



基于MTD类分布小样本增强的带压开采煤层底板破坏深度预测

刘世伟 赵家鑫 袁乐忠 杨宇 陈茹齐 于昊

引用本文：

刘世伟, 赵家鑫, 袁乐忠, 等. 基于MTD类分布小样本增强的带压开采煤层底板破坏深度预测[J]. 煤炭科学技术, 2024, 52(S2): 175–185.

LIU Shiwei, ZHAO Jiaxin, YUAN Lezhong. Prediction of failure depth of coal seam floor in pressure mining based on small sample enhancement of MTD class distribution[J]. Coal Science and Technology, 2024, 52(S2): 175–185.

在线阅读 View online: <https://doi.org/10.12438/cst.2023-1360>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度神经网络的煤矿井下低光照图像增强算法

Deep neural network-based image enhancement algorithm for low-illumination images underground coal mines
煤炭科学技术. 2023, 51(9): 231–241 <https://doi.org/10.12438/cst.2022-1626>

基于BP神经网络的覆岩“两带”高度预测研究

Research on height prediction of “two zones” of overburden based on BP neural network in Wuyang Mine
煤炭科学技术. 2021, 49(8): 53–59 <http://www.mtkxjs.com.cn/article/id/25b4e31b-0c95-40e7-8cec-5a393c55ac7a>

基于循环神经网络的煤矿工作面瓦斯浓度预测模型研究

Research on prediction model of gas concentration based on RNN in coal mining face
煤炭科学技术. 2020, 48(1) <http://www.mtkxjs.com.cn/article/id/103948b5-8d79-4797-8ccb-4e67a686f21f>

承压水体上煤层底板下位隐伏断层采动突水机制研究

Study on mining water inrush mechanism of buried fault under coal seam floor above confined water body
煤炭科学技术. 2023, 51(2): 283–291 <https://doi.org/10.13199/j.cnki.est.2022-1698>

煤层底板采动破坏机制的岩性效应研究

Study on lithologic effect of mining-induced failure mechanism of coal seam floor
煤炭科学技术. 2023, 51(11): 166–178 <https://doi.org/10.13199/j.cnki.est.2023-0309>

奥灰承压水上采场底板沿工作面倾向破坏特征分析

Analysis of inclined failure characteristics of floor along working face in Ordovician limestone confined water stope
煤炭科学技术. 2022, 50(12): 206–214 <https://doi.org/10.13199/j.cnki.est.2020-1573>



关注微信公众号，获得更多资讯信息



移动扫码阅读

刘世伟, 赵家鑫, 袁乐忠, 等. 基于 MTD 类分布小样本增强的带压开采煤层底板破坏深度预测[J]. 煤炭科学技术, 2024, 52(S2): 175-185.

LIU Shiwei, ZHAO Jiaxin, YUAN Lezhong, et al. Prediction of failure depth of coal seam floor in pressure mining based on small sample enhancement of MTD class distribution[J]. Coal Science and Technology, 2024, 52 (S2): 175-185.

基于 MTD 类分布小样本增强的带压开采煤层底板 破坏深度预测

刘世伟^{1,2}, 赵家鑫¹, 袁乐忠¹, 杨宇¹, 陈茄齐¹, 于昊¹

(1. 河北工程大学 水利水电学院, 河北 邯郸 056038; 2. 水资源智慧调控与综合管理省部共建协同创新中心, 河北 邯郸 056038)

摘要: 机器学习预测方法在带压开采煤层底板破坏深度预测评价中得到广泛应用, 然而在预测模型构建过程中常常存在着实测数据获取成本高、采集难度大以及随机性强等问题, 基于少量样本构建的模型预测的性能往往受到预测精度和泛化能力的严重制约。通过文献调研, 收集获得 50 组实测数据样本, 引入 MTD (Mega-Trend-Diffusion) 类分布虚拟样本生成技术, 生成虚拟样本, 进一步实现煤层底板破坏深度实测样本的扩充增强; 采用 SaDE-ELM (Self Adaptive Differential Evolution-Extreme Learning Machine)、GA-PSO-BP (Genetic Algorithm-Particle Swarm Optimization-Back Propagation Neural Network)、BP (Back Propagation Neural Network) 等机器学习算法, 构建虚拟样本增强前后煤层底板破坏深度预测模型, 并对增强前后的模型预测精度进行对比分析。结果表明: 所提方法生成的虚拟样本与实测样本的分布基本一致; 采用虚拟样本增强后的预测模型精度均显著提高, 其中基于 MTD 类分布小样本增强的 PCA-SaDE-ELM (Particle Swarm Optimization-Self Adaptive Differential Evolution-Extreme Learning Machine) 预测模型的预测效果最优, 增强前后误差可降低 42.95%~51.27%。采用 MTD 类分布生成技术进行小样本增强, 并构建带压煤层底板破坏深度预测模型能够更加精确地预测多因素影响下煤层底板破坏深度, 通过与规范经验预测结果和滑移线场理论预测结果进行对比分析, 所提方法预测得到的云驾岭矿 19105 工作面底板破坏深度相对较大, 有利于工作面安全生产管理。相关研究成果对奥灰承压水上煤层的安全高效回采提供了有利支撑。

关键词: 带压开采; 底板破坏深度; 小样本; 数据增强; 神经网络

中图分类号: TD713 文献标志码: A 文章编号: 0253-2336(2024)S2-0175-11

Prediction of failure depth of coal seam floor in pressure mining based on small sample enhancement of MTD class distribution

LIU Shiwei^{1,2}, ZHAO Jiaxin¹, YUAN Lezhong¹, YANG Yu¹, CHEN Jiaqi¹, YU Hao¹

(1. College of Water Conservancy and Hydropower, Hebei University of Engineering, Handan 056038, China; 2. Collaborative Innovation Center for Smart Regulation and Comprehensive Management of Water Resources jointly built by the Provincial and Ministerial Departments, Handan 056038, China)

Abstract: The prediction method of machine learning has been widely used in the prediction and evaluation of the failure depth of coal seam floor in pressure mining. However, there are often some problems in the construction of the prediction model, such as high acquisition cost, difficulty in collecting and strong randomness of the measured data. The prediction performance of the model built based on a small number of samples is often severely restricted by the prediction accuracy and generalization ability. Through literature research, 50 sets of measured data samples were collected, and MTD similar distribution virtual sample generation technology was introduced to generate virtual samples to further expand and enhance the measured samples of coal seam floor failure depth. Machine learning algorithms such

收稿日期: 2023-10-26 责任编辑: 朱恩光 DOI: 10.12438/cst.2023-1360

基金项目: 河北省自然科学基金资助项目(E2020402087, E2021402023, E2024402061)

作者简介: 刘世伟(1989—), 男, 汉族, 河北邯郸人, 副教授, 硕士生导师, 博士。Tel: 15531081921, E-mail: liu1989shiwei@163.com

通讯作者: 赵家鑫(1999—), 男, 河北邯郸人, 硕士研究生。Tel: 19833753826, E-mail: jiaxiner99@163.com

as ADE-ELM, GA-PSO-BP and BP were used to build a prediction model of coal seam floor failure depth before and after virtual sample enhancement, and the prediction accuracy of the model before and after enhancement was compared and analyzed. The results show that the distribution of virtual samples generated by this method is basically consistent with that of measured samples. The accuracy of the prediction models enhanced with virtual samples is significantly improved, among which the PCA-ADE-ELM prediction model enhanced with small distributed samples of MTD class has the best prediction effect, and the error of the enhanced model can be reduced by 42.95% ~ 51.27%. MTD similar distribution virtual sample generation technology is used to enhance small samples, and the prediction model of failure depth of coal seam floor under pressure can be built to more accurately predict the failure depth of coal seam floor under the influence of multiple factors. Through comparison and analysis with the standard empirical prediction results and the slip line field theory prediction results, the failure depth of 19105 working face of Yunjialing Mine predicted by this method is relatively large, which is conducive to the safe production management of working face. The relevant research results provide favorable support for the safe and efficient mining of confined above-water coal seam of Ordovician limestone.

