不均衡样本下基于最优化理论的冲击倾向性分类预测方法

贾金兑，丁自伟，张超凡，高成登，巩欣伟，许庆钊，李龙勃
（西安科技大学 能源与矿业工程学院，陕西 西安 710054）

摘要：煤层冲击倾向性等级分类预测是煤矿冲击地压风险评估的重要前提．为提高复杂条件冲击倾向性分类的预测效率及精度，深入挖掘整合构建了含221组样本案例及49组σ_c单特征缺失的样本数据集．针对统计实测数据集中存在的类别不均衡（如强、弱冲击样本相对稀缺等）问题，提出了不均衡样本下基于最优化理论的元启发式算法智能优化支持向量机（SVM）冲击倾向性分类预测方法．结果表明：采用基于K-means的改进SMOTE算法进行空值填充、异常值剔除、重采样与聚类合成新样本等数据预处理，解决了数据集的不均衡问题．提出基于新型元启发式优化算法CFOA（捕鱼优化算法）策略优化的CNN（卷积神经网络）-SVM预测模型，通过智能优化算法对CNN控制分类器核函数进行最优化理论调优，进一步调整了SVM超参数的优化搜索策略；经样本数据类均衡预处理、优化算法改进SVM调优，与其他8种模型对比分析，验证了所提优化SVM模型的优良性能，经数据预处理后CFOA-CNN-SVM模型在测试集中综合性能最优，准确度最高，达96.3%；对于3类标签样本，预测精度分别为0.97，0.95和0.97，召回率分别为1.00，0.95和0.95，F_1值分别为0.99，0.95和0.96．最后通过实际案例验证了所提模型的泛化能力，其在处理样本不均衡及单一特征数据缺失、小样本试验数据中表现出了良好的预测效果．研究成果对于提高煤层冲击倾向性预测数据库的质量和改善模型预测性能具有重要的基础价值．

关键词：冲击倾向性；不均衡样本；SMOTE算法；元启发式优化；支持向量机；分类预测
中图分类号：TD 324　　**文献标志码**：A　　**文章编号**：1000-1964（2026）03-0862-15

Coal bursting liability classification prediction method based on optimization theory under unbalanced samples

JIA Jindui, DING Ziwei, ZHANG Chaofan, GAO Chengdeng,
GONG Xinwei, XU Qingzhao, LI Longbo
(School of Energy and Mining Engineering, Xi′an University of Science and Technology, Xi′an, Shaanxi 710054, China)

Abstract: Classification prediction of coal seam bursting liability grade is a crucial prerequisite for risk assessment of coal mine rock bursts. To enhance the prediction efficiency and accuracy of coal bursting liability classification under complex conditions, a dataset comprising 221 sample cases and 49 cases with missing σ_c single features was constructed through deep mining and integration. Aiming at the problems of category imbalance in the statistical measured data set, such as the relative

收稿日期：2025-10-25　　**修回日期**：2025-12-24　　**DOI**：10.13247/j.cnki.jcumt.20250472
基金项目：国家自然科学基金项目（52074209）
第一作者：贾金兑（1996—），男，山西省晋城市人，讲师．E-mail：jiajindui@stu.xust.edu.cn
通信作者：丁自伟（1987—），男，山东省临沂市人，教授，博士生导师．E-mail：zwding519@163.com
引用格式：贾金兑，丁自伟，张超凡，等．不均衡样本下基于最优化理论的冲击倾向性分类预测方法[J]．中国矿业大学学报，2026，55（3）：862-876.
JIA Jindui, DING Ziwei, ZHANG Chaofan, et al. Coal bursting liability classification prediction method based on optimization theory under unbalanced samples[J]. Journal of China University of Mining & Technology, 2026, 55(3): 862-876.

scarcity of strong and weak bursting liability samples, a meta-heuristic algorithm, intelligent optimization support vector machine (SVM), a bursting liability classification prediction method based on optimization theory under unbalanced samples was proposed. The results show that the improved SMOTE algorithm based on K-means is used for data preprocessing, such as null value filling, outlier elimination, resampling and clustering synthesis of new samples, which solves the imbalance problem of data sets. A CNN (Convolutional Neural Network)-SVM prediction model based on a new meta-heuristic optimization algorithm, CFOA (Fishing Optimization Algorithm) strategy optimization, is proposed. The kernel function of the CNN control classifier is optimized by an intelligent optimization algorithm, and further adjusts the SVM hyper-parameter optimization search strategy. The SVM tuning is improved by sample data class equalization preprocessing, and optimization algorithm. Compared with the other 8 models, the excellent performance of the proposed optimized SVM model is verified. The CFOA-CNN-SVM model, after data preprocessing, has the best comprehensive performance in the test set, with the highest accuracy rate of 96.3 %. For the three types of label samples, the prediction accuracy rates are 0.97, 0.95 and 0.97, the recall rates are 1.00, 0.95 and 0.95, and the F_1 values are 0.99, 0.95 and 0.96. Finally, the generalization ability of the model is further verified by practical cases. It shows a good prediction effect in dealing with sample imbalance and feature missing statistical data, and small sample test data. The research results have important basic value for improving the database quality of coal seam bursting liability prediction and improving the model prediction performance.

Key words: bursting liability; unbalanced sample; SMOTE algorithm; meta-heuristic optimization; support vector machine; classification prediction

冲击地压是深部开采面临的重大工程地质灾害之一，易导致巷道毁坏、设备损毁及人员伤亡等灾难性事故发生[1-2]. 冲击倾向性是指煤体内部积聚弹性能并突然释放的固有属性，直接决定了冲击地压发生的可能性与强度[3-4]. 开展冲击地压风险准确评估，是实现冲击地压灾害有效管控的重要前提，而煤层冲击倾向性测试分析与等级分类预测正是此类冲击地压风险评估的源头性基础工作[5].

在预测方法方面，《冲击地压测定、监测与防治方法第2部分：煤的冲击倾向性分类及指数的测定方法》(GB/T 25217.2—2010)是常用的一种冲击倾向性综合判别方法，给出了81种所有可能的指标等级组合，明确了73种组合所对应的、唯一确定的冲击倾向性等级，其优势在于标准化和一致性，但存在决策逻辑不透明、难以优化等缺点，且存在边界盲区[6-7]. 此外，我国厚及特厚煤层分布广泛，软硬复合煤层冲击倾向性迥异，常表现为软煤无冲击、硬煤强冲击，现有方法无法对复合煤层整体进行统一的冲击倾向性界定.

近年来，随着机器学习技术的发展，基于数据驱动的机器学习模型为冲击倾向性预测提供了新的思路，该方法一定程度上降低了指标权重赋值的主观性，提高了实际冲击倾向性等级的预测效率及精度[8]. 目前，学者们开发了支持向量机(SVM)、随机森林(RF)、梯度提升决策树(GBDT)、决策树和K近邻(KNN)等模型，这些模型在冲击倾向性预测中均表现出较好的性能[9]. 其中，SVM通过确定最优超平面以区分不同类别的数据指标，在处理高维、非线性问题方面表现较为出色[10-11]. 文献[12]构建了基于改进萤火虫算法优化支持向量机(IGSO-SVM)的预测模型，针对有限样本条件下非线性问题预测准确度达90%. 文献[13]收集了132个样本数据集，采用启发式遗传算法(GA)和粒子群优化算法(PSO)对SVM模型进行改进，建立了冲击地压分类预测方法. 文献[14]开发了改进哈里斯鹰优化(IHHO)预测模型，在136个案例数据中进行了测试，准确度达94.12%.

然而，基于文献统计所得各矿井的实测数据呈现出高度的离散性与非线性特征，用于驱动预测模型的实测数据普遍存在样本量有限的问题，如通常小于300例、类别不均衡(强/弱冲击倾向性样本稀缺)及单一特征指标数据缺失等问题，导致早期机器学习模型泛化能力弱、预测精度不足. 由于通常存在的数据不均衡问题，以及不同等级样本案例数量差异较大，模型的预测性能和泛化能力显著降低[15].

在数据预处理方面，文献[16-17]指出限于对大量训练数据的依赖性和质量的敏感性，机器学习方法面临数据不平衡、模型泛化能力有限及依赖大量训练数据等问题；文献[18-19]通过改进不平衡数据集的过采样方法，提升了数据库质量与预测结果的准确性；文献[20]提出了一种基于数据预处理与聚类-分类算法相结合的三步法冲击预测模型，通过处理缺失值、异常值和类不平衡问题，提高了预测结果的准确性；文献[21]基于Copula蒙特卡洛模拟(Copula-MC)提出了一种新的过采样算法，用于平衡冲击倾向性数据集的类别；文献[22]通过t-分布邻域嵌入(t-SNE)降维和聚类重新标记数据，提高了模型的泛化能力.

因此，探索更有效的数据重采样技术、模型优化方法和模型验证策略，建立适应不均衡样本分布的数据驱动预测模型，提出基于数据预处理-模型优化-指标验证的综合预测方法意义重大.

