异大、类内差异小的噪声职业健康损害特征子集^[12],实现对噪声职业健康损害评估特征量的重要度排序,计算方法见下式:

$$F_k = \frac{\sum_{1 \leq s < q \leq l} ((n_s + n_q - n_{sq})/N)(u_k^s - u_k^q)^2}{\sum_{j=1}^l \frac{1}{n_j - 1} \sum_{k=1}^{n_j} (x_{kj}^j - u_k^j)^2} \quad (1)$$

式中: F_k 为第 k 个特征在数据集上的 Fisher Score 值; n_s 和 n_q 分别为 s 、 q 类别中的样本数; u_k^s 、 u_k^q 分别为第 k 个特征在类别 s 和 q 的样本均值; N 为去除重复特征值的样本总数; n_j 为第 j 类样本数;分母为类内散度之和;分子为多类别中任意 2 种类别的组合,并进行求和运算。

采用 Fisher Score 方法,计算得出各影响因素 Fisher Score 值,见表 2。Fisher Score 值越大,对应特征辨识度越强,重要度也越大^[13]。由表 2 可知:筛选并删除 Fisher Score<0.050 的噪声健康损害影响因素,将选取的因素子集作为第 2 阶段的输入,降低第 2 阶段冗余性指标筛选运算规模。

表 2 噪声健康损害影响因素 Fisher Score 值

Table 2 Influencing factors of noise health damage

Fisher Score value			
因素	Fisher Score	因素	Fisher Score
X_1	0.289	X_{12}	0.322
X_2	0.250	X_{13}	0.662
X_3	0.827	X_{14}	0.176
X_4	0.165	X_{15}	1.592
X_5	0.147	X_{16}	3.663
X_6	0.016	X_{17}	0.228
X_7	0.752	X_{18}	0.640
X_8	0.093	X_{19}	0.064
X_9	0.051	X_{20}	0.014
X_{10}	0.008	X_{21}	0.013
X_{11}	0.306	X_{22}	0.014

在第 2 阶段冗余性分析中,基于最大信息系数

进一步挖掘变量间依赖关系,综合最大信息系数及近似马尔可夫毯方法删除冗余特征,最终选取高相关、低冗余的综采工作面噪声职业健康损害特征子集。由文献[14]得出,网格划分的上限值 $B(n) = n^{0.6}$ 时变量之间依赖关系最为理想,其中, n 为样本数量,故采用该值。

基于最大信息系数计算方法,其他 17 个噪声健康损害影响因素与风险类别相关性及其影响因素间相关性分别见表 3 及图 3。由表 3 及图 3 可知:噪声暴露量、噪声监测点合格率、管理机构及人员等因素与噪声职业健康风险类别相关度高,分别为 0.706、0.379、0.356,专业技能、消声设备使用率、设备布局等因素与健康风险类别相关度较低,仅为 0.018、0.052、0.053;管理结构及人员与职业危害监测冗余性较高,因素间相关度为 0.787。

表 3 噪声健康损害影响因素与风险类别相关性

Table 3 Correlation between influencing factors of noise health damage and risk categories

因素	重要度	因素	重要度
X_1	0.246	X_{12}	0.052
X_2	0.264	X_{13}	0.186
X_3	0.267	X_{14}	0.131
X_4	0.157	X_{15}	0.379
X_5	0.018	X_{16}	0.706
X_7	0.268	X_{17}	0.206
X_8	0.246	X_{18}	0.356
X_9	0.053	X_{19}	0.109
X_{11}	0.072	—	—

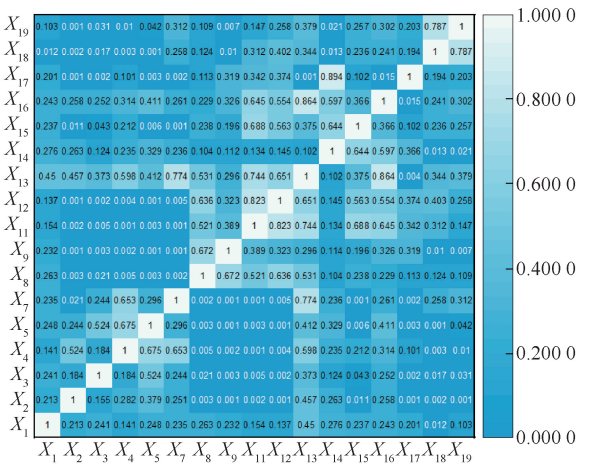


图 3 噪声健康损害影响因素间相关性系数矩阵

Fig. 3 Correlation coefficient matrix between influencing factors of noise health damage