Key words: coal mining above confined water; failure depth of coal seam floor; small sample; data augmentation; neural network

0 引言

煤炭在我国一次能源结构中的占比仍超过 50%，是保障我国能源安全的压舱石^[1]，而随着开采深度的增加，煤层赋存地质环境愈加恶劣，加之复杂的开采扰动影响，导致底板奥灰水突涌风险加剧，严重威胁着深部煤层安全高效回采^[2]。据不完全统计，华北地区约 20% 的煤炭资源遭受底板岩溶水的影响^[3]，存在带压开采问题，以邯郸和邢台地区的煤矿尤甚，一旦管理不当常常诱发工作面突水灾害。为此，国内外学者开展诸多研究^[4-6]，并取得了可喜成果。其中突水系数理论得到了广泛认可和应用，底板有效隔水层厚度是突水系数理论的关键参数之一，合理确定底板破坏深度对煤层底板突水风险评价^[7-9]具有重要意义。

煤层底板破坏深度传统计算方法有：现场实测法^[10]，公式法^[11-12]，数值模拟法^[13-14]以及室内模型试验^[15]等。近年来，机器学习被广泛引入到煤层底板破坏深度预测研究中，于小鸽等^[16]提出的 BP 神经网络预测模型；施青龙等^[17]构建了 PCA-GWO-BP (Principal Component Analysis - Grey Wolf Optimization - Back Propagation Neural Network) 神经网络预测模型；邵良杉等^[18]构建了基于 PSO-ELM-Boosting (Particle Swarm Optimization- Extreme Learning Machine- Boosting) 的底板破坏深度预测模型；WANG 等^[19]基于模糊神经网络建立了底板破坏深度预测模型。相比传统方法，智能预测模型考虑了更多的影响因素，对原始数据进行了更深层次的挖掘；预测精度和效率都有很大提高。但上述成果中没有考虑底板奥灰水的影响。同时煤层底板破坏相关实测数据的获取还存在着成本高、难度大以及随机性强等问题。少量的样本难以覆盖整个有效空间，存在信息

不完整和不充足等缺陷。小样本问题严重影响煤层底板破坏深度预测模型的精度。通过撷取小样本数据间隙中存在的潜在信息，产生适当数量虚拟样本，即虚拟样本生成技术 (Virtual Sample Generation, VSG)，实现小样本数据增强，进而能够提高训练样本对总体数据特征的表征能力和模型的学习与泛化能力。

基于整体趋势扩散的虚拟样本生成技术 (Mega-Trend-Diffusion, MTD)^[20-21] 是一种常用的 VSG 技术，其属于随机样本生成技术的一种改进形式，可实现对真实数据信息间隔填补，主要通过隶属函数来估计样本变量的真实数据空间，即通过隶属函数计算其相应的虚拟样本信息的左边界 LB 和右边界 RB，从而在该范围内生成虚拟样本信息。但是该方法较少考虑真实样本与虚拟样本之间变量的相似特征，从而导致预测模型的过度拟合现象。此外，实际中先验知识的获取存在一定的困难性，实测样本中的变量之间往往并非完全独立，导致虚拟样本的生成难以满足变量间的独立性假设。且由于虚拟样本的引入存在误差的累积效应，虚拟样本的生成数量也是影响模型预测精度的因素之一。综上，少量实测样本、变量独立性以及虚拟样本生成数量均是影响煤层底板破坏深度预测模型精度的关键因素，如何消除原始样本数据特征量之间的关联性、生成与实测样本信息分布类似的虚拟样本以及确定合理的虚拟样本数量等问题值得进一步深入研究。

鉴于此，基于文献^[22]调研获得带压开采煤层底板破坏深度实测数据和影响因子集，共计 50 组数据；借助 PCA 算法优化实测数据样本空间，消除原始数据样本中各变量间的相关性；分析 PCA (Principal Component Analysis) 优化后的数据样本中主成分分布特征，通过引入基于 MTD 类分布生成技术，生成类似分布的虚拟样本，并通过 K-S (Kolmogorov-

Smirnov) 检验验证虚拟样本与实测样本分布的相似性, 进而实现对实测数据小样本的扩充; 采用 SaDE-ELM 混合优化算法, 生成虚拟样本输出, 构建影响煤层底板破坏深度的混合数据样本; 基于 SaDE-ELM、GA-PSO-BP、BP 3 种算法, 构建底板破坏深度预测模型, 对比分析数据增强前后的模型预测精度。相关研究成果对煤层底板突水风险评价和预测预警能力的提升具有一定的支撑作用。

1 基于 MTD 小样本增强的底板破坏深度预测模型

1.1 实测样本数据的 PCA 优化

采用已搜集到的 50 组煤层底板破坏深度相关数据, 见文献^[22]。实测数据集中包含底板破坏深度影响因子和评价指标集 2 部分。其中影响因子集包含工程地质因素 (导水构造发育程度 f_1 、陷落柱发育程度 f_2 、断层发育程度 f_3 和断层落差 H_1)、水文地质因素 (底板含水层水压力 P 、初始底板隔水层厚度 h 、隔水层砂性百分比 p_1 、隔水层泥性百分比 p_2 、隔水层灰性百分比 p_3 、断裂导水特性 k_1 、构造充水性 k_2 和最大涌水量 Q) 和工作面开采条件 (开采深度 H 、煤层倾角 α 、煤层厚度 M 、工作面长度 L_1 、走向长度 L_2 、工作面采高 m 、月推进步距 L_3) 等 19 种影响因子。值得说明的是, 此处的初始底板隔水层厚度是指地质勘探的结果, 不包含开采扰动的影响。