本文围绕数据预处理与过采样方法、机器学习分类算法、性能评估及案例验证等方面展开研究：构建了包含221组最新样本案例的冲击倾向性数据集，采用基于K均值聚类算法(K-means)的改进合成少数类过采样技术(SMOTE)算法进行空值填充、异常值剔除、重采样与聚类合成新样本等数据预处理；其次，改进提出基于最优化理论的元启发式智能优化算法冲击倾向预测模型，通过对比9种机器学习算法，分析评估了不同组合模型的预测性能；最后，通过实际案例验证了模型的泛化能力.

1 数据集准备

1.1 数据收集

1) 指标选取

从力学强度、破坏时间、变形特征、刚度以及能量等指标出发，目前学者们已对冲击倾向性分类方法进行了诸多研究，并取得了一定成果[23-24]. 国标提出了基于动态破坏时间、弹性能量指数、冲击能量指数和单轴抗压强度的四参数测定法，据此可将冲击倾向性划分为无、弱、强3个等级，这是实验室冲击倾向性测试的基础，现有文献实测数据也多基于此.

本文选取动态破坏时间(D_T)、弹性能量指数(W_{et})、冲击能量指数(K_E)和单轴抗压强度(σ_c)4个指标作为模型的输入特征参数，将冲击倾向性等级划分为无冲击倾向性(1类)、弱冲击倾向性(2类)、强冲击倾向性(3类)，并将此等级作为机器学习模型的训练与预测标签，见表1.

表1 冲击倾向性分类依据

Table 1 Classification basis of coal bursting liability

标签	分类等级	D_T/ms	W_{et}	K_E	σ_c/MPa
1	无冲击倾向性	500	<2	<1.5	<7
2	弱冲击倾向性	>50～500	2～<5	1.5～<5.0	7～<14
3	强冲击倾向性	≤50	≥5	≥5.0	≥14

2)数据集构建

通过深入挖掘与筛选整合，构建了包含221组试验数据的样本案例，所收集数据均来自中国范围内的煤矿冲击倾向性样本. 构建的数据集中有13例来自文献[25]，20例来自文献[26]，127例来自文献[27]，16例来自现场相关测试报告，其他45例样本均来自在中国知网检索的相关文献. 数据仅涉及煤矿开采地下工程领域，样本来源保持一致，具有完整的4个特征参数、1个分类等级及对应标签，部分原始数据见表2.

表2 收集的冲击倾向性数据库

Table 2 The collected coal bursting liability database

序号	输入特征参数				分类等级	标签
	D_T/ms	W_{et}	K_E	σ_c/MPa		
1	973.00	2.70	1.80	5.90	无	1
2	357.00	2.90	4.60	11.50	弱	2
3	49.00	4.60	6.00	16.50	强	3
4	291.00	12.50	3.10	19.30	强	3
5	72.00	13.20	2.40	19.30	强	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
217	308.00	2.43	1.65	12.06	弱	2
218	17 225.00	0.18	0.96	0.12	无	1
219	122.40	8.42	2.34	14.00	弱	2
220	140.00	6.95	1.55	17.19	弱	2
221	170.60	10.86	1.91	24.89	强	3

1.2 数据分析

数据集各等级冲击倾向性占比如图1a所示. 由于实际研究对于无冲击倾向性关注不足，导致1类案例数据较少；而强冲击倾向性样本本身也较为稀缺，获取该类样本数据较为困难. 因此1类与3类案例数量偏少，这种数据的不平衡性，正是制约实现模型精准预测普遍存在的客观问题.

本文数据库样本总数为221例，其中1类数据占比11.3%(25例)，2类数据占比54.3%(120例)，3类数据占比34.4%(76例)，1，2，3类数据数量比为1.00∶4.80∶3.04. 因此需要采用合适的采样策略来平衡数据集，以减少不平衡现象对模型预测可靠性的影响.

基于数据集特征，建立了4个指标参数的统计数据分布小提琴图，可视化展示了不同冲击倾向性

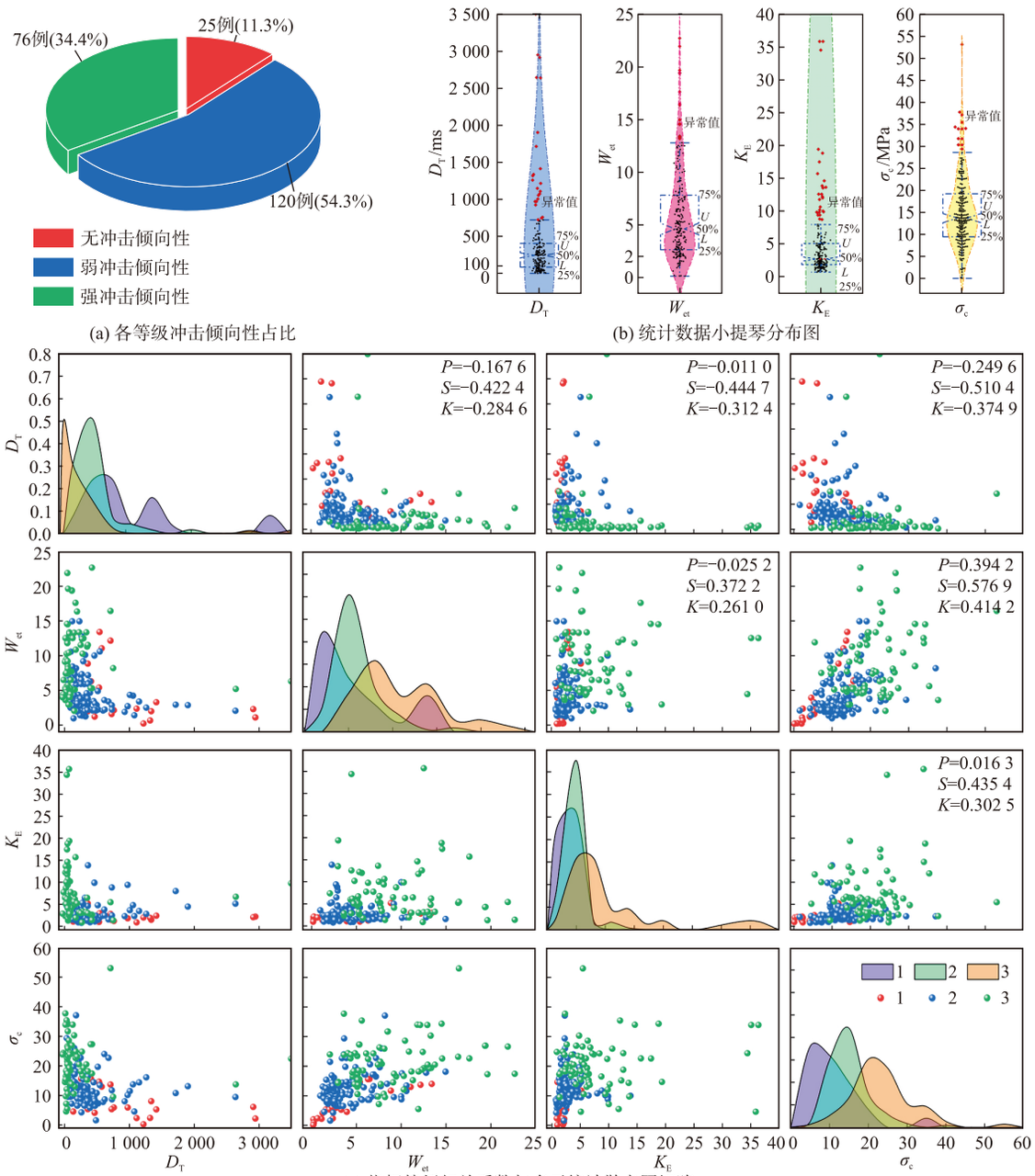
样本的分布情况和统计信息，包含上四分位数、下四分位数、中位数、离散值，如图1b和表3所示．4个特征指标D_T，W_{et}，K_E与σ_c的均值分别为805.19 ms，6.16，14.15与15.08 MPa，中位数分别为248.57 ms，4.64，2.74与13.73 MPa．冲击倾向性越大，4个特征指标的0.25，0.50，0.75分位数值越大．

为定量分析样本数据的多样性和4个特征指标的定量化关系，绘制了多指标散点图矩阵和相关性系数图，如图1c所示．其中，4个特征指标间的3个相关性系数，即皮尔逊系数P、斯皮尔曼系数S与肯德尔等级系数K，反映了特征指标D_T分别与W_{et}，K_E，σ_c呈一定的负相关关系，而W_{et}，K_E，σ_c三者间表现出一定的正相关关系．此外，散点图矩阵外层的低密度和杂乱点以及小提琴图中的极值点，揭示了样本的离散值或异常值．

表3　不同冲击倾向性样本特征统计参数

Table 3　Statistical parameters of sample characteristics with the different coal bursting liability

参数	均值	中位数	标准差	最小值	最大值
D_T/ms	805.19	248.57	3 463.83	5.70	33 224.00
W_{et}	6.16	4.64	5.26	0.18	49.00
K_E	14.15	2.74	135.79	0.76	2 025.00
σ_c/MPa	15.08	13.73	7.91	0.12	53.00