依据近似马尔可夫毯计算方法,删除噪声职业健康损害评估指标体系中冗余特征,主元素选取过程如图 4 所示,图中,以指标 X_1 为例,指标 X_1 为指标 X_9 的近似马尔可夫毯,描述为 $X_1 \rightarrow X_9$,删除以 X_1

为主元素的相关指标,因此 X_1 为主元素,同理,筛选出剩余主元素 X_2 、 X_3 、 X_7 、 X_8 、 X_{15} 、 X_{16} 、 X_{17} 和 X_{18} ,此时,由主元素组成的特征子集构成噪声职业健康损害评估最优特征子集,子集内部特征间冗余性低,特征与噪声健康风险类别间相关度高,筛选后的指标包含了较为丰富的噪声健康损害影响信息。最终选取岗位类别、个体年龄、工龄时长、防护意识、设备自动化程度、噪声监测点合格率、噪声暴露量、混响时间和管理机构及人员 9 个特征作为综采工作面噪声职业健康风险分类预测模型的关键指标。

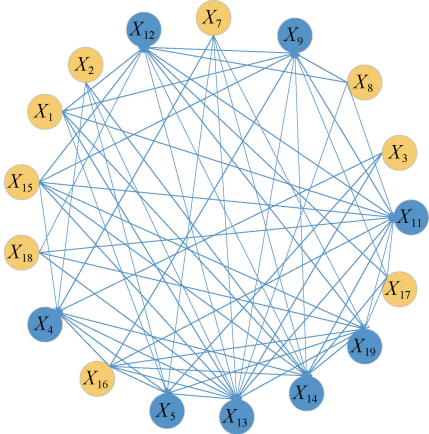


图 4 近似马尔可夫毯主元素选取过程
Fig. 4 Selection process of approximate Markov blanket main elements

2 GBDT 噪声健康风险分类模型

2.1 GBDT 算法模型

GBDT 隶属于集成学习中的 boosting 方法,其含义为采用梯度提升的优化思想训练出决策树模型,并通过寻找一系列合适的弱分类器进行有效线性组合来达到提升模型性能的目的。GBDT 算法通过回归树和目标值间的残差进行归类分析,每次迭代都会顺着残差减少的梯度方向构建一棵新决策树,随着迭代次数增加,决策树数量不断增加,计算值逐渐逼近目标值。因此,即使样本中有错误点,模型训练结果也不会受影响,具有良好的鲁棒性^[15]。

运用 GBDT 理论预测综采工作面噪声职业健康风险,其分类的基本步骤如图 5 所示。以实测的健康风险数据为基础,将噪声职业健康损害评估关键指标作为模型输入序列,按照 4 : 1 划分数据样本,通过 5 折交叉验证法验证 GBDT 算法模型性能,输出不同训练样本下的准确率并保存 GBDT 训练模型。最终将训练模型用于煤矿职业健康风险分类测试数据集,输出作业人员职业健康风险分类类别,包括严重风险、可接受风险和潜在风险。