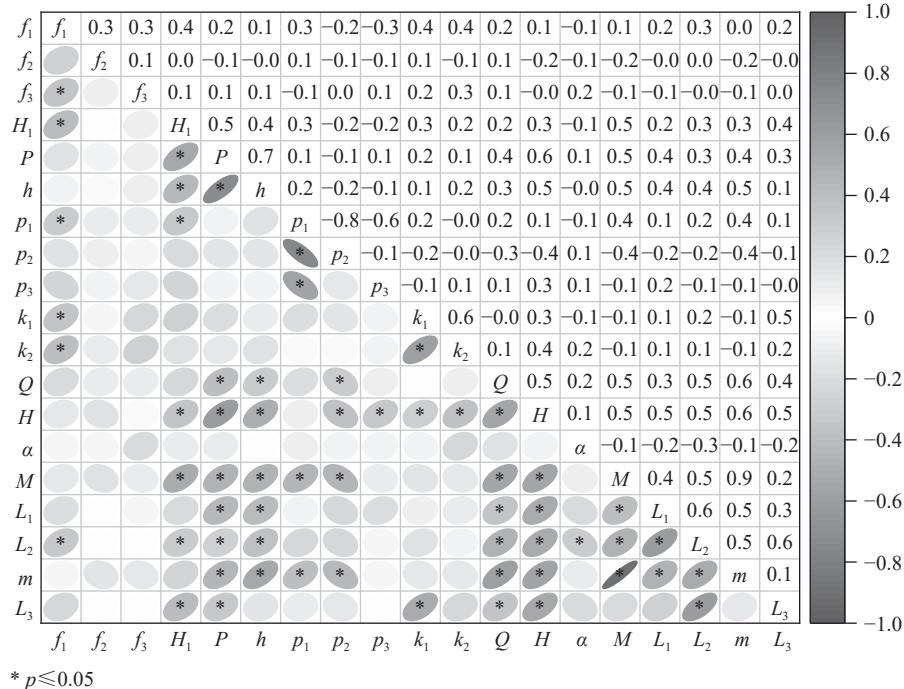


图 1 影响因子相关性热图
Fig.1 Influence factor correlation heat map

对实测数据集进行皮尔逊相关性检验, 如图 1 所示。由图分析可知标有星号(*)的某些变量间的确存在相关性, 如底板含水层水压力 P 和初始底板隔水层厚度 h , 工作面采高 m 和开采深度 H 、最大涌水量 Q 等。考虑到 MTD 类分布生成技术的独立性假设要求, 采用 PCA 算法对 50 组原始数据样本进行主成分分析, 降低样本数据之间的关联性。借助 SPSS 软件对影响因子集进行 PCA 降维优化处理, 确定协方差矩阵 R 的特征值及对应的特征向量, 结果见表 1、表 2。根据主成分累计贡献率选取前 9 个主成分作为输入变量, 并根据式(1)计算各主成分。

$$A_{mk} = [a'_{ij}]_{m \times n} \frac{W_{nk}}{\sqrt{\lambda_j}} \quad (1)$$

式中: a'_{ij} 为标准化后的数据, $i=1, 2, \dots, m; j=1, 2, \dots, n$; m 为样本个数; n 为样本特征数; A_{mk} 为样本的主成分; W_{nk} 为特征向量根据特征值大小按列降序排列, 选取前 k 列组成的矩阵; λ_j 为特征向量, $j=1, 2, \dots, n$ 。

1.2 数据集划分

SPXY (Sample set partitioning based on joint x-y distance) 算法^[23] 是广泛应用的数据集划分方法, 是基于 KS (Kennard-Stone) 算法提出的一种改进方法, 该方法能够同时计算不同样本的 x 向量方向和 y 向量方向的欧氏距离, 并通过正则化将 x 和 y 方向的距离结合, 更加全面的评估和划分数据集, 能够有效地减小过拟合问题, 提高机器学习模型的泛化能力。

表1 总方差解释
Table 1 Total variance explanation

主成分	特征值	方差百分比/%	累积/%
I	5.541	29.163	29.163
II	2.464	12.971	42.134
III	2.121	11.163	53.297
IV	1.534	8.072	61.370
V	1.322	6.957	68.327
VI	1.215	6.396	74.723
VII	0.933	4.912	79.635
VIII	0.823	4.331	83.966
IX	0.743	3.910	87.876
X	0.588	3.097	90.973
X I	0.453	2.382	93.355
X II	0.310	1.633	94.989
X III	0.286	1.508	96.497
X IV	0.267	1.404	97.901
X V	0.176	0.925	98.827
X VI	0.122	0.641	99.468
X VII	0.079	0.415	99.882
X VIII	0.022	0.118	100.000
X IX	0.000	0.000	100.000

笔者采用该方法进行样本的训练集和测试集划分。

1.3 基于 MTD 类分布的虚拟样本生成

1) 基于 MTD 类分布生成技术的基本原理

MTD 的基本原理^[20-21]如图 2 所示, 所生成的虚拟样本的左边界 LB 和右边界 RB 按式(2)和式(3)计算。

$$LB = \begin{cases} CL - Skew_L \times \sqrt{-2 \times \frac{\hat{S}_x^2}{N_L} \times \ln(10^{-20})} & LB \leq \min \\ \min & LB > \min \end{cases} \quad (2)$$

$$RB = \begin{cases} CL + Skew_R \times \sqrt{-2 \times \frac{\hat{S}_x^2}{N_R} \times \ln(10^{-20})} & RB \geq \min \\ \min & RB < \min \end{cases} \quad (3)$$

其中:

$$CL = (\max - \min) / 2$$

$$Skew_L = \frac{N_L}{N_L + N_R}$$

$$Skew_R = \frac{N_R}{N_L + N_R}$$

$$\hat{S}_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

表2 成分矩阵
Table 2 Component matrix

主成分	I	II	III	IV	V	VI	VII	VIII	IX
f_1	0.377	0.526	-0.410	0.078	0.086	0.324	0.131	-0.246	0.072
f_2	-0.039	0.159	-0.346	-0.009	0.350	0.613	-0.325	0.429	0.123
f_3	0.035	0.521	0.098	0.386	0.106	0.366	0.052	-0.405	-0.354
H_1	0.620	0.202	-0.237	0.097	-0.408	0.043	-0.109	0.076	-0.347
P	0.695	0.070	0.300	0.220	-0.348	0.091	-0.271	0.193	-0.061
h	0.661	-0.042	0.187	0.221	-0.387	0.125	-0.342	-0.014	0.194
p_1	0.453	-0.162	-0.768	0.204	0.159	-0.265	-0.071	-0.008	0.052
p_2	-0.515	0.167	0.323	-0.147	-0.575	0.327	0.308	0.068	0.133
p_3	-0.036	0.034	0.751	-0.124	0.482	-0.011	-0.279	-0.073	-0.245
k_1	0.280	0.755	-0.161	-0.148	-0.003	-0.325	-0.147	0.064	0.030
k_2	0.219	0.741	0.150	0.148	0.046	-0.301	0.052	-0.145	0.314
Q	0.662	-0.087	0.151	0.182	0.318	0.233	0.387	0.213	-0.021
H	0.789	0.098	0.408	-0.027	0.140	-0.224	-0.005	0.125	0.011
α	-0.116	0.151	0.283	0.745	0.125	-0.091	0.260	0.276	0.185
M	0.782	-0.464	-0.056	0.157	-0.064	-0.030	0.125	-0.098	-0.143
L_1	0.629	-0.064	0.245	-0.312	0.092	0.194	-0.163	-0.234	0.346
L_2	0.717	0.017	-0.031	-0.452	0.062	0.193	0.266	-0.085	0.120
m	0.765	-0.515	0.038	0.109	0.026	0.002	0.107	-0.138	0.068
L_3	0.559	0.380	0.003	-0.473	-0.021	-0.118	0.248	0.330	-0.247

式中: \min 为样本特征最小边界; \max 为样本特征最大边界; LB 为虚拟样本左边界; CL 为样本数据中心点; RB 为虚拟样本右边界; N_L, N_R 分别为中心点数据 CL 左边和右边的样本数量; \hat{s}_x^2 为样本方差; x_i 为样本, $i=1, 2, \dots, n$; \bar{x} 为样本均值; n 为样本数量。