(a) 各等级冲击倾向性占比　　(b) 统计数据小提琴分布图

(c) 指标特征相关系数与多元统计散点图矩阵

图1　样本数据库特征的可视化结果

Fig. 1　Visualization results of sample database features

1.3 数据预处理

针对数据集的缺失值、异常值和类不平衡问题，提出了一种指标特征参数异常数据的预处理方法．主要步骤为：1)缺失值处理，2)异常值剔除，3)类不平衡处理．

1.3.1 缺失值处理

K近邻(KNN)插补是一种基于邻近样本的插补方法，主要用于处理数据集中的缺失值，尤其适用于分类特征数据．首先，确定用于插补的邻近样本的数量；其次，结合数据集中每个非缺失值样本与其他所有样本之间的距离，为每个缺失值样本确定K个最邻近的非缺失值；最后，选择K个邻近样本中出现次数最多的类别，作为缺失值的插补值．

1.3.2 异常值剔除

采用基于聚类方法和四分位距(I_{QR})的策略剔除数据中的异常值．通过聚类方法将数据分组，识别出远离大多数数据的离群点；并采用基于I_{QR}的异常值检测方法进行筛选．I_{QR}通过计算数据集第一四分位数(Q_1)和第三四分位数(Q_3)，确定数据集的中间50%范围，即：

$$I_{QR}=Q_3-Q_1. \tag{1}$$

通常采用以下公式确定异常值的下界L和上界U为：

$$L=Q_1-1.5\times I_{QR}. \tag{2}$$

$$U=Q_3+1.5\times I_{QR}. \tag{3}$$

1.3.3 类不平衡处理

少数类样本(如3类)往往数量较少，而多数类样本(如2类)相对较多，这种样本数量的不平衡性会导致模型在训练过程中对多数类过度敏感，而忽视了少数类的特征．尤其针对少数类样本进行预测时，少数类别的重复复制可能导致对该类样本的过拟合．

为解决数据集的不平衡问题，文献[28]提出了一种SMOTE算法的过采样技术，即在保留原有少数类样本基础上，通过人工合成新样本来增加少数类样本数量，实现数据不同类别的平衡．SMOTE过采样策略的原理是基于KNN算法，如图2所示．具体步骤如下：

1)确定近邻数：选择合理K值，确定每个少数类样本的K个最近邻样本．

2)计算最近邻样本：采用K近邻算法计算K个最近邻样本，即哪些样本被用于合成新样本．

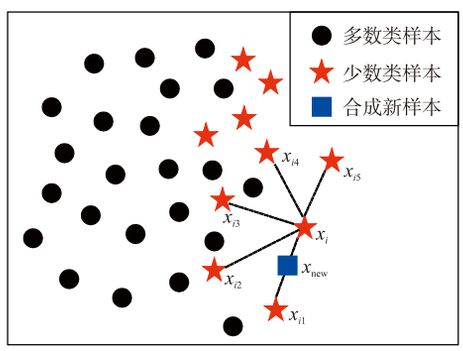
3)合成新样本：从每个少数类的K个最近邻中随机选择N个样本进行插值．对于原始样本x_i和其最近邻样本x_{ij}，新样本x_{new}的合成表达式为：

$$x_{new}=x_i+(x_{ij}-x_i)\cdot r, \tag{4}$$

式中 r为0～1之间的随机数．

4)重复插值：重复步骤3)，直到合成足够数量的新样本，使得少数类和多数类数量达到平衡．

5)合并数据集：将合成的新样本与原始数据集进行合并，形成一个新的平衡数据集．



注：x_i为少数类中的1个原始样本；x_{i1}～x_{i5}表示少数类样本x_i按距离排序的前5个近邻样本．

图2 SMOTE过采样方法及改进策略[21]

Fig. 2 Improvement strategies for SMOTE oversampling

通过SMOTE合成引入新样本，增强了模型对少数类样本的识别能力，解决了数据集的不平衡问题．然而，其也可能导致过拟合，特别是在合成样本过多的情况下，需综合考虑合成新样本的数量，以确保模型的泛化能力．

针对类不平衡问题，平衡方法包括过采样和欠采样策略．欠采样是指在每次采样中多数类的样本数必须等于少数类的样本数．过采样则通过增加少数类样本数量来实现类别平衡．传统SMOTE算法通过在少数类样本之间进行线性插值来合成新样本，但由于其主要关注少数类样本，可能导致新合成的样本过于集中，无法有效捕捉少数类样本的多样性，易在样本的类边界处产生重叠数据．

基于此，本文提出了一种基于K-means聚类的改进SMOTE算法，采用K-means聚类算法结合SMOTE过采样以重新平衡数据集．利用K-means聚类算法对少数类样本进行聚类，然后在每个聚类内部应用SMOTE算法合成新样本，以更好捕捉少数类样本的内部差异．具体步骤如下：

1)聚类阶段：使用K-means聚类算法对少数类样本进行聚类，从M个样本中随机选择k个对象作为初始聚类中心．计算每个样本与聚类中心的距离，将样本分配给最近的聚类中心．重复分配，直至聚类中心不再发生变化或达到预设迭代

次数.

聚类过程的数学表达式[29]为：

$$J(\boldsymbol{c},\boldsymbol{u})=\min\sum_{i=1}^{M}\left\| x_i-u_{c_i}\right\|^2, \tag{5}$$

式中：$J(\boldsymbol{c},\boldsymbol{u})$为目标函数，表示所有样本与其对应聚类中心的距离平方和；$\boldsymbol{u}$为聚类中心向量；u_{c_i}为第c_i个聚类的中心；$\boldsymbol{c}$为所有样本的聚类归属向量；c_i为样本x_i所属的聚类编号；M为样本总数.

2)过滤阶段：评估每个簇的类分布特征，保留具有高比例的少数类样本的簇，剔除类分布不均衡的簇，将更多样本分配给少数样本稀疏的群集.

3)过采样阶段：对少数类样本进行过采样. 在每个选定的簇中采用SMOTE算法，在聚类内部的样本之间线性插值合成新的样本.

鉴于模型在不平衡数据集上训练时可能会偏向于多数类，从而忽视少数类的特征，提出了改进K-means SMOTE方法来处理这种不平衡性. 类平衡前后样本预处理的数据分布特征如图3所示.

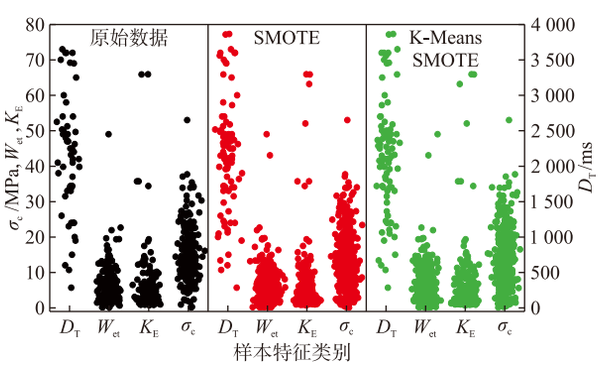


图3　完整样本特征的数据统计分布

Fig. 3　Data statistical distribution of complete sample features

经K-means SMOTE预处理方法进行数据处理后，扩充数据统计结果如表4所示. 由原始的221例(如1.2节所述，3类样本数比例为1.00∶4.80∶3.04)扩充至360例，其中每类样本均为120例，实现了数据样本的类别平衡.

表4　类平衡前后样本数量的比值

Table 4　Sample features ratio before/after data cluster balance

数据集	各标签的样本数量/例			训练集3类样本数之比	测试集3类样本数之比
	1	2	3		
原始数据	25	120	76	20∶96∶61	5∶24∶15
SMOTE算法	120	120	120	96∶96∶96	24∶24∶24
K-Means SMOTE算法	120	120	120	96∶96∶96	24∶24∶24

在此基础上，采用分层抽样策略，按7∶3的比例划分训练集与测试集，其中训练集共252例(每类84例)，测试集共108例(每类36例)，确保训练集和测试集中均保持各标签样本数1∶1∶1的类别平衡.

2　研究方法

2.1　预测模型算法基础

2.1.1　SVM理论基础

支持向量机(SVM)作为一种有监督的机器学习模型，广泛应用于分类决策问题研究，其核心思想在于利用有限的样本信息，在特征空间中找到一个最优超平面，该超平面能够最大化地分开不同类别数据点. 通过最大化2个类别之间的间隔，使得最近邻点(支持向量)到超平面的距离最大化.

在分类识别问题中，对于线性可分数据，SVM通过寻找一个线性超平面，使得不同类别的数据点被有效分隔开. 对于非线性可分的数据，SVM通过核函数将数据映射到高维空间，在高维空间中寻找最优超平面，使得样本数据线性可分. 该方法基于统计学理论基础，直接建立了特征指标与冲击倾向性等级之间的非线性映射关系，避开了指标权重的确定. SVM基本优化问题[13]如下：

$$\min_{w,b,\xi}\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{m}\xi_i. \tag{6}$$

受以下条件约束：

$$y_i\left[\boldsymbol{w}\cdot\phi(x_i)+b\right]\geqslant 1-\xi_i,\quad \xi_i\geqslant 0,\quad i=1,\cdots,m, \tag{7}$$

式中：$\boldsymbol{w}$为权重向量；b为偏置项；ξ_i为松弛变量；C为惩罚参数；y_i为第i个样本的类别标签；$\phi(x_i)$为将原始数据映射到高维特征空间的函数；m为样本总数.