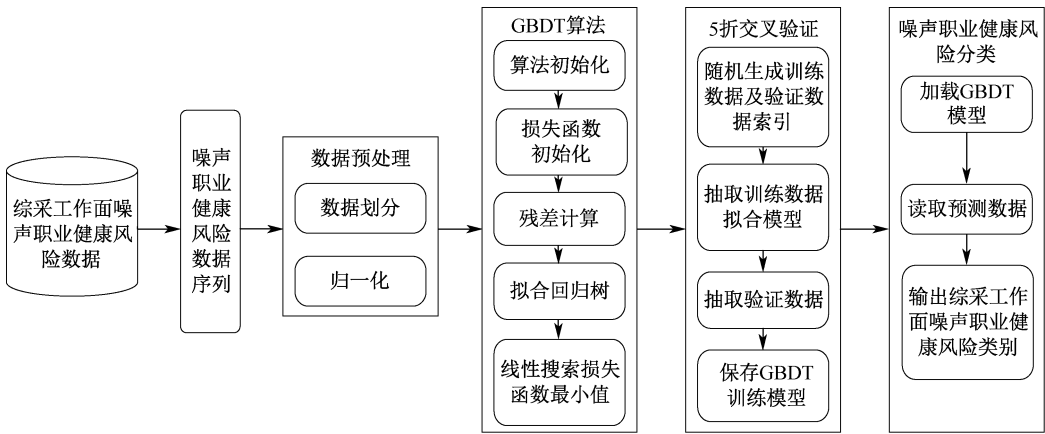


图 5 综采工作面噪声职业健康风险分类预测模型
Fig. 5 Classification prediction model of occupational health risk for noise in fully mechanized mining face

2.2 性能评价指标

由于数据样本较少,采用 5 折交叉验证法完成噪声健康风险分类预测模型性能验证,并且在算法模型验证过程中,同时参考 Kappa 系数及总体分类精度衡量综采工作面噪声职业健康风险分类预测模

型性能。
总体分类精度为正确分类总样本数与所选取样本数量的比值,Kappa 系数用于衡量真实分类与预测结果的一致性,较为全面地反映了分类器评估效果。

3 煤矿噪声健康风险分类试验结果与分析

3.1 基于 GBDT 煤矿噪声健康风险分类

3.1.1 数据来源与预处理

由于在井下综采工作面布置有采煤机、刮板输送机、破碎机、带式输送机及转载机等机械设备,且设备分布较为集中,因此,选择接噪作业人员较为集中的地点作为噪声监测点。选取陕北地区 X 煤矿综采工作面作为研究对象,以《煤矿作业场所职业

病危害防治规定》和《工作场所物理因素测量 第 8 部分:噪声》(GBZ/T189.9—2007)为测量依据,以 YSD130 矿用防爆噪声仪为测量工具,测量采煤机司机、刮板机司机、转载机司机等 7 个接噪岗位的噪声^[16]。不同作业岗位的噪声暴露强度均测量 3 次并取均值,将 8 h 等效声级 $L_{EX,8h}$ 的测量数值作为不同作业岗位的噪声暴露量,并调查统计设备配置、环境情况以及各岗位作业人员的年龄、工龄时长、防护意识等个人信息,最终得到 612 组有效数据。X 煤矿部分噪声健康风险分类实测样本数据见表 4。

表 4 X 煤矿部分噪声健康风险分类实测样本数据

Table 4 Measured sample data of noise health risk classification from X coal mine										
数据编号	工种	监测地点	年龄/岁	工龄/a	防护意识/%	设备自动化程度/%	监测点合格率/%	噪声暴露量 $L_{EX,8h}$ /dB(A)	混响时间/s	管理机构及人员比例/%
1	采煤机司机	采煤机司机操作位	46	15	82	86	90	94.1	0.52	88
2	超前支护工	超前支护工操作位	49	18	86	85	79	82.1	0.65	81
3	刮板机司机	刮板机司机操作位	50	20	94	81	86	93.5	0.49	75
4	转载机司机	转载机司机操作位	37	11	81	87	93	92.7	0.73	83
5	破碎机司机	破碎机旁	42	13	89	88	85	95.3	0.55	90
6	乳化泵工	设备列车旁	45	10	90	93	89	96.4	0.62	85
7	带式输送机司机	带式输送机司机操作位	35	8	79	89	83	85.8	0.83	91
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
612	带式输送机司机	带式输送机司机操作位	52	19	92	82	81	88.6	0.77	87

3.1.2 GBDT 模型分类结果分析

采用 GBDT 模型判定综采工作面作业工人噪声职业健康风险类别,为提高预测模型准确率,采用 5 折交叉验证法和网格寻优法优化 GBDT 模型的迭代次数、学习率、树的复杂度等超参数,依据文献[17]所取 GBDT 最优超参选取区间,选定的参数搜索范围见表 5。通过 sklearn 中 GridSearchCV 方法的 best_estimator_属性进行参数寻优,目标参数调试过程如图 6 所示,最终调试确定模型的超参数为:迭代次数 500,学习率 0.05,决策树最大深度为 5,划分所需最小样本数为 2。

