

柴敬,王润沛,杜文刚,等. 基于XGBoost的光纤监测矿压时序预测研究[J]. 采矿与岩层控制工程学报, 2020, 2(4): 043035.

CHAI Jing, WANG Runpei, DU Wengang, et al. Study on time series prediction of rock pressure by XGBoost in optical fiber monitoring[J]. Journal of Mining and Strata Control Engineering, 2020, 2(4): 043035.

基于XGBoost的光纤监测矿压时序预测研究

柴敬^{1,2}, 王润沛¹, 杜文刚¹, 雷武林¹, 朱旭宝¹

(1. 西安科技大学 能源学院, 陕西 西安 710054; 2. 教育部西部矿井开采及灾害防治重点实验室, 陕西 西安 710054)

摘要:为掌握矿山压力显现规律,针对岩石内部变形预测难的问题,采用XGBoost算法和分布式光纤技术相结合的方法,基于相似材料模型试验,制作三维立体模型,并在模型内部预埋3根垂直光纤,模拟工作面开采。引入光纤平均频移变化度作为判断周期来压指标,构建矿压监测数据的相空间,采用XGBoost方法建立矿山压力显现规律时间序列预测模型。同时,在训练样本和测试样本不变的前提下,建立BP神经网络回归模型(BPNN)、支持向量机回归模型(SVM),并与XGBoost回归模型(XGBR)预测结果进行对比分析。试验结果表明,BPNN模型易产生“过拟合”现象,SVM计算速度缓慢且依赖于超参数的选择,XGBoost的计算速度和预测精度均优于其他模型。

关键词:相空间重构;分布式光纤;XGBoost;SVM;BPNN

中图分类号:TD353

文献标志码:A

文章编号:2096-7187(2020)04-3035-08

Study on time series prediction of rock pressure by XGBoost in optical fiber monitoring

CHAI Jing^{1,2}, WANG Runpei¹, DU Wengang¹, LEI Wulin¹, ZHU Xubao¹

(1. College of Energy Engineering, Xi'an University of Science and Technology, Xi'an 710054, China; 2. Key Laboratory of Western Mine Exploitation and Hazard Prevention, Ministry of Education, Xi'an 710054, China)

Abstract: Understanding the ground pressure characterization requires the knowledge of the deformation inside rock mass. In this paper, a three-dimensional physical model was created to simulate longwall mining excavation. The international deformation of the model was measured using a combination of the XGBoost algorithm and distributed fiber technology. The weighting period was detected by introducing the concept of the average frequency shift of the fiber. The XGBoost method was used to establish a time series prediction model of the pressure manifestation. On the premise that the training samples and test samples remain unchanged, the BP neural network regression model (BPNN) and the support vector machine regression model (SVM) were established, and the prediction results of the XGBoost regression model (XGBR) were compared and analyzed. The test results show that the BPNN model is prone to "overfitting", the SVM calculation speed is slow and depends on the choice of hyperparameters, and the calculation

收稿日期:2019-10-29

修回日期:2020-03-06

责任编辑:许书阁

基金项目:国家自然科学基金资助项目(41027002, 51804244)

作者简介:柴敬(1964—),男,宁夏平罗人,教授,博士生导师。主要从事采矿工程、实验岩石力学及光纤传感方面的研究工作。Tel:029-85583059,

E-mail: chaj@xust.edu.cn

speed and prediction accuracy of XGBoost is better than other models.

Key words: phase space reconstruction; distributed fiber; XGBoost; SVM; BPNN

煤矿开采过程中,关键层会出现初次破断和周期破断,上覆岩层大面积悬空引起大步距破断,从而导致矿山压力显现剧烈,甚至发生冲击矿压等矿山灾害。因此,为了减少灾害事故的发生,掌握矿压规律并对其进行有效的预测和防治是煤矿安全生产的重要课题之一。光纤传感技术近些年在采矿工程、地质工程等领域具有广泛应用,其具有体积小、质量轻、精度高、可以实时监控等技术优点,为解决岩体内部变形难以监测、矿压规律难以表征等问题提供了技术支撑^[1]。