对于多分类问题，SVM的RBF决策函数为：

$$K(x_i,x_j)=\exp(-\gamma\left\| x_i-x_j\right\|^2), \tag{8}$$

式中：$K(x_i,x_j)$为径向基核函数(RBF)；γ为核函数参数；x_i,x_j为任意2个样本的特征向量；$\left\| x_i-x_j\right\|$为样本x_i与x_j之间的欧氏距离.

其中，惩罚参数C控制着模型对误分类的惩罚程度，而γ决定了核函数的宽度，两者共同影响着决策边界的确定. 通过不同策略搜索最佳的C和γ参数对模型泛化能力和预测性能至关重要.

2.1.2　新型元启发式CNN优化算法

新型元启发式算法(又称捕鱼优化算法，简写为CFOA算法)通过搜索代理不断更新位置来探

索全局情况，从而找到最佳的解决方案，元启发则相当于“海洋”，算法模型如图 4 所示.

CFOA 算法的主要步骤包括：

1）初始化：在搜索空间中随机分布一定数量的“鱼”（潜在解）.

2）撒网：捕鱼者（算法）在“海洋”中撒网，每个网对应一个捕鱼者.

3）评估：评估网中鱼的质量（目标函数值），质量最好的鱼被认为是当前的全局最佳解.

4）更新位置：根据捕鱼者的经验和当前捕获的鱼的质量，更新捕鱼者的位置和网的大小.

5）迭代：重复上述过程，直到满足停止条件，如最大迭代次数或达到预定的解质量.

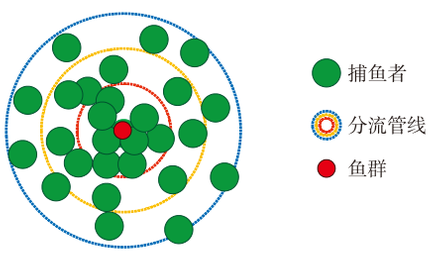


图 4 捕鱼优化算法模型示意[30]

Fig. 4 Fishing optimization algorithm model diagram

CFOA 算法优化通过动态调整捕鱼者的位置和网的大小，从而提高捕获高质量鱼（优质解）的概率. CFOA 元启发式算法用于优化 CNN-SVM 的参数，包括惩罚参数 C 和核函数参数 γ. 其中，C 控制分类器的间隔大小，γ 控制径向基函数（RBF）核的密度. 模型通过训练卷积神经网络（CNN）来提取特征，并使用智能优化算法来调整 SVM 的超参数，以达到最佳的分类效果.

2.1.3 对比 SVM 组合优化分类模型

1）牛顿拉夫森优化算法

牛顿拉夫森优化算法（NRBO）是一种新型的元启发式优化策略，核心机制包括 NRSR 搜索规则和陷阱规避算子（TAO）. NRSR 利用牛顿-拉夫森迭代法来增强 NRBO 探索能力，并加速其收敛过程. TAO 帮助 NRBO 规避陷入局部最优解. 该方法具有搜索速度快及寻优能力强等特点[31].

2）ISSA-SVM 和 KPCA-ISSA-SVM 策略

多策略麻雀算法（ISSA）通过模拟麻雀的觅食行为，为 SVM 超参数优化提供了一种有效的搜索策略. 核主成分分析（KPCA）与 ISSA 结合方法，通过降低数据维度，在保留数据核心特征的同时，提供了更高质量的输入参数特征.

3）QPSO-SVM 和 ASL-QPSO-SVM 策略

量子粒子群优化算法（QPSO）通过模拟量子行为和粒子群动态，利用量子力学原理来引导粒子群的搜索过程. 改进量子粒子群自适应算法（ASL-QPSO），通过引入自适应机制，能够根据搜索过程的进展动态调整搜索策略[32].

2.2 基于最优化 SVM 的分类预测模型

本文构建了基于元启发式算法 CFOA 策略优化超参数的 CNN-SVM 预测模型，提出了基于“数据预处理-组合模型优化-数据指标验证”的三步式冲击倾向性等级预测方法.

模型采用 D_T，W_{et}，K_E，σ_c 这 4 个指标作为输入特征参数，基于原始 221 组样本数据集，通过所提 K-means SMOTE 过采样方法扩充至总计 360 组数据集样本；通过空值填充、异常值剔除、重采样与聚类合成新样本等数据预处理，解决了数据集的不均衡问题，增强了模型对不平衡数据的处理能力. 类平衡前后数据集均按照 70% 和 30% 比例划分训练集和测试集，其中训练集用于模型确定和超参数优化，测试集则用于评估模型性能，训练集和测试集之间相互独立.

为了验证不同组合分类模型的预测分类效果，采用分类问题中的模型性能评估指标进行验证，以衡量评估预测模型的泛化能力. 这些指标包括准确度 A、精度 P、召回率 R 和 F_1 值，计算方法如图 5 所示.

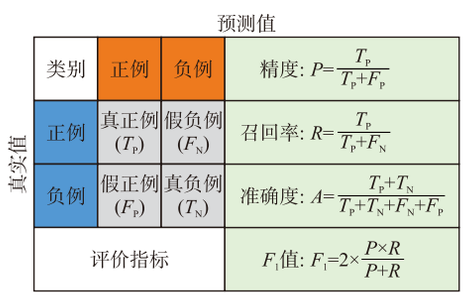


图 5 评价指标的计算

Fig. 5 Calculation of evaluation index

1）准确度：表示在模型中预测分类正确的样本数占总样本数的比例.

2）精度：表示模型预测为正类且实际为正类的样本数占所有预测为正类的样本数的比例.

3）召回率：表示模型正确预测为正类的样本数占所有实际正类样本数的比例.

4）F_1 值：F_1 值是精度和召回率的调和平均数，表征精度和召回率的平衡性能.

3　结果分析与讨论

3.1　数据预处理效果

为了证实CFOA-CNN-SVM模型分类预测结果的可靠性，采用9种基于SVM的不同组合策略预测模型，对预测效果进行对比分析与验证．基于此，构建了3类共9种组合分类模型，包括SVM基础模型、3种超参数优化模型（ISSA-SVM，QPSO-SVM和NRBO-SVM）、2种数据特征处理和超参数优化模型（KPCA-ISSA-SVM和ASL-QPSO-SVM）、2种新型元启发式算法超参数优化模型（CFOA-CNN-SVM与NRBO-CNN-SVM）．

本文采用9种不同的优化组合分类模型对原始数据集、SMOTE数据集及K-means SMOTE生成数据集进行预测．将构建的3组数据集按7∶3的比例划分为训练集和测试集，随后将划分好的70％训练数据集分别输入到9个分类模型中进行训练．

首先，为了验证所提数据预处理方法SMOTE和K-means SMOTE策略的效果，以模型预测精度为评价指标，对比分析了9种模型对原始数据库和预处理后数据的预测精度，见表5．

表5　不同数据预处理方法对模型预测精度的影响

Table 5　The influence of different data preprocessing methods on the prediction accuracy of the model

模型编号	预测模型	原始数据预测精度			SMOTE预处理数据预测精度			K-means SMOTE预处理数据预测精度		
		训练集	测试集	排名	训练集	测试集	排名	训练集	测试集	排名
1	SVM	0.75	0.71	9	0.77	0.73	9	0.86	0.84	9
2	ISSA-SVM	0.88	0.74	8	0.88	0.79	7	0.95	0.84	8
3	KPCA-ISSA-SVM	0.90	0.77	7	0.91	0.81	4	0.96	0.88	5
4	QPSO-SVM	0.82	0.79	6	0.86	0.80	6	0.93	0.86	7
5	ASL-QPSO-SVM	0.83	0.79	5	0.87	0.81	5	0.94	0.88	4
6	CNN-SVM	0.89	0.82	2	0.91	0.86	3	0.97	0.89	3
7	CFOA-CNN-SVM	0.92	0.82	1	0.93	0.91	1	0.97	0.96	1
8	NRBO-CNN-SVM	0.95	0.80	3	0.96	0.87	2	0.97	0.94	2
9	NRBO-SVM	0.86	0.80	4	0.88	0.77	8	0.94	0.86	6

分析可知：SVM模型在所有模型中预测精度最低，在原始数据训练和测试集中的预测精度分别为0.75和0.71；经SMOTE处理后，训练集和测试集精度分别为0.77和0.73，较原始数据分别提高了2.14％和1.94％；经K-means SMOTE策略处理后预测精度达到0.86和0.84，较原始数据分别提高了11.27％和13.05％．CFOA-CNN-SVM在所有模型中的识别精度最高，在原始数据训练集和测试集中预测精度分别为0.92和0.82；经SMOTE处理后，训练集和测试集预测精度达到0.93和0.91，较原始数据分别提高了0.60％和8.92％；经K-means SMOTE方法处理后训练集和测试集预测精度分别达到0.97和0.96，较原始数据提高了4.96％和14.48％．