为验证 GBDT 噪声职业健康风险评估模型的正确性,测试剩余的 20%样本数据,由第 1.2 节可知:噪声暴露量对健康风险影响较大。因此,将其作为散点图纵坐标标志,散点图横坐标为测试集样本编

表 5 超参数及选择范围

Table 5 Hyperparameters and selection range	
超参数	数值范围
迭代次数	[100,1 000],步长为 100
模型学习率	0.001,0.01,0.02,0.03,⋯,0.1
树最大深度	[1,9],步长为 2
划分所需最小样本数	[1,5],步长为 1

号,最终健康风险散点图及归一化后的混淆矩阵结果如图 7 所示。由图 7 可知:GBDT 算法在严重健康风险类别的预测准确率为 0,其原因是在噪声监测数据样本中存在个别类别为严重健康风险的错误噪声数据;GBDT 算法对可接受风险和潜在健康风险的分类准确率分别为 98.8%和 97.2%,分类预测效果较好。

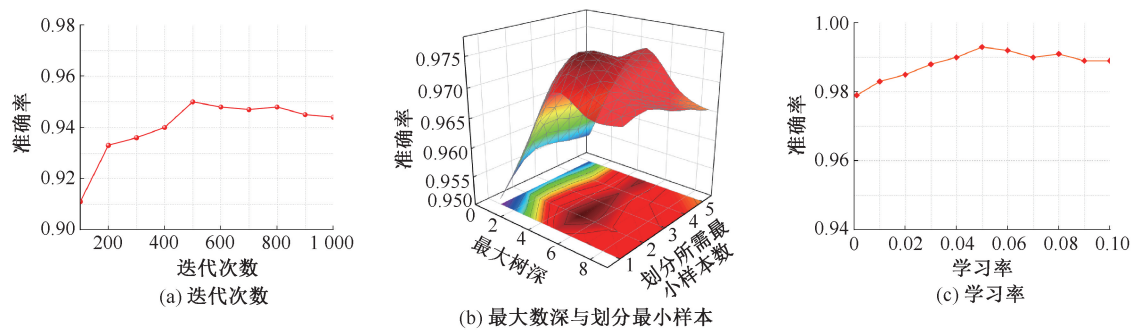


图 6 超参数寻优

Fig. 6 Hyperparameter optimization

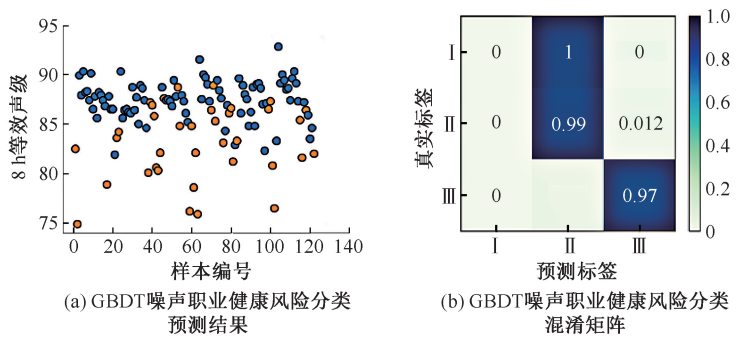


图 7 GBDT 算法煤矿职业健康噪声风险分类结果

Fig. 7 Coal mine occupational health of noise risk classification results based on GBDT algorithm

3.2 对比算法试验结果

为体现本文方法优越性,采用决策树算法、支持向量机算法 (Support Vector Machine, SVM)、粒子群算法优化最小二乘支持向量机算法 (Particle Swarm optimization Least Squares Support Vector Machine, PSO-LSSVM)、随机森林算法 (Random Forest, RF)、遗传算法优化随机森林算法 (Genetic Algorithm optimization Random Forest, GA-RF) 对数据样本进行建模预测。其中, SVM 基于 RBF 核, 使用“一对其余”方法构造多分类器, 惩罚系数 $C=1.0$, 核函数参数 $\gamma=0.118$, 决策树和随机森林最大深度为 5, 随机森林决策树个数为 50, PSO-LSSVM 模型最优参数为 $\text{gamma}=85.89$, $\text{sigma}=0.5$ 。