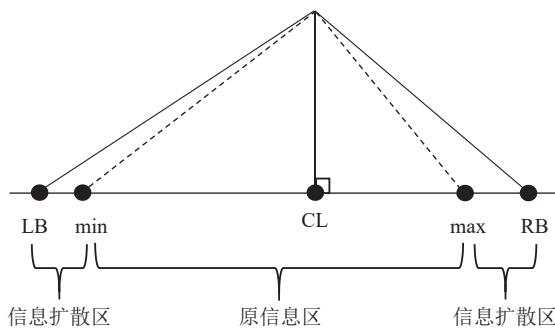


图 2 MTD 原理
Fig.2 MTD schematic diagram

一般情况下, 虚拟样本的生成方法采用插值法, 根据插值方法的不同可分为三角分布插值法, 均匀分布插值法、正态分布插值法以及混合插值方法。采用上述方法生成的虚拟样本中的变量分布类型为既定类型, 没有考虑到真实样本数据的潜在分布规律等情况的影响, 虚拟样本与真实样本的分布不同, 训练出来的模型可能会过度拟合虚拟样本的特征, 无法准确预测真实样本。因此基于 MTD 提出类分布虚拟样本生成技术, 其核心是通过分析实测样本数据潜在分布类型, 进而生成与实测样本相似分布的虚拟样本集, 可按照式(4)计算生成。

$$X_{VSG} = LB + g(RB - LB) + \Delta x \quad (4)$$

式中: g 为与实测样本分布相似的随机数, $g = (0, 1)$; Δx 为修正量, 由于真实样本存在边界分布不平衡的问题, 可能出现少量极端样本, 这会导致 MTD 生成的拓展域上的虚拟样本不能很好的反应实测样本的分布, 生成的虚拟样本分布情况与实测样本存在一定偏移, 故对虚拟样本增加一个修正因子 Δx 。

2) 实测样本训练集的潜在分布特征

实测样本训练集中主成分 I ~ IX 的频率直方图和累计分布曲线, 如图 3 所示。由图中分析可知, 除主成分 I 出现明显两端分布大、中间分布少的特点外, 其他主成分分布两头低、中间高, 与正态分布相似。故主成分 I 采用双正态分布组合的形式进行模拟, 其他主成分采用正态分布进行模拟。

3) 虚拟样本合理性 K-S 检验

为进一步验证生成的虚拟样本与实测样本分布

的相似性, 采用 K-S 检验方法^[24] 对虚拟样本与实测样本训练集进行分布检验, 该方法适用于样本量小的非参数检验。

设原假设: 两个样本来自同一连续分布; 备择假设: 两个样本来自不同的连续分布。虚拟样本生成数量分别为 50、200 和 500, 限于图幅要求, 文中仅给出虚拟样本数量为 200 时, 虚拟样本与实测样本中主成分的累积分布曲线对比结果, 如图 4 所示。由图中分析可知, 九个主成分的分布显著性水平均在 0.8 以上, 故接受虚拟样本与实测样本同分布假设。

3) 虚拟样本输出生成

通过 SPXY 算法划分生成的训练集为 (x, y) , 采用 MTD 类分布技术获得得到的虚拟样本为 $(T_x, y_{(T_x)})$ 。为了给虚拟样本输入生成更合理的虚拟样本输出, 首先采用 SaDE-ELM 算法对训练集 (x, y) 进行回归训练, 通过调整模型超参数, 使预测均方根误差 RMSE 降低到 0.1 以下, 利用训练好的模型为虚拟样本生成输出值。

1.4 底板破坏深度预测模型构建

将虚拟样本集与训练集组成的混合训练集, 分别采用 SaDE-ELM、BP、GA-PSO-BP 3 种智能算法构建数据增强前后的带压开采煤层底板破坏深度预测模型, 并对模型预测精度进行对比分析, 模型实现的具体流程如图 5 所示。

2 模型验证

为验证引入虚拟样本对模型的优化效果, 采用绝对误差 E_a 、相对误差 δ 、均方根误差 RMSE 三个指标评价模型预测精度, 评价指标的计算方法见式(5)~式(7)。

$$E_a = |\hat{b}_i - b_i| \quad (5)$$

$$\delta = \frac{|\hat{b}_i - b_i|}{b_i} \times 100\% \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_i^n (\hat{b}_i - b_i)^2}{n-1}} \quad (7)$$