经K-means SMOTE方法数据预处理后，SVM基础模型训练样本预测精度为0.86，而常规超参数优化机器学习模型（QPSO-SVM，ISSA-SVM）的预测精度为0.93～0.95．基于数据降维处理和超参数优化的机器学习模型（KPCA-ISSA-SVM，ASL-QPSO-SVM）预测精度分别达0.96和0.94；CNN-SVM模型预测精度为0.97；超参数优化CNN与SVM组合分类模型预测精度均达到0.97．

如图6所示，不同模型对3种不同预处理数据集的预测精度存在差异，表明采用数据集类别平衡算法进行数据预处理，能有效提升模型预测精度．本文所提K-means SMOTE算法在多数情况下均表现出较高准确率，总体性能优于SMOTE算法和未经处理的原始数据集．除在NRBO-CNN-SVM模型中预测精度未有显著提升外，在其他8种分类模型中，经过K-means SMOTE策略处理的数据集预测精度均明显优于传统SMOTE算法．

综上，针对数据集缺失值、异常值和类不平衡问题，采用基于K-means改进SMOTE算法进行空值填充、异常值剔除、重采样与聚类合成新样本等数据处理，解决了数据集不均衡问题，经预处理的数据集在多种优化分类模型下均能获得较优预测精度．

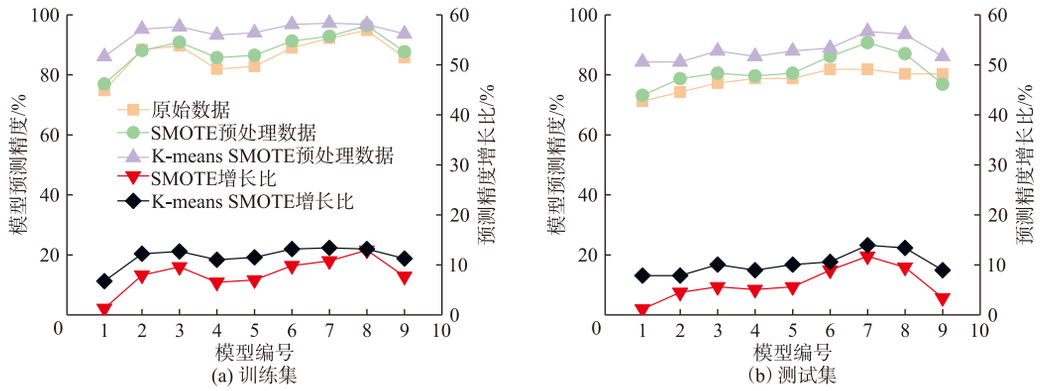


图 6 3类预处理数据集的模型预测精度

Fig. 6 The model prediction accuracy of the data sets of the three types of preprocessing

3.2 使用9种机器学习算法预测冲击倾向性等级

将剩余30%测试集数据导入9种机器学习算法训练的分类模型进行模型有效性测试与验证，8种不同组合分类模型预测结果混淆矩阵如图7所示，不同组合分类模型的预测指标评价见表6.

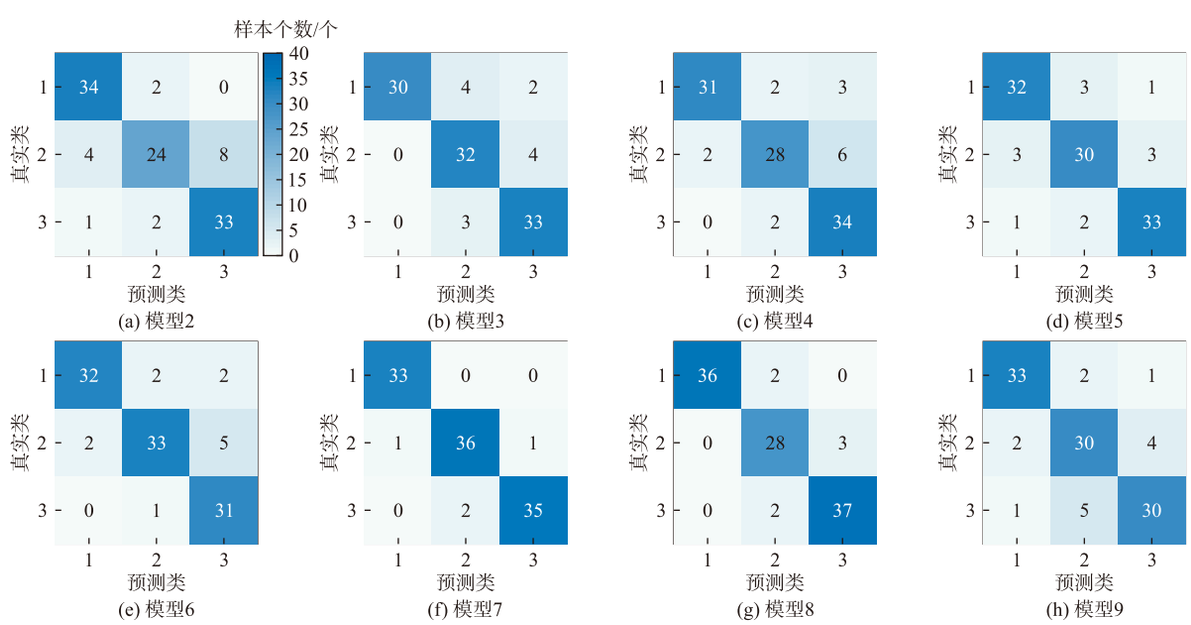


图 7 不同组合分类模型预测结果的混淆矩阵

Fig. 7 Confusion matrix of prediction results of different combined classification models

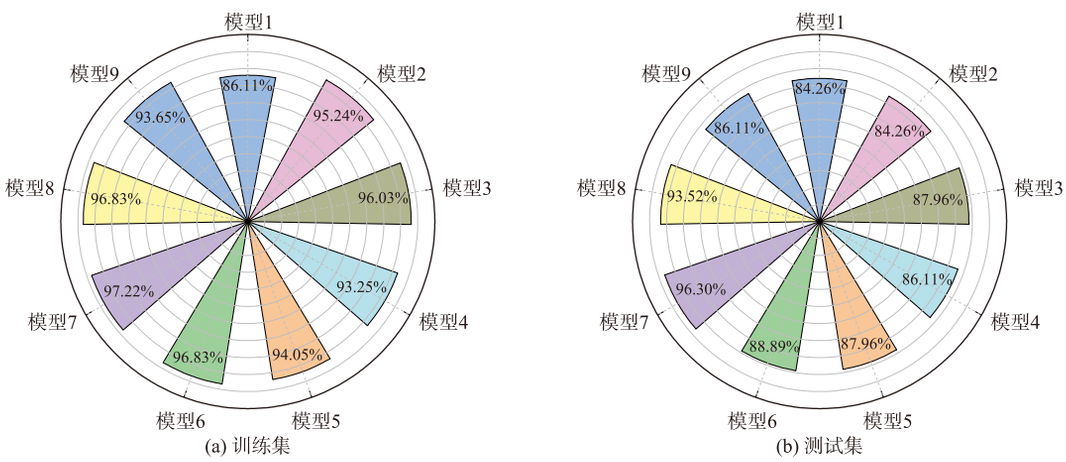
表 6 不同组合分类模型中的预测指标评价

Table 6 Evaluation of predictors in different combined classification models

模型编号	预测模型	准确度/%		各标签样本预测精度			各标签样本召回率			各标签样本F_1值		
		训练集	测试集	1	2	3	1	2	3	1	2	3
1	SVM	86.11	84.26	0.94	0.81	0.79	0.89	0.72	0.92	0.91	0.77	0.85
2	ISSA-SVM	95.24	84.26	0.87	0.86	0.81	0.94	0.67	0.92	0.91	0.75	0.86
3	KPCA-ISSA-SVM	96.03	87.96	1.00	0.82	0.85	0.83	0.89	0.92	0.91	0.85	0.88
4	QPSO-SVM	93.25	86.11	0.94	0.88	0.79	0.86	0.78	0.94	0.90	0.82	0.86
5	ASL-QPSO-SVM	94.05	87.96	0.89	0.86	0.89	0.89	0.83	0.92	0.89	0.85	0.90
6	CNN-SVM	96.83	88.89	0.94	0.92	0.82	0.89	0.83	0.97	0.91	0.87	0.89
7	CFOA-CNN-SVM	97.22	96.30	0.97	0.95	0.97	1.00	0.95	0.95	0.99	0.95	0.96
8	NRBO-CNN-SVM	96.83	93.52	1.00	0.88	0.93	0.95	0.90	0.95	0.97	0.89	0.94
9	NRBO-SVM	93.65	86.11	0.92	0.81	0.86	0.92	0.83	0.83	0.92	0.82	0.85

分析可知：对于测试集而言，SVM 模型预测准确度为 84.26%，经过常规超参数优化的机器学习模型（QPSO-SVM，ISSA-SVM）的预测准确度为 84.26%～86.11%，而 KPCA-ISSA-SVM，ASL-QPSO-SVM 模型预测准确度为 87.96%，CNN-SVM 模型预测准确度为 88.89%，元启发式 CFOA-CNN-SVM，NRBO-CNN-SVM 优化模型预测准确度分别达 96.30% 与 93.52%. 其中，元启发式智能优化算法超参数优化 CNN 模型对模型参数进行精细调优，显著提高了预测的精确度和鲁棒性. 对比分析其他 8 种改进的优化 SVM 可知，CFOA-CNN-SVM 模型在测试集中准确度最高，达 96.30%，在不同组合分类模型中的预测性能最优.