在综采工作面噪声职业健康风险分类预测模型中, 以 5 折交叉检验得到的 Kappa 系数及准确率作为模型效率的衡量标准, 对 GBDT、PSO-LSSM、GA-RF 等 6 种算法依据数据样本进行建模预测, 通过计算混淆矩阵, 6 种评估预测方法训练结果如图 8 所示, 由图 8 可知: 经过相关参数优化, 6 种算法模型在噪声职业健康风险分类数据集上均有较好的表现, 平均 Kappa 系数均大于 0.810。其中, GBDT 准确率最大达 99.6%, 平均准确率和 Kappa 系数分别

为 98.3% 和 0.958; 决策树算法的准确率和 Kappa 系数为这 6 种评估预测方法中的最小值, 5 次交叉验证结果中, 决策树的平均准确率为 93.5%, 平均 Kappa 系数为 0.834; SVM 算法考虑了数据样本中的支持向量, 可处理特征数较多的数据, 鲁棒性良好, 因此, 其相对于决策树模型精度较高, 平均准确率和 Kappa 系数分别为 94.5% 和 0.865; PSO-LSSVM 算法的平均准确率为 95.4%, Kappa 系数为 0.888; RF 算法的平均准确率和 Kappa 系数分别为 94.7% 和 0.873; GA-RF 算法的准确率和 Kappa 系数为这 6 种算法中变化范围最大的算法, 准确率范围为 93.4%~99.1%, Kappa 系数范围为 0.823~0.977。

相比于其他 5 种模型, GBDT 噪声健康风险分类预测模型可通过增加决策树的数目减少模型偏差, 并且该模型利用多棵决策树间线性组合降低了模型误差, 相较于单一决策树模型更加稳定, 多分类器集成丰富了 GBDT 算法模型多样性。因此, 在综采工作面噪声健康风险分类方面, GBDT 算法模型表现能力比其他 5 种算法模型更好, 取得了较好的预测效果。

图 9 和图 10 分别为 SVM、Decision Tree、PSO-LSSVM、RF、GA-RF 算法对数据集的预测散点图和混

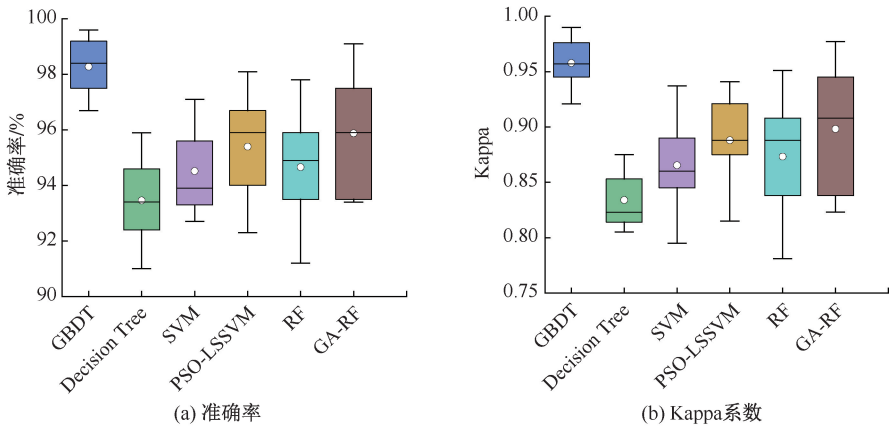


图 8 噪声职业健康风险预测模型结果对比

Fig. 8 Comparison of noise occupational health risk prediction model results

淆矩阵图。由图 10 可知:几种算法在严重健康风险类别的预测准确率与 GBDT 算法结果相同且均为 0;由图 10a 和图 10b 可知:SVM 和决策树算法对可接受风险类别预测准确率较高,但对噪声健康风险类别为潜在风险的分类准确率均较低,仅为 83.0%和 78.0%;在图 10c 中,经 PSO 优化的 LSSVM 算法健康