式中: n 为测试集样本个数; E_a 为绝对误差; δ 为相对误差; RMSE 为均方根误差; b_i 为真实输出; \hat{b}_i 为预测输出。

虚拟样本的数量对模型计算效率和预测精度具有一定程度的影响, 不同虚拟样本数量下模型的误差分布变化规律, 如图 6 所示。由图中分析可知, 虚

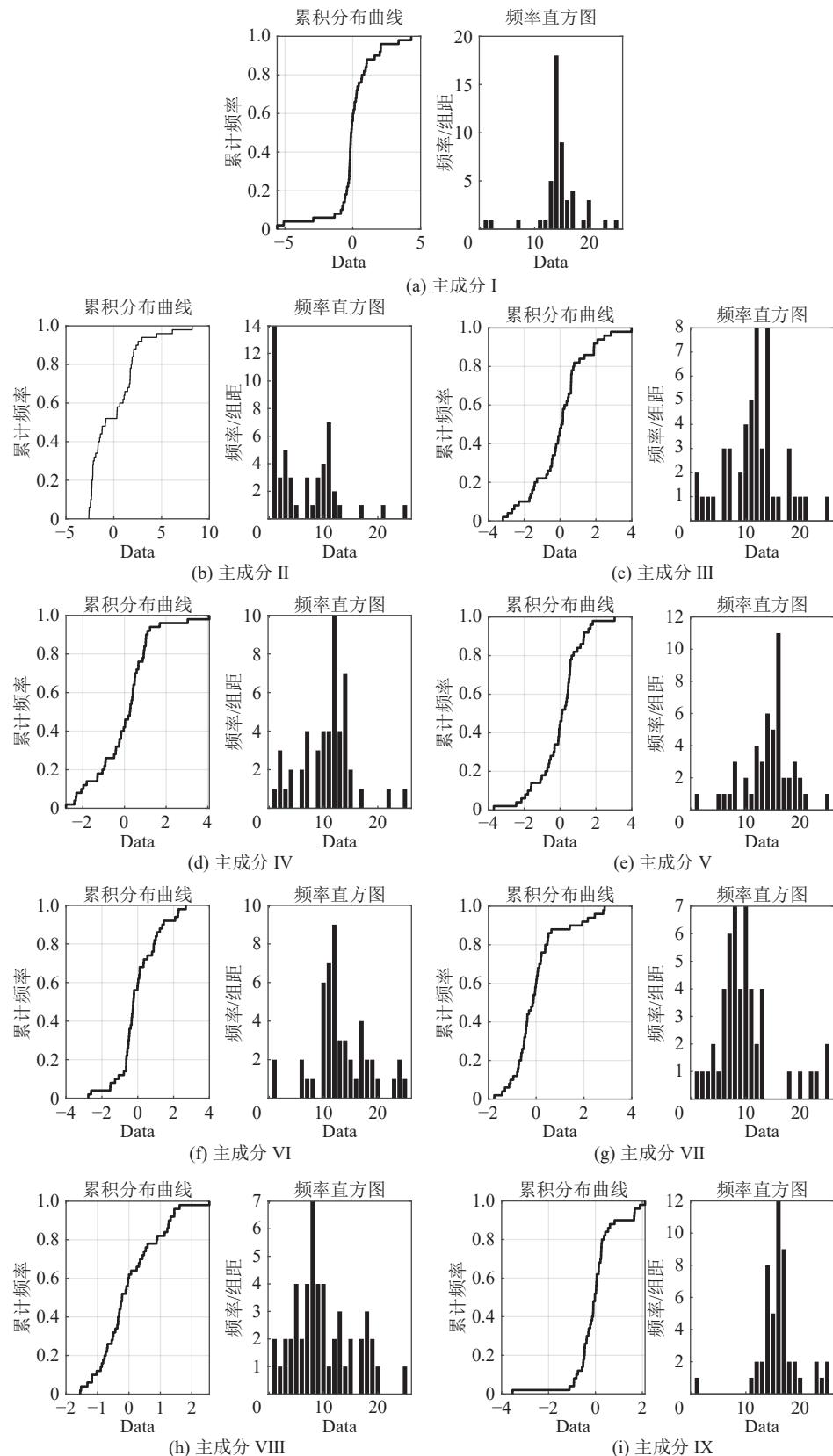


图3 主成分的频率直方图和累积分布曲线

Fig.3 Frequency histogram and cumulative distribution curve of principal components

拟样本数量超过80后,模型预测误差变化基本稳定,而未考虑虚拟样本增强的模型预测误差始终波动变

化,综合考虑计算效率及模型预测误差的影响,本次分析中虚拟样本数量均取100。

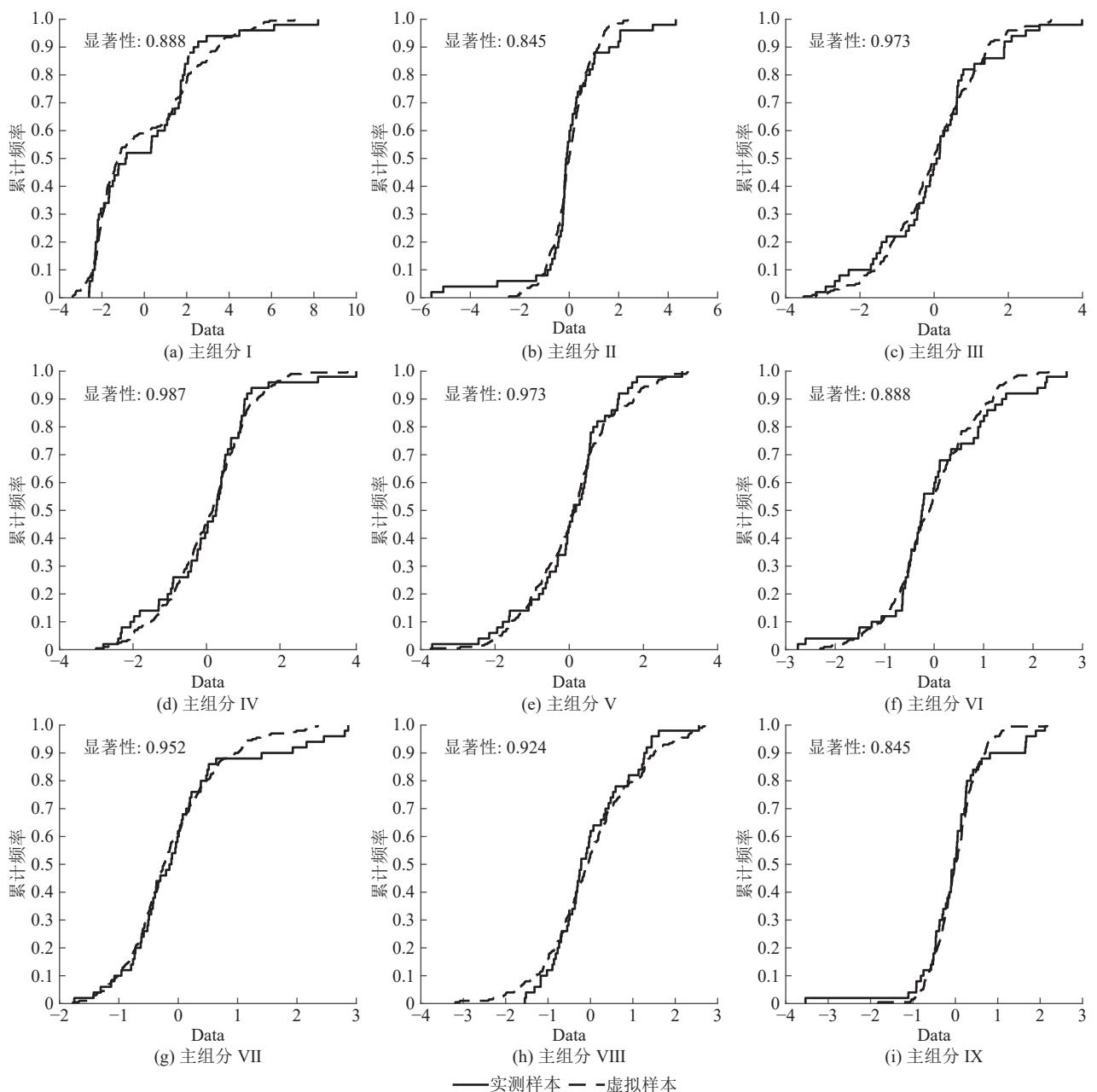


图 4 虚拟样本数为 200 时实测样本和虚拟样本中的主成分累积分布曲线对比分析

Fig.4 Comparative analysis of principal component cumulative distribution curves in measured samples and virtual samples with a virtual sample size of 200

根据 SPXY 算法样本集划分结果, 对黑山矿 3 号 (真实破坏深度: 9.34 m)、夏庄矿 3 号 (7.83 m)、夏庄矿 5 号 (12.58 m)、夏庄矿 7 号 (9.34 m)、夏庄矿 8 号 (13.66 m)、夏庄矿 10 号 (13.66 m)、夏庄矿 11 号 (13.66 m)、夏庄矿 12 号 (11.50 m)、夏庄矿 13 号 (6.10 m)、双山大井 4 号 (8.80 m) 10 个样本进行预测。为避免模型的随机性, 进行 5 次试验, 结果取平均。采用虚拟样本增强前后, 各模型的误差对比分析结果如图 7—图 9 和表 3 所示。由图表分析可知, 数据增强前, SaDE-ELM 模型预测的底板破坏深度

分别为 9.335 7、7.868 1、12.594 0、9.254 9、13.625 5、13.584 6、13.679 3、11.550 9、6.120 8、8.804 6 m, GA-PSO-BP 模型预测的底板破坏深度分别为 9.558 8、7.779 0、12.714 2、9.262 4、13.526 2、13.640 0、13.490 1、11.360 7、5.988 0、8.788 1 m, BP 模型预测的底板破坏深度分别为 9.466 7、6.512 2、10.806 1、8.358 2、12.721 9、11.721 9、11.635 3、9.219 3、6.149 6、8.083 0 m; 数据增强后, SaDE-ELM 模型预测的底板破坏深度分别为 9.343 5、7.840 1、12.564 8、9.310 6、13.648 6、13.613 5、13.705 9、11.522 8、6.111 4、8.801 5 m, GA-PSO-BP