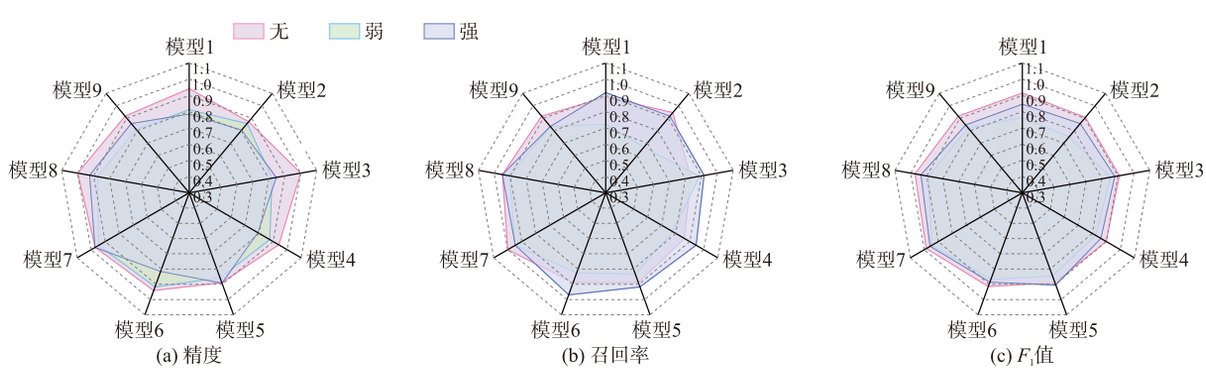
不同分类预测模型的指标评价结果如图 8～9 所示. 分析可知：基础 SVM 模型对训练集、测试集的预测准确度分别为 86.11% 和 84.26%，对于无、弱、强冲击倾向性的数据样本，精度分别为 0.94，0.81 和 0.79，召回率分别为 0.89，0.72 和 0.92，F_1 值分别为 0.91，0.76 和 0.85. CFOA-CNN-SVM 模型对训练集、测试集的预测准确度分别达 97.22% 和 96.30%. 对于无、弱、强冲击倾向性的数据样本，CFOA-CNN-SVM 模型的预测精度分别为 0.97，0.95 和 0.97，召回率分别为 1.00，0.95 和 0.95，F_1 值分别为 0.99，0.95 和 0.96. 在测试数据集中，相较于 SVM 基础模型，CFOA-CNN-SVM 模型的预测准确度提高了 12.04%；此外，对于 3 种冲击倾向性数据样本的预测精度分别提升 0.03，0.14 和 0.18，召回率分别提高 0.11，0.23 和 0.03，F_1 值分别增加 0.08，0.19 和 0.11. 与其他 8 种模型相比，CFOA-CNN-SVM 分类预测模型在冲击倾向性等级识别方面优势显著.



(a) 训练集　　(b) 测试集

图 8　测试集中不同组合模型的预测准确度

Fig. 8　The prediction accuracy of different combination classification models in the test set



(a) 精度　　(b) 召回率　　(c) F_1值

图 9　测试集中不同组合分类模型预测多指标评价

Fig. 9　Multi-index evaluation of different combination classification models in test set

4　实际案例验证

4.1　面向不完整样本数据的案例效果验证

为了验证所提方法在不完整样本（如特征缺失统计数据）条件下模型的可靠性，在前文 221 组原始数据集的基础上，新增了 49 组含缺失特征数据（如 σ_c 参数）的样本数据集进行效果验证，结果见表 7.

表7 不完整样本数据的分类预测案例验证

Table 7 Verification of classification prediction effect of incomplete sample data

序号	输入特征参数			分类等级	标签
	D_T/ms	W_{et}	K_E		
222	28.00	4.80	75.00	强	3
223	537.00	2.08	2.38	弱	2
224	5.21	5.20	126.00	强	3
225	409.00	2.16	1.39	弱	2
226	3 894.00	1.22	1.80	无	1
⋮	⋮	⋮	⋮	⋮	⋮
266	21.60	8.20	87.00	强	3
267	318.00	3.60	0.90	无	1
268	77.00	5.80	4.60	弱	2
269	31.00	4.30	5.30	强	3
270	46.00	8.40	7.90	强	3

在新生成的σ_c特征参数缺失案例验证数据集中，原始数据集样本总数为270例．其中，1类数据占比11.9%(32例)，2类数据占比50.0%(135例)，3类数据占比38.1%(103例)．

采用改进K-means算法聚类合成少数类的过采样方法，即K-means SMOTE采样策略，进行数据清洗、分类与处理，解决样本缺失值、异常值和类比例失衡问题．经过K-means SMOTE方法进行数据预处理后，数据集样本案例数量从270例增加到405例，其中，无和强冲击倾向性样本分别增加了103组和32组，1，2，3类样本数比例为135∶135∶135．

针对不完整样本案例的数据预处理，K-means SMOTE策略不仅能够有效处理数据集中的缺失值和异常值，还通过合成新的少数类样本调整了样本的类比例失衡问题，提高了模型对少数类的预测能力．在该案例中，K-means SMOTE算法的应用使得数据集的样本数量从270例增加至405例，增强了预测模型训练效果的可靠性．

按70%和30%比例划分训练集和测试集，采用最优组合分类模型(CFOA-CNN-SVM)进行分类预测结果验证．将30%测试集案例数据置入优化模型中进行有效性测试和验证，结果如表8所示．

表8 面向不完整样本统计数据的模型分类预测效果

Table 8 Model classification prediction effect for incomplete sample statistical data

数据	各标签样本比例	数据集类型	准确度/%	各标签样本预测精度			各标签样本召回率			各标签样本F_1值		
	1∶2∶3			1	2	3	1	2	3	1	2	3
原始数据	25∶120∶76	训练	88.36	0.92	0.85	0.93	0.50	0.95	0.93	0.65	0.90	0.93
		测试	79.01	0.86	0.77	0.85	0.75	0.88	0.69	0.80	0.82	0.75
SMOTE预处理	135∶135∶135	训练	93.66	0.94	0.92	0.95	0.92	0.91	0.98	0.93	0.92	0.96
		测试	85.12	0.83	0.79	0.95	0.81	0.84	0.90	0.82	0.81	0.92
KM-SMOTE预处理	135∶135∶135	训练	94.72	0.97	0.93	0.95	0.94	0.93	0.97	0.96	0.93	0.96
		测试	91.74	0.97	0.95	0.94	0.88	0.88	0.87	0.90	0.96	0.93

分析可知：对于不完整数据样本而言，最优组合分类模型预测测试准确度为79.01%；而经SMOTE处理后预测准确度达85.12%，经K-means SMOTE处理后预测准确度达91.74%，相较于未经处理的数据样本预测准确度分别提高了6.11%和12.73%．这表明经K-means SMOTE处理的数据集在模型中的辨识度提升显著．对于无、弱、强冲击倾向数据样本，优化组合分类模型的平均精度、召回率、F_1值分别达0.94，0.88和0.93．

4.2 矿井侧小样本试验数据效果验证

为了验证所提方法在煤层冲击倾向性实际预测过程中的准确性．在第3节已经训练好的预测模型(模型原始数据集对应为360组)基础上，重新引入了10组全新的煤样真实试验数据，并将这些试验小样本数据导入到已训练好的模型中，进行未知试验数据的预测．基于机器学习的煤层煤样冲击倾向性分类结果验证方法如图10所示．图中，$\boldsymbol{X}_c$为第c个搜索代理的当前位置向量(c=1，2，…，N)，N为搜索代理总数；$\boldsymbol{X}'_c$为更新后的新位置向量；箭头“→”表示位置更新过程．

试验样品取自陕西省永陇矿区某矿3#煤层22盘区和23盘区，该矿采用厚煤层综采放顶煤工作面布置方式，煤层开采厚度为12 m．依据实地调研，煤层具有弱冲击倾向性．按照国标方法开展岩石力学试验，获得了10组全新的煤层真实试验数据．

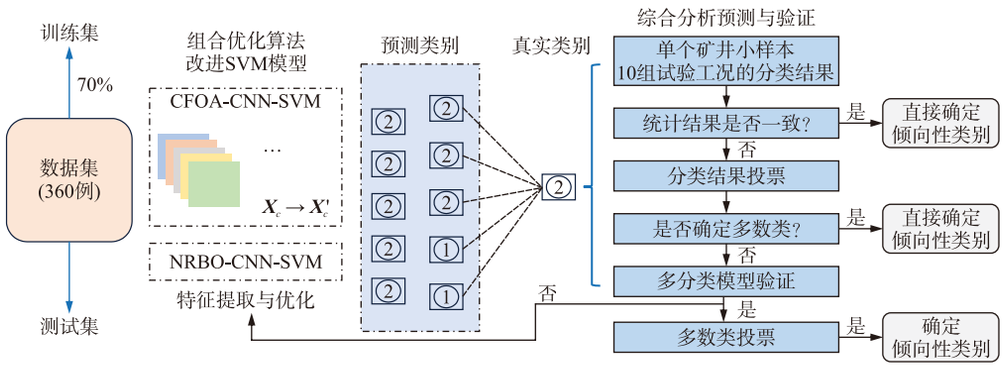


图 10　基于机器学习的煤样冲击倾向性分类结果验证

Fig. 10　Verification of coal bursting liability classification results of coal samples based on machine learning

面对多指标分布在多个等级的复杂情况时，采用标准中推荐的综合判定法，存在 8 种组合情况难以判定的问题. 基于此，本文提出了一种适用于矿井侧小样本、未知真实试验数据的优化模型及其验证方法，通过收集或试验有限数据库样本，实现矿井侧复杂煤层煤样冲击倾向性的高效精确分类预测，验证结果见表 9.