风险预测准确率相较 SVM 和决策树算法更高,对可接受风险分类准确率达 100%,但潜在风险类别准确率一般;在图 10d 和图 10e 中,对潜在风险的预测准确率由 RF 算法 83.0%的准确率提高至 GA-RF 算法的 94.4%,但对可接受风险类别的预测准确率由 RF 算法 99.0%的准确率下降至 GA-RF 算法的 97.6%。

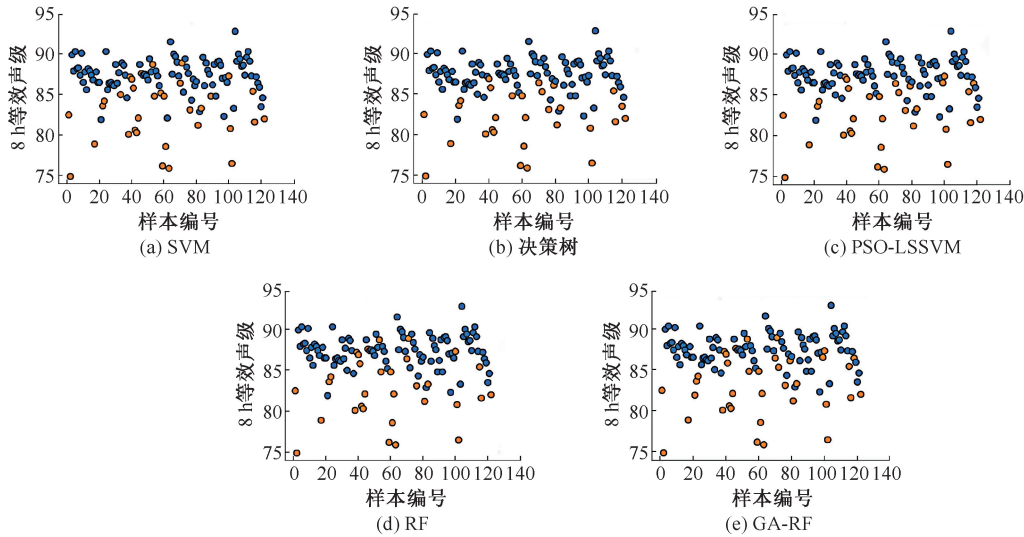


图 9 噪声职业健康风险分类预测结果

Fig. 9 Prediction results of noise occupational health risk classification

分析可知:与传统的机器学习算法相比,GBDT 算法的总体准确率和 Kappa 系数均为最大,且对潜在健康风险的分类准确率最高;GA-RF 算法的平均准确率和 Kappa 系数均次于 GBDT 算法,且对潜在健康风险的分类准确率为 94.4%;PSO-LSSVM 算法对可接受风险预测准确率为 6 种算法中的最高值,但对潜在健康风险的分类准确率一般;RF 算法对可接受风险和潜在健康风险的预测准确率均低于 PSO-LSSVM 算法;SVM 算法虽然与 RF 算法对潜在

健康风险的预测准确率相同,但可接受风险的预测准确率却低于 RF 算法;决策树算法的总体准确率和 Kappa 系数均为最小值,且对潜在健康风险的预测准确率为 78.0%,为最低预测准确率。可见:综采工作面噪声职业健康风险分类预测模型评估准确率次序为:GBDT > GA-RF > PSO-LSSVM > RF > SVM > 决策树,且 GBDT 算法对样本数据量均衡性要求不高。因此,参数优化后的 GBDT 算法模型适用于综采工作面噪声职业健康风险评估,其分类精度在 3 类健