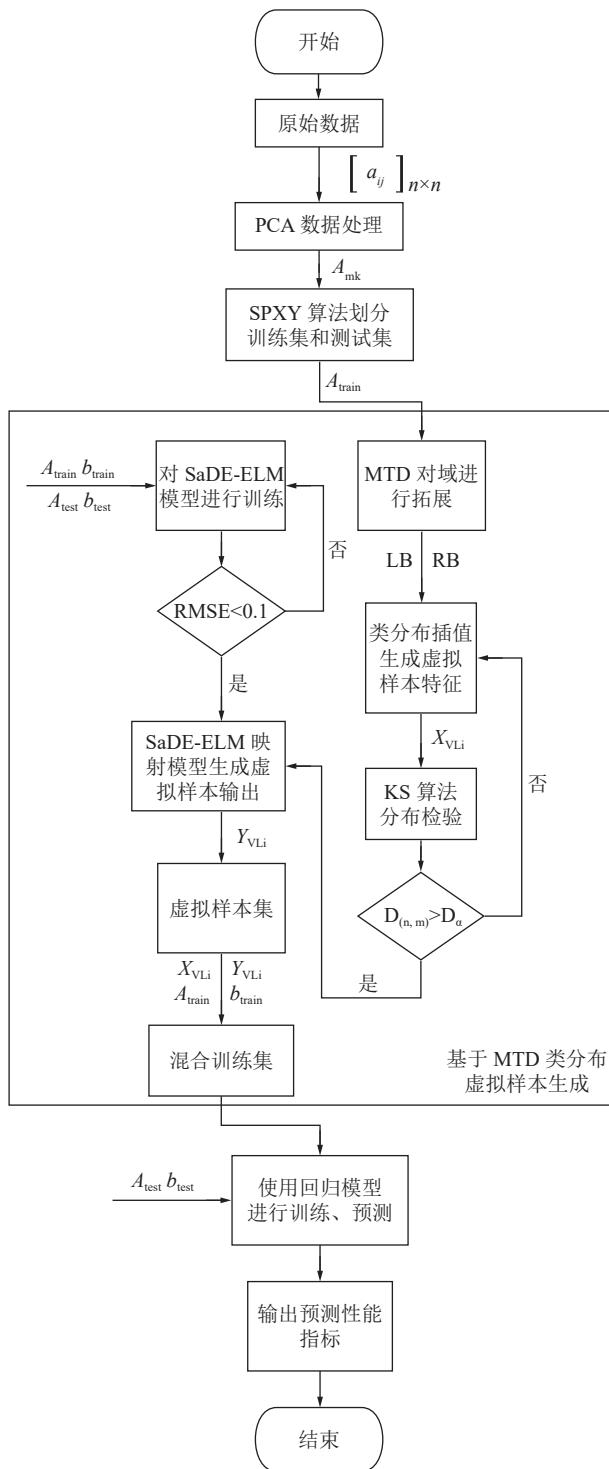


图 5 模型实现流程

Fig.5 Model implementation process

模型预测的底板破坏深度分别为 9.2529、7.7339、12.6455、9.2745、13.7134、13.5855、13.5567、11.4919、6.3055、8.7799 m, BP 模型预测的底板破坏深度分别为 10.1114、7.9542、12.8855、9.6846、12.3618、13.7132、13.6569、11.4998、6.7525、8.8053 m。相较于虚拟样本增强前, 增强后 SaDE-ELM 模型、GA-PSO-BP 模型、BP 模型的平均绝对误差分别降低约

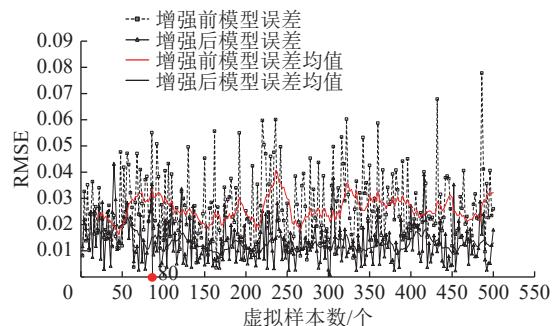


图 6 模型预测误差随虚拟样本数量的变化规律

Fig.6 Variation of model prediction error with number of virtual samples

42.95%、27.09%、70.7%，平均相对误差分别降低约 47.08%、16.3%、65.9%，均方根误差分别降低约 51.27%、27.77%、36.46%，可见采用虚拟样本增强后的预测模型精度显著提升，且 SaDE-ELM 模型预测精度最优。

3 工程应用

以邯郸云驾岭煤矿九号煤层为研究对象。根据云驾岭矿 9 煤层工程地质调查结果，目前正在生产的工作面为 19105 工作面。相邻的 19103 工作面已被完全开采。19105 工作面存在多处局部断层，最小落差为 1 m。煤底板以砂岩为主，抗拉强度 5 MPa，单轴抗压强度 33.26 MPa。煤层下方存在奥陶系灰岩含水层，工作面距该含水层顶界面 31.41 m。含水层底部界面水压为 1.49 MPa。工作面采动期间，平均涌水量约为 5 m³/h，最大涌水量约为 10 m³/h。根据 MENG 等^[25], p_1 、 p_2 、 p_3 的值由式 (8) 计算得出。根据地质勘探结果和 MENG^[25] 的方法计算，可以确定 19105 工作面影响因子值，见表 4。

$$p_i = \frac{h_i}{h_1} \times 100\% \quad (8)$$

式中： p_i 为岩层厚度占不透水层总厚度的比例； h_i 为层状岩层厚度。

采用规范经验方法 (CWIC)^[25]、基于滑移场理论的解析解 (ASSF)^[26] 和本方法对 19105 工作面底板突水风险进行预测和评估，如图 10 所示。可以看出，对于云驾岭煤矿 19105 工作面实际工况，本方法预测结果为 13.4794 m, CWIC 预测结果为 8.09 m, ASSF 预测结果为 12.85 m。所提方法的预测结果略大于 CWIC 和 ASSF，主要是综合考虑了地质和建设因素的影响。该方法的评价结果偏向危险性，有利于工程安全。