依据 10 组真实煤样案例冲击倾向性预测投票结果，性能最优的机器学习模型将这 10 组煤样分为了 2 类(80%)、1 类(20%)，其中 2 类为多数类标签，无需进一步进行多个模型的验证投票. 综合判定这 10 组真实煤样的冲击倾向性预测结果为 2 类，对应煤层回采过程中可能发生轻微冲击地压现象，与现场情况基本吻合.

表 9　单个矿井小样本真实试验数据案例下的冲击倾向性预测结果

Table 9　Prediction results of coal bursting liability under small sample test cases of single mine

煤层取样位置	案例	D_T/ms		W_{et}		K_E		σ_c/MPa		综合指标评估	机器学习方法
		数值	等级	数值	等级	数值	等级	数值	等级		
22 盘区	1	294	2	2.76	2	2.71	2	13.40	2	2	2
	2	288	2	3.36	2	1.83	2	12.40	2	2	2
	3	448	2	2.46	2	1.88	2	15.32	3	2	2
	4	384	2	0.73	1	7.19	3	16.21	3	3	2
	5	437	2	2.60	2	6.31	3	13.92	2	2	2
23 盘区	6	272	2	3.15	2	7.21	3	19.08	3	*	2
	7	238	2	3.21	2	5.36	3	16.22	3	*	2
	8	158	2	1.84	1	2.70	2	15.32	3	2	2
	9	402	2	0.79	1	7.33	3			2/3	1
	10	300	2			3.92	2			2/3/*	1
3#煤层冲击倾向性综合评估与预测结果										2	2

注：“*”表示依据现行标准无法明确判定等级.

综上，提出基于“数据预处理-组合模型优化-数据指标验证”的三步式冲击倾向性分类预测方法，通过统计收集或现场试验的有限数据样本进行精准预测. 该方法针对全新小样本数据也能维持较高预测准度，是对国标综合判定规则的智能化扩展及验证，增强了其在处理边界模糊案例时的应用能力，在面对厚/特厚复合软硬煤层的整体冲击倾向性界定等复杂预测问题时，也能够被可靠、有效地应用验证.

4.3　讨论

与表征煤岩介质固有属性的“冲击倾向性”指标不同，“冲击危险性”多以巷道为评估对象，特指真实地质条件与采动应力环境中发生冲击地压的可能性与潜在破坏强度. 根据评估对象差异，可分为“采区冲击危险性评价”、“回采工作面冲击危险性评价”及“掘进巷道冲击危险性评价”等[33].

明确煤岩介质冲击倾向性指标是评价采矿工程岩体冲击危险性的内在基础与前提条件. 现有研究表明,冲击地压发生的临界载荷 P_{cr} 主要受煤岩单轴抗压强度 σ_c 与其冲击倾向性指数 K 的影响. 文献[34]对冲击地压发生的力学模型进行解析,给出了冲击地压发生的临界载荷理论公式:

$$\frac{P_{cr}}{\sigma_c}=\frac{1}{2}\left(1+\frac{1}{K}\right), \tag{9}$$

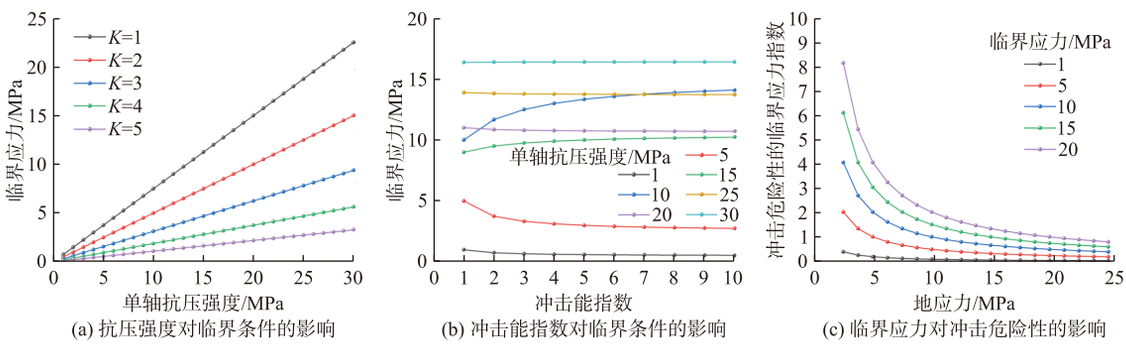
式中 K 表征煤岩介质的冲击倾向性,与测试冲击能量指数 K_E 本质相同.

依据前文统计数据确定冲击倾向性指数 K 与抗压强度 σ_c 取值范围,据此计算得到临界冲击载荷. 基于此,将冲击危险性定义为该处岩体的实际地应力 P 与经修正的临界载荷 $\bar{P}_{cr}$ 之间的相对关系为:

$$K_{cr}=\frac{P}{\bar{P}_{cr}}, \tag{10}$$

式中 K_{cr} 为表征冲击危险性的临界应力指数. 实际地应力 P 越高,临界应力指数 K_{cr} 越大,冲击危险性越高,越可能会发生冲击地压现象.

研究统计表明,多数煤层埋深为 100~1 000 m. 基于相对应的实际地应力数据,换算得到其临界应力指数 K_{cr} 的分布范围. 以本文验证数据为例,引入冲击危险性指标并进行相应换算,揭示冲击倾向性与冲击危险性间的关联,如图 11 所示.



(a) 抗压强度对临界条件的影响 (b) 冲击能指数对临界条件的影响 (c) 临界应力对冲击危险性的影响

图 11 冲击倾向性指标与冲击危险性的理论关系验证

Fig. 11 Verification of the theoretical relationship between coal bursting liability index and impact risk

分析可知:抗压强度对临界应力的影响呈线性增长趋势,K 值越小增幅越显著,临界应力水平越高. 冲击能指数对临界应力的影响呈非线性变化特征,在 K=1~5 范围内,临界应力呈缓慢增加趋势,超过该范围后趋于总体稳定;强度越大临界应力整体越高. 临界应力与冲击危险性呈负相关关系,随地应力增大临界应力指数逐渐降低,临界应力越高时冲击危险性越低,越不易发生冲击地压.

5 结 论

1)构建了含 221 组样本案例的冲击倾向性数据集,针对其缺失值、异常值和类不平衡问题,采用基于 K-means 的改进 SMOTE 算法进行空值填充、异常值剔除、重采样与聚类合成新样本等数据预处理,解决了数据集的不均衡问题,增强了模型对少数类样本的识别能力,识别精度最大提高了 14.48%.

2)建立了智能元启发式算法 CFOA 策略优化超参数的 CNN-SVM 预测模型,并提出了基于“数据预处理-组合模型优化-数据指标验证”的三步式冲击倾向性等级预测方法. 与其他 8 种优化 SVM 模型对比,采用 K-means SMOTE 技术进行数据预处理后的优化 CFOA-CNN-SVM 模型的准确度、召回率等评估指标均表现优异,准确度最高达 96.30%,表现出较强的分类预测性能与泛化能力.

3)实际案例验证表明,本文提出的方法在面向不完整、样本量有限的复杂预测场景中仍能保持较高的预测性能. 不完整数据下(270 例)预测准确度达 91.74%,相比数据未经预处理时提高了 12.73%;提出基于现有训练模型的矿井侧小样本数据预测方法进行验证,通过引入 10 组真实测试数据,导入已训练好的模型(221 例)进行预测,综合判定为 2 类,结果基本吻合.

参考文献:

[1] 窦林名,田鑫元,曹安业,等. 我国煤矿冲击地压防治现状与难题[J]. 煤炭学报,2022,47(1):152-171.
DOU Linming, TIAN Xinyuan, CAO Anye, et al. Present situation and problems of coal mine rock burst prevention and control in China [J]. Journal of China Coal Society, 2022, 47(1): 152-171.

[2] 潘俊锋,齐庆新,刘少虹,等. 我国煤炭深部开采冲击

地压特征、类型及分源防控技术[J].煤炭学报,2020,45(1):111-121.
PAN Junfeng, QI Qingxin, LIU Shaohong, et al. Characteristics, types and prevention and control technology of rockburst in deep coal mining in China[J]. Journal of China Coal Society, 2020, 45(1): 111-121.

[3] 齐庆新,李一哲,赵善坤,等.我国煤矿冲击地压发展70年:理论与技术体系的建立与思考[J].煤炭科学技术,2019,47(9):1-40.
QI Qingxin, LI Yizhe, ZHAO Shankun, et al. Seventy years development of coal mine rockburst in China: Establishment and consideration of theory and technology system[J]. Coal Science and Technology, 2019, 47(9): 1-40.