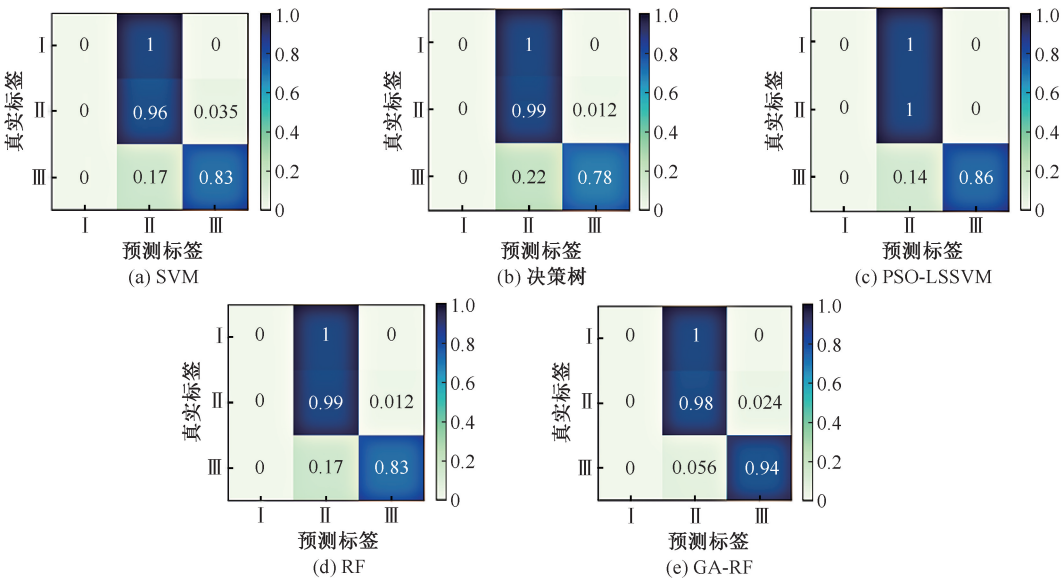


图 10 噪声职业健康风险分类混淆矩阵

Fig. 10 Confusion matrix noise occupational health risk classification

康风险预测中具有显著优势,能在实际综采工作面噪声职业健康风险评估中得到较好应用。

4 结 论

1) 从人-机-环-管系统理论出发,结合 Fisher Score、最大信息系数及近似马尔可夫毯的特征选择方法,确定综采工作面噪声职业健康风险分类的关键指标,即岗位类别、个体年龄、工龄时长、防护意识、设备自动化程度、噪声监测点合格率、噪声暴露

量、混响时间和管理机构及人员。

2) 构建基于 GBDT 算法的综采工作面噪声职业健康风险分类预测模型,该模型可有效精准预测判定噪声风险级别,为保障接噪人员的身心健康和煤矿企业安全高效生产提供一定借鉴。

3) 文中研究聚焦于煤矿综采面噪声风险级别分类,尚未系统研究噪声所导致的职业健康损害,对防噪预警方面的研究也存在缺失。在后续的研究中进一步完善煤矿综采工作面噪声相关研究体系。

参 考 文 献

[1] 顾大钊, 李全生. 基于井下生态保护的煤矿职业健康防护理论与技术体系[J]. 煤炭学报, 2021, 46(3): 950-958.
GU Dazhao, LI Quansheng. Theoretical framework and key technologies of underground ecological protection based on coal mine occupational health prevention[J]. Journal of China Coal Society, 2021, 46(3): 950-958.

[2] SUN Kan, AZMAN A S. Evaluating hearing loss risks in the mining industry through MSHA citations[J]. Journal of Occupational and Environmental Hygiene, 2018, 15(3): 246-262.

[3] SUN Kan, AZMAN A S, CAMARGO H E, et al. Risk assessment of recordable occupational hearing loss in the mining industry[J]. International Journal of Audiology, 2019, 58(11): 761-768.

[4] 常波峰, 黄伟, 刘宽, 等. 陕西省某煤矿企业噪声职业健康风险评估[J]. 公共卫生与预防医学, 2024, 35(1): 70-73.
CHANG Bofeng, HUANG Wei, LIU Kuan, et al. Occupational health risk assessment of noise in a coal mining enterprise in Shaanxi province[J]. Journal of Public Health and Preventive Medicine, 2024, 35(1): 70-73.

[5] 王海椒, 邹晓雪, 刘丽华, 等. 模糊层次分析法在煤矿生产性噪声危害分级评价中的应用[J]. 中国工业医学杂志, 2019, 32(4): 286-289.
WANG Haijiao, ZOU Xiaoxue, LIU Lihua, et al. Application of fuzzy analytic hierarchy process in classification and evaluation of productive noise hazard[J]. Chinese Journal of Industrial Medicine, 2019, 32(4): 286-289.