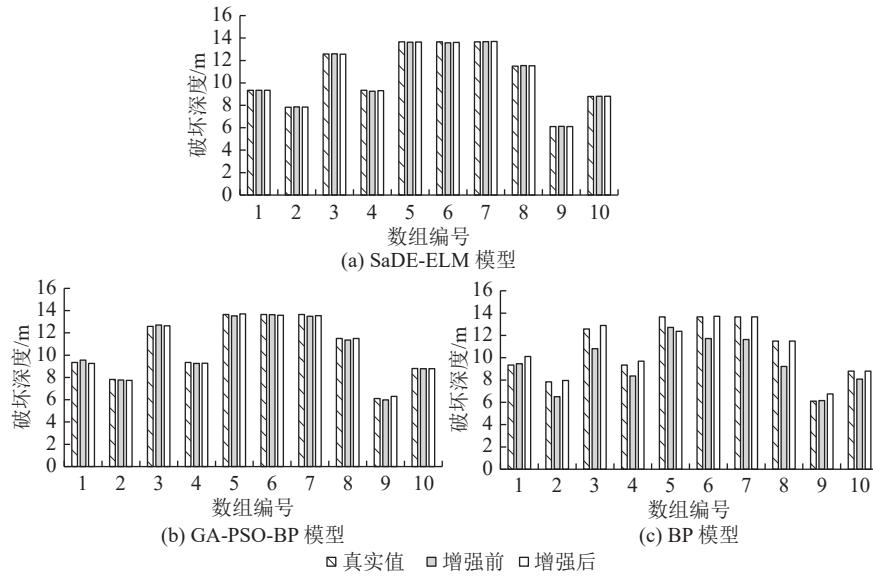


图 7 模型预测结果

Fig.7 Model prediction results

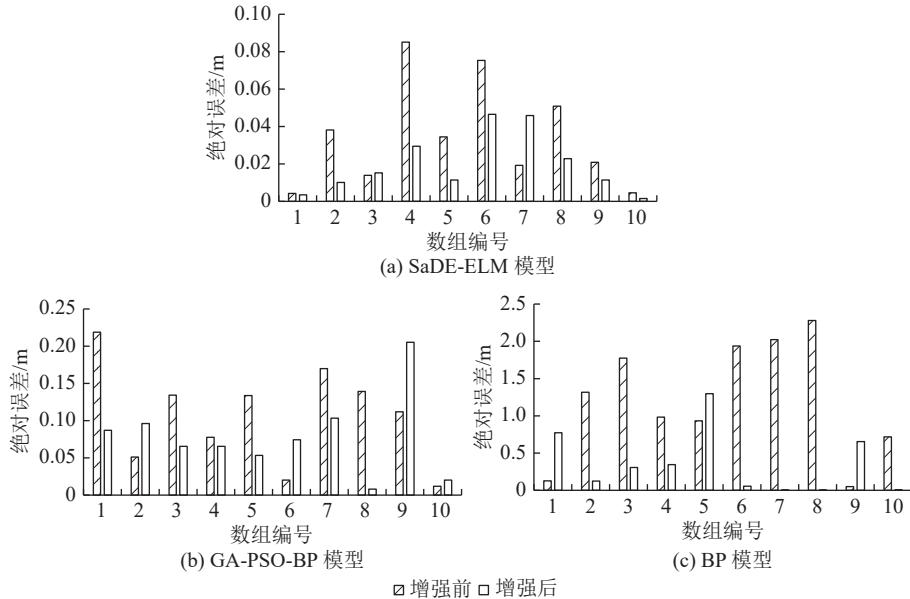


图 8 模型预测绝对误差

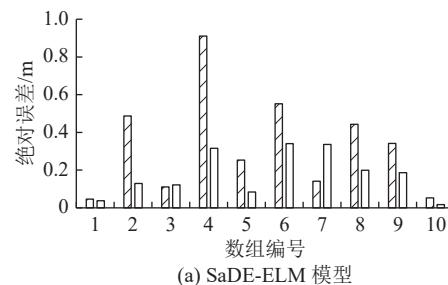
Fig.8 Absolute error of model prediction

4 结 论

1) 针对底板破坏深度实测样本量少、预测模型精度低和泛化能力弱等问题, 通过采用 MTD 类分布

虚拟样本生成技术, 增强煤层底板破坏深度实测样本, 通过 K-S 检验方法检验了虚拟样本数据集的合理性。

2) 采用 SaDE-ELM、GA-PSO-BP、BP 3 种算法,



(a) SaDE-ELM 模型

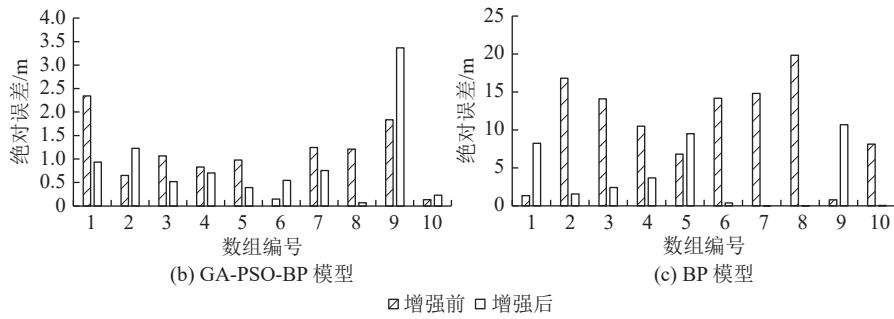


图9 模型预测相对误差
Fig.9 Relative error of model prediction

表3 模型预测误差

Table 3 Model predictive error

预测性能指标	数据样本	SaDE-ELM		GA-PSO-BP		BP	
		误差	误差降低/%	误差	误差降低/%	误差	误差降低/%
平均绝对误差/m	实测训练集	0.034 678	42.953 28	0.106 853	27.092 14	1.214 402	70.699 71
	增强训练集	0.019 783		0.077 904		0.355 823	
平均相对误差/%	实测训练集	0.333 537	47.082 35	1.044 385	16.301 62	10.743 78	65.898 09
	增强训练集	0.176 5		0.874 134		3.663 832	
均方根误差	实测训练集	0.057 211	51.273 44	0.243 7	27.769 8	1.539 483	36.455 41
	增强训练集	0.027 877		0.176 025		0.978 258	

表4 19105工作面影响因子值
Table 4 Impact factor values of 19105 working face

影响因子	值	影响因子	值	影响因子	值
f_1	0.5	P / MPa	1.49	H / m	346
f_2	0	h_i / m	34.41	α / °	13
f_3	1	p_1	1	M / m	2.88
H_1 / m	1	p_2	0	L_1 / m	69
Q / (m ³ · h ⁻¹)	10	p_3	0	L_2 / m	199
		k_1	0.5	m_1 / m	2.88
		k_2	0.5	L_3 / m	30

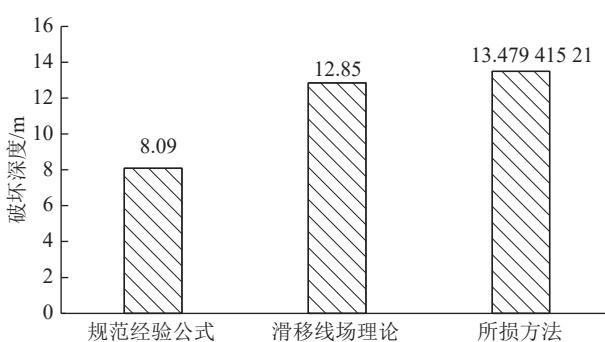


图10 19105工作面破坏深度预测结果对比
Fig.10 Comparison of failure depth prediction results of 19105 working face

构建了虚拟样本增强前后煤层底板破坏深度预测模型。

3)采用虚拟样本增强后的煤层底板破坏深度

预测模型精度显著提高,模型预测误差可降低42.95%~51.27%,其中基于MTD类分布小样本增强的PCA-SaDE-ELM预测模型的预测效果最优。

4)相较于其他方法,所提方法预测获得的云驾岭19105工作面底板破坏深度相对较大,有利于工作面安全生产管理。

参考文献(References):

- [1] 张吉雄, 张强, 周楠, 等. 煤基固废充填开采技术研究进展与展望[J]. 煤炭学报, 2022, 47(12): 4167-4181.
ZHANG Jixiong, ZHANG Qiang, ZHOU Nan, et al. Research progress and prospect of coal based solid waste backfilling mining technology[J]. Journal of China Coal Society, 2022, 47(12): 4167-4181.