[4] 钱七虎.岩爆、冲击地压的定义、机制、分类及其定量预测模型[J].岩土力学,2014,35(1):1-6.
QIAN Qihu. Definition, mechanism, classification and quantitative forecast model for rockburst and pressure bump[J]. Rock and Soil Mechanics, 2014, 35(1): 1-6.

[5] CHEN C, ZHOU J. A new empirical chart for coal burst liability classification using Kriging method[J]. Journal of Central South University, 2023, 30(4): 1205-1216.

[6] 温廷新,于凤娥.基于NRS-ACPSO-SVM的冲击地压危险性预测模型[J].中国安全科学学报,2017,27(10):19-25.
WEN Tingxin, YU Fenge. NRS-ACPSO-SVM based model for prediction of rock burst risk[J]. China Safety Science Journal, 2017, 27(10): 19-25.

[7] 张　曼,陈建宏,周智勇.基于SVM的冲击地压分级预测模型及R语言实现[J].中国地质灾害与防治学报,2018,29(4):64-69.
ZHANG Man, CHEN Jianhong, ZHOU Zhiyong. Grading prediction model of rock burst based on SVM and its R language description[J]. The Chinese Journal of Geological Hazard and Control, 2018, 29(4): 64-69.

[8] 赵洪波,茹忠亮.冲击地压预测的PSO-SVM模型[J].岩石力学与工程学报,2007,26(增1):3479-3483.
ZHAO Hongbo, RU Zhongliang. PSO-SVM Model for prediction of rock burst[J]. Chinese Journal of Rock Mechanics and Engineering, 2007, 26(Sup 1): 3479-3483.

[9] SUN J H, WANG W J, XIE L K. Predicting short-term rockburst intensity using a weighted probability stacking model with optimal feature selection and Bayesian hidden layer[J]. Tunnelling and Underground Space Technology, 2024, 153: 106021.

[10] PU Y Y, APEL D B, WANG C, et al. Evaluation of burst liability in kimberlite using support vector machine[J]. Acta Geophysica, 2018, 66(5): 973-982.

[11] 姚　囝,张义礼,刘　洋,等.基于PCA,CBLOF和SVMSMOTE算法组合的岩爆烈度等级分级预测[J].岩石力学与工程学报,2025,44(5):1230-1241.
YAO Nan, ZHANG Yili, LIU Yang, et al. Rock burst intensity grading prediction based on the combination of PCA, CBLOF and SVMSMOTE algorithms[J]. Chinese Journal of Rock Mechanics and Engineering, 2025, 44(5): 1230-1241.

[12] 田　睿.基于机器学习的岩爆烈度等级预测模型研究与应用[D].包头:内蒙古科技大学,2020.
TIAN Rui. Research and application of rockburst intensity classification prediction model based on machine learning algorithms[D]. Baotou: Inner Mongolia University of Science & Technology, 2020.

[13] ZHOU J, LI X B, SHI X Z. Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines[J]. Safety Science, 2012, 50(4): 629-644.

[14] LI M L, LI K G, QIN Q C. A rockburst prediction model based on extreme learning machine with improved Harris Hawks optimization and its application[J]. Tunnelling and Underground Space Technology, 2023, 134: 104978.

[15] 温廷新,王泽锋.基于K-means SMOTE和IDBO-RF岩爆烈度等级预测模型[J].中国安全生产科学技术,2024,20(6):140-146.
WEN Tingxin, WANG Zefeng. Prediction model of rockburst intensity levels based on K-means clustering optimization and improved dung beetle optimization algorithm[J]. Journal of Safety Science and Technology, 2024, 20(6): 140-147.

[16] 陈　结,高靖宽,蒲源源,等.冲击地压预测预警的机器学习方法[J].采矿与岩层控制工程学报,2021,3(1):57-68.
CHEN Jie, GAO Jingkuan, PU Yuanyuan, et al. Machine learning method for predicting and warning of rockbursts[J]. Journal of Mining and Strata Control Engineering, 2021, 3(1): 57-68.

[17] 窦林名,何学秋.煤矿冲击矿压的分级预测研究[J].中国矿业大学学报,2007,36(6):717-722.
DOU Linming, HE Xueqiu. Technique of classification forecasting rock burst in coal mines[J]. Journal of China University of Mining & Technology, 2007, 36(6): 717-722.

[18] WU M, YE Y C, WANG Q H, et al. Development of rockburst research: A comprehensive review[J]. Applied Sciences-Basel, 2022, 12(3): 974.

[19] LIU Q S, XUE Y G, LI G K, et al. Application of KM-SMOTE for rockburst intelligent prediction[J]. Tunnelling and Underground Space Technology, 2023, 138: 105180.

[20] ZHANG H W, XIA Y Y, LIN M Q, et al. A three-step rockburst prediction model based on data preprocessing combined with clustering and classification algorithms[J]. Bulletin of Engineering Geology and the Environment, 2024, 83(7): 266.

[21] XUE Y G, LI G K, LI Z Q, et al. Intelligent prediction of rockburst based on Copula-MC oversampling architecture[J]. Bulletin of Engineering Geology and the Environment, 2022, 81(5): 209.

[22] LI J, FU H L, HU K X, et al. Data preprocessing and machine learning modeling for rockburst assessment[J]. Sustainability, 2023, 15(18): 13282.

[23] BASNET P M S, MAHTAB S, JIN A, et al. A comprehensive review of intelligent machine learning based predicting methods in long-term and short-term rockburst prediction [J]. Tunnelling and Underground Space Technology, 2023, 142: 105434.

[24] 鞠文君,卢志国,高富强,等.煤岩冲击倾向性研究进展及综合定量评价指标探讨[J].岩石力学与工程学报,2021, 40(9): 1839-1856.
JU Wenjun, LU Zhiguo, GAO Fuqiang, et al. Research progress and comprehensive quantitative evaluation index of coal rock bursting liability[J]. Chinese Journal of Rock Mechanics and Engineering, 2021, 40(9): 1839-1856.

[25] 郭金栋.变权重灰色归类识别模型在煤层冲击倾向性评价中的应用[J].华北科技学院学报,2017, 14(1): 44-49.
GUO Jindong. Application of variable weight and grey classification recognition model in rock burst tendency evaluation of coal seam[J]. Journal of North China Institute of Science and Technology, 2017, 14(1): 44-49.

[26] 潘俊锋,简军峰,刘少虹,等.黄陇侏罗纪煤田冲击地压地质特征与防治[J].煤矿开采,2019, 24(1): 110-115.
PAN Junfeng, JIAN Junfeng, LIU Shaohong, et al. Geological characteristic and control of rock burst of Huanglong Jurassic coal mine field[J]. Coal Mining Technology, 2019, 24(1): 110-115.

[27] 李岳峰.煤样冲击倾向性分类的多指标评判模型比较及优选[D].昆明:昆明理工大学,2022.
LI Yuefeng. Coal sample impact tendency classification multi-indicator evaluation model comparison and optimization[D]. Kunming: Kunming University of Science and Technology, 2022.

[28] CHAWLA N V, BOWYER K W, Hall L O, et al. SMOTE: Synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 321-357.

[29] GEORGIOS D, FERNANDO B, FELIX L. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE[J]. Information Sciences, 2018, 465: 1-20.

[30] JIA H M, WEN Q X, WANG Y H, et al. Catch fish optimization algorithm: a new human behavior algorithm for solving clustering problems[J]. Cluster Computing: The Journal of Networks Software Tools and Applications, 2024, 27(9): 13295-13332.

[31] SOWMYA R, PREMKUMAR M, JANGIR P. Newton-Raphson-based optimizer: A new population-based metaheuristic algorithm for continuous optimization problems[J]. Engineering Applications of Artificial Intelligence, 2024, 128: 107532.

[32] 陈绍杰,盛守前,韩 磊,等.基于机器学习的覆岩离层注浆充填地表移动变形动态预测:以霍尔辛赫煤业3801工作面为例[J].中国矿业大学学报, 2025, 54(6): 1261-1275.
CHEN Shaojie, SHENG Shouqian, HAN Lei, et al. Dynamic prediction of surface movement and deformation based on machine learning for overburden bed separation grouting filling: A case study of the 3801 working face in Huoerxinhe Coal Industry[J]. Journal of China University of Mining & Technology, 2025, 54(6): 1261-1275.

[33] 朱小景,潘一山,李 祁,等.巷道冲击地压软化区能量极值判别准则及试验研究[J].中国矿业大学学报,2021,50(5):975-982.
ZHU Xiaojing, PAN Yishan, LI Qi, et al. Softening zone energy extremum criterion and experimental study of roadway rock burst[J]. Journal of China University of Mining & Technology, 2021, 50(5): 975-982.

[34] 潘一山,代连朋.煤矿冲击地压发生理论公式[J].煤炭学报,2021,46(3):789-799.
PAN Yishan, DAI Lianpeng. Theoretical formula of rock burst in coal mines[J]. Journal of China Coal Society, 2021, 46(3): 789-799.

(责任编辑 沈春妮 宋 爽)