在岩石破裂失稳过程中,产生的应力使布里渊频移发生变化,基于此原理可实现对岩体内部变形过程的连续监测。经大量试验验证,光纤频移变化度是衡量岩体破裂失稳过程的一个表征量,可以表征矿压显现规律^[2]。由于岩石破裂失稳过程属于非线性变化,光纤传感技术监测到的数据是随机的、非线性的。采集来的光纤监测数据可以看作随开挖次数变化的随机一维的非线性时间序列^[3],常规的预测方法进行时序预测往往精度和效果都不是很好,而机器学习是能够有效解决复杂动态行为的技术。

其中BP神经网络是解决非线性时序的主要手段之一,由于其具有良好的非线性映射、函数拟合、自适应性以及自学习能力被广泛应用在煤矿动力灾害的预测中^[4-7]。但其难以克服维度灾难、“过拟合”等问题,且其泛化性能较差。支持向量机具有结构风险最小化原则成为岩土领域主要研究对象^[8],SVM的特点是具有解决小样本、非线性等优势,由于其是一个凸优化问题可以寻找出局部最优解,这些优势是BP神经网络所不具备的,因此泛化性能比BP神经网络好,已广泛应用到解决各种非线性分类、回归、时间序列预测等问题中^[9-11]。但SVM在处理大样本时计算速度慢,并且其核函数参数和惩罚函数的等超参数大多依据经验选择^[12-14],估计输出及特征识别没有概率意义等缺点,其实际操作中多以经验值为主要依据。鉴于此,需要有一种更有效的算法解决上述问题,2015年由CHEN Tianqi^[15]等提出的一种基于二阶导信息的极限梯度提升算法(Extreme Gradient Boosting, XGBoost),使用

二阶导数信息和加入正则化方法使其在泛化性和计算性能上获得极大的提升,同时在数据特征的选取方面借鉴随机森林的列抽样,即特征随机抽取,起到了降低过拟合风险和减少计算量的作用,其并行计算的特点将大大提高计算效率,适合处理大量复杂数据,具有优异的模型精度和运行效率。在工业互联网行业,将其应用于推荐、预测等实际业务中,如腾讯使用XGBoost算法对微信购物中的网络广告点击率进行预估,以及阿里巴巴电商推荐系统^[16]。然而,学术界对于XGBoost算法近几年才展开讨论,2017年张钰^[17]等将XGBoost应用于轴承机械结构故障诊断中;2017—2019年杨修德^[18-19]等应用于电力系统预测中;在岩土工程领域较少,笔者引入XGBoost方法进行采场覆岩变形中的矿山压力显现规律预测建模,同时与支持向量机和神经网络的方法做对比,结合预测结果分析XGBoost在采场覆岩变形预测中的特点。希望进一步将此算法应用于煤矿开采中。

1 光纤频移表征矿压预测原理

1.1 分布式光纤矿压监测系统

根据光纤传感原理,布里渊频移、应变、温度之间存在线性关系,当温度不变时,光纤只受应变的作用,在实验室中,控制温度在5℃以内,引起的光纤布里渊频移误差可以不计。

布里渊频移值的变化是传感光纤对岩体变形破裂过程中信息变化的反应。当煤层未开挖时,覆岩中光纤的布里渊频移量无明显变化,当煤层开挖时会引起上覆岩层移动导致应力变化。覆岩产生小变形时,不同位置的光纤频移值变化量较小;覆岩产生大变形时,岩层会产生较大离层甚至破断垮落,这时的光纤频移变化量较大。由此可以得到光纤频移变化的大小反映岩层移动的剧烈程度,提出光纤频移变化度表征工作面来压规律。根据文献[2]的方法,将数据归一化后,按工作面开挖前后2次推进的光纤频值做差,得到布里渊频移变化度的概念,定义为

$$D_i = \frac{1}{n} \left(\sum_{i=1}^n |B_j| - \sum_{i=1}^n |B_{j-1}| \right) \quad (1)$$

式中, D_i 为工作面开挖距离 l 时的光纤平均频移变化度; n 为光纤上采样点的个数; B_j 为第 j 次开挖时光纤上某个位置的频移值; B_{j-1} 为第 $j-1$ 次开挖时光纤上某个位置对应的频移值; j 为工作面开挖次数。

通过前后 2 次频移差值来表示该时间段岩体变形程度, 即光纤上采样点的任意位置处的频移变化的绝对值可以表示该位置岩层在某一段时间内的变形大小。其曲线会随开挖次数出现突峰, 因此可以作为上覆岩层初次来压或周期来压的判断指标。