- [2] SUN W J, ZHOU W F, JIAO J. Hydrogeological classification and water inrush accidents in China's coal mines[J]. *Mine Water and the Environment*, 2016, 35(2): 214–220.
- [3] WANG P P, JIANG Y D, REN Q S. Roof hydraulic fracturing for preventing floor water inrush under multi aquifers and mining disturbance: A case study[J]. *Energies*, 2022, 15(3): 1187.
- [4] DUAN H F, ZHAO L J. New evaluation and prediction method to determine the risk of water inrush from mining coal seam floor[J]. *Environmental Earth Sciences*, 2021, 80(1): 30.
- [5] 靳德武. 我国煤层底板突水问题的研究现状及展望[J]. *煤炭科学技术*, 2002, 30(6): 1–4.
JIN Dewu. Research status and outlook of water outburst from seam floor in China coal mines[J]. *Coal Science and Technology*, 2002, 30(6): 1–4.
- [6] 袁亮. 我国煤炭资源高效回收及节能战略研究[J]. *中国矿业大学学报(社会科学版)*, 2018, 20(1): 3–12.
YUAN Liang. Strategies of high efficiency recovery and energy saving for coal resources in China[J]. *Journal of China University of Mining & Technology (Social Sciences)*, 2018, 20(1): 3–12.
- [7] LIU Y, LIU S L, HUO Z C, et al. Failure characteristics of coal seam floor and risk assessment of water inrush caused by underground coal mining[J]. *Energy Exploration and Exploitation*, 2023, 41(2): 677–695.
- [8] LIU Y, ZHU J Z, LIU Q M, et al. Mechanism analysis of delayed water inrush from Karst collapse column during roadway excavation based on seepage transition theory: A case study in PanEr coal mine[J]. *Energies*, 2022, 15(14): 4987.
- [9] HE T, LI G D, SUN C, et al. Floor failure characteristics of thick coal seam mining above confined aquifer[J]. *Mining, Metallurgy & Exploration*, 2022, 39(4): 1553–1562.
- [10] ZHANG P S, OU Y C, SUN B Y, et al. A case study of floor failure characteristics under fully mechanised caving mining conditions in extra-thick coal seams[J]. *Journal of Geophysics and Engineering*, 2020, 17(5): 813–826.
- [11] 李杨杨, 张士川, 孙熙震, 等. 煤层采动底板突水演变过程可视化试验平台研制与试验研究[J]. *煤炭学报*, 2021, 46(11): 3515–3524.
LI Yangyang, ZHANG Shichuan, SUN Xizhen, et al. Development and experimental study on visualization test platform for water inrush evolution process of coal seam mining floor[J]. *Journal of China Coal Society*, 2021, 46(11): 3515–3524.
- [12] WANG W M, YUAN Y, LIANG X K, et al. Experimental study on floor damage and slurry material ratio optimization in deep and high confined water mining[J]. *Processes*, 2022, 10(9): 1806.
- [13] QI Y, WANG W, GE J Q, et al. Development characteristics of the rock fracture field in strata overlying a mined coal seam group[J]. *PLoS One*, 2022, 17(10): e0268955.
- [14] LIU W T, DU Y H, LIU Y B, et al. Failure characteristics of floor mining-induced damage under deep different dip angles of coal seam[J]. *Geotechnical and Geological Engineering*, 2019, 37(2): 985–994.
- [15] 郭惟嘉, 张士川, 孙文斌, 等. 深部开采底板突水灾变模式及试验应用[J]. *煤炭学报*, 2018, 43(1): 219–227.
GUO Weijia, ZHANG Shichuan, SUN Wenbin, et al. Experimental and analysis research on water inrush catastrophe mode from coal seam floor in deep mining[J]. *Journal of China Coal Society*, 2018, 43(1): 219–227.
- [16] 于小鸽, 韩进, 施龙青, 等. 基于 BP 神经网络的底板破坏深度预测[J]. *煤炭学报*, 2009, 34(6): 731–736.
YU Xiaoge, HAN Jin, SHI Longqing, et al. Forecast of destroyed floor depth based on BP neural networks[J]. *Journal of China Coal Society*, 2009, 34(6): 731–736.
- [17] 施龙青, 张荣邀, 韩进, 等. 基于 GWO 改进的 PCA-BP 神经网络煤层底板破坏深度预测模型[J]. *矿业研究与开发*, 2020, 40(2): 88–93.
SHI Longqing, ZHANG Rongqiao, HAN Jin, et al. Prediction model of failure depth of coal seam floor based on PCA-BP neural network improved by GWO[J]. *Mining Research and Development*, 2020, 40(2): 88–93.
- [18] 邵良杉, 周玉. 基于 PSO-ELM-Boosting 模型的底板破坏深度预测[J]. *中国安全科学学报*, 2018, 28(4): 24–29.
SHAO Liangshan, ZHOU Yu. Prediction of destroyed floor depth based on SO-ELM-Boosting model[J]. *China Safety Science Journal*, 2018, 28(4): 24–29.
- [19] WANG Z C, ZHAO W T, HU X. Analysis of prediction model of failure depth of mine floor based on fuzzy neural network[J]. *Geotechnical and Geological Engineering*, 2019, 37(1): 71–76.
- [20] LI D C, CHEN C C, CHANG C J, et al. A tree-based-trend-diffusion prediction procedure for small sample sets in the early stages of manufacturing systems[J]. *Expert Systems with Applications*, 2012, 39(1): 1575–1581.
- [21] CHAO G Y, TSAI T I, LU T J, et al. A new approach to prediction of radiotherapy of bladder cancer cells in small dataset analysis[J]. *Expert Systems with Applications*, 2011, 38(7): 7963–7969.
- [22] 余国锋. 基于微震和神经网络的煤层底板突水预警技术研究[D]. 淮南: 安徽理工大学, 2022.
YU Guofeng. Study on early warning technology of water inrush in coal seam floor based on microseism and neural network[D]. Huainan: Anhui University of Science & Technology, 2022.
- [23] ZHANG Y X, WANG X Z, WANG Y F, et al. Detection of tomato water stress based on terahertz spectroscopy[J]. *Frontiers in Plant Science*, 2023, 14: 1095434.
- [24] 宫凤强, 王天成, 黄天朗. 基于正态信息扩散原理的极值型工程参数概率分布推断方法[J]. *中南大学学报(自然科学版)*, 2020, 51(6): 1692–1702.
GONG Fengqiang, WANG Tiancheng, HUANG Tianlang. An inference method for probability distribution of extreme value engineering parameters based on normal information diffusion principle[J]. *Journal of Central South University (Science and Technology)*, 2020, 51(6): 1692–1702.
- [25] MENG Z P, LI G Q, XIE X T. A geological assessment method of floor water inrush risk and its application[J]. *Engineering Geology*, 2012, 143: 51–60.
- [26] ZHANG W, ZHANG D S, QI D H, et al. Floor failure depth of upper coal seam during close coal seams mining and its novel detection method[J]. *Energy Exploration & Exploitation*, 2018, 36(5): 1265–1278.