- [6] MILAD A, SAEID Y, ALI A. Developing and validating a risk assessment method for noise-induced hearing loss in workers[J]. Heliyon, 2024, 10(22): DOI: 10. 1016/J. HELIYON. 2024. E40475.
- [7] ASGHARI M, GORJI R, MORADZADEH R, et al. A risk model for occupational noise-induced hearing loss in workers[J]. Work, 2023, 77(3): 1 017-1 022.
- [8] QI Haihui, DU Yiyi, TIAN Yu, et al. Research on early warning model of the hearing loss of workers exposed to noise[J]. Chinese Journal of Industrial Hygiene and Occupational Diseases, 2023, 41(1): 47-51.
- [9] 王坤. 机械加工车间环境质量及其职业健康损害评价研究[D]. 重庆: 重庆大学, 2019.
WANG Kun. Study on environmental quality and occupational health damage assessment in mechanical processing workshop[D]. Chongqing: Chongqing University, 2019.
- [10] 陈春艳, 龚志起. 钢构件生产车间噪声健康损害评价[J]. 科学技术创新, 2021(10): 32-35.
- [11] 李丹丹. 不同类型巷道中噪声危害评价研究[D]. 廊坊: 华北科技学院, 2020.
LI Dandan. Study on noise hazard evaluation in different types of laneways[D]. Langfang: North China Institute of Science and Technology, 2020.
- [12] 赵玲, 龚加兴, 黄大荣, 等. 基于 Fisher Score 与最大信息系数的齿轮箱故障特征选择方法[J]. 控制与决策, 2021, 36(9): 2 234-2 240.
ZHAO Ling, GONG Jiaying, HUANG Darong, et al. Fault feature selection method of gearbox based on Fisher Score and maximum information coefficient[J]. Control and Decision, 2021, 36(9): 2 234-2 240.
- [13] 贾改妮, 蒋方, 杨明, 等. 煤矿井下不同作业场所的职业健康损害研究[J]. 中国安全科学学报, 2023, 33(4): 221-229.
JIA Gaini, JIANG Fang, YANG Ming, et al. Research on occupational health damage in different working places in underground coal mine[J]. China Safety Science Journal, 2023, 33(4): 221-229.
- [14] RESHEF D N, RESHEF Y A, FINUCANE H K, et al. Detecting novel associations in large data sets[J]. Science, 2011, 334(6062): 1 518-1 524.
- [15] 马陇飞, 萧汉敏, 陶敬伟, 等. 基于梯度提升决策树算法的岩性智能分类方法[J]. 油气地质与采收率, 2022, 29(1): 21-29.
MA Longfei, XIAO Hanmin, TAO Jingwei, et al. Intelligent lithology classification method based on GBDT algorithm[J]. Petroleum Geology and Recovery Efficiency, 2022, 29(1): 21-29.
- [16] 高晓旭, 高璐, 潘相旭, 等. 基于 t-SSA-BP 的煤矿噪声职业健康损害预测[J]. 中国安全科学学报, 2023, 33(12): 214-222.
GAO Xiaoxu, GAO Lu, PAN Xiangxu, et al. Prediction of occupational health damage caused by coal mine noise based on t-SSA-BP[J]. China Safety Science Journal, 2023, 33(12): 214-222.
- [17] 韩启迪, 张小桐, 申维. 基于梯度提升决策树(GBDT)算法的岩性识别技术[J]. 矿物岩石地球化学通报, 2018, 37(6): 1 173-1 180.
HAN Qidi, ZHANG Xiaotong, SHEN Wei. Lithology identification technology based on gradient boosting decision tree (GBDT) algorithm[J]. Bulletin of Mineralogy, Petrology and Geochemistry, 2018, 37(6): 1 173-1 180.



作者简介: 高晓旭 (1978—),男,山西吕梁人,博士,教授,硕士生导师,主要从事采矿安全与职业健康等方面的研究。E-mail:396783933@qq.com。