1.2 XGboost 时序预测原理

根据 Takens 理论进行相空间重构, 对光纤频移变化度数据集进行一维投影, 将光纤频移变化度作为随开挖距离变化的时间序列 \mathbf{X} , 光纤频移变化度构成的时间序列数据集为 $\{x_1, \dots, x_i, \dots, x_{n-1}\}$, x_i 为第 i 个采样点开挖一次的光纤频移变化度, $i=1, 2, \dots, n-1$ 。选取嵌入维数 m 和延迟时间 τ 是相空间重构的关键, 这里嵌入维数采用预测误差最小法进行确定, 延迟时间采用互信息法进行确定。由相空间重构后构成的学习样本集为

$$\mathbf{Y} = f(x_i, x_{i+1}, \dots, x_{i+(m-1)\tau}) \quad (2)$$

其中, $i=1, 2, \dots, n-1$, 共 $n-1+(m-1)\tau$ 个样本。已知上一次开挖光纤频移变化度 x_i 预测下一次开挖光线频移变化度 x_{i+1} , 通过学习得到一个函数 $f(x)$, 使得样本以外的 \mathbf{X} 通过 $f(x)$ 的映射找到对应的 \mathbf{Y} 值, 且使真实值与 \mathbf{Y} 尽可能地接近。

将 m 作为输入特征建立训练样本, \mathbf{X} 和 \mathbf{Y} 对应关系为

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_{1+(m-1)\tau} \\ x_2 & x_3 & x_4 & \dots & x_{2+(m-1)\tau} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n-1} & x_n & x_{n+1} & \dots & x_{(n-1)+(m-1)\tau} \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} x_n \\ x_{n+1} \\ \vdots \\ x_{n+(m-1)\tau} \end{bmatrix} \quad (3)$$

XGBoost 为由 2015 年华盛顿大学 CHEN Tianqi^[15] 等提出, 其关键是利用损失函数的二阶泰勒展开, 并加入正则化降低模型复杂度, 避免“过拟合”。将样本集 \mathbf{X} 和 \mathbf{Y} 作为 XGBoost 的输入值和输出值。建立由 K 棵决策树组成的加法模型, 即

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (4)$$

式中, $f_k(x_i)$ 为第 k 棵决策树, 所有树组成函数空间 F ; x_i 为输入模型中的特征向量; \hat{y}_i 为输出结果。

加入正则化的加法模型的损失函数为

$$obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 = \sum_{i=1}^n l\left[y_i, \hat{y}_i^{(t-1)} + f_i(x_i)\right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (5)$$

其中, $obj^{(t)}$ 表示目标函数; T 为决策树叶子的个数; w 为叶子的权重; γ 和 λ 为叶子和权重的惩罚项。对损失函数二阶泰勒展开, $x = \hat{y}_i^{(t-1)}$ 泰勒展开将 $f_i(x_i)$ 看作 Δx , 则目标函数为

$$obj^{(t)} \approx \sum_{i=1}^n \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_i(x_i) + \frac{1}{2} h_i f_i(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (6)$$

$$g_i = \partial_{\hat{y}_i^{(t-1)}} (y_i - \hat{y}_i^{(t-1)})^2 = 2(\hat{y}_i^{(t-1)} - y_i)^2 \quad (7)$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 (\hat{y}_i^{(t-1)} - y_i)^2 = 2 \quad (8)$$

其中, g_i 为平方损失函数的一阶导数; h_i 为二阶段导数信息; $l(y_i, \hat{y}_i^{(t-1)})$ 为常数项信息, 去掉常数项信息, 并将决策树定义为 $f_i(x) = w_{q(x)}$, 定义集合 $I_j = \{i | q(x_i) = j\}$ 为所有被划分到树叶子节点 j 的训练样本的集合。根据树叶子节点重新组织为 T , 其损失函数为

$$obj^{(t)} = \sum_{i=1}^n \left[g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2 \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 = \sum_{i=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (9)$$

对式(9)中的 w_j 求偏导并等于 0, 把解得的 w_j 代入式(9)得到最终的目标函数为

$$obj^{(t)} = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (10)$$

1.3 XGBoost 预测建模

根据重构相空间理论和 XGBoost 原理, 建立矿压监测数据 XGBoost 时序预测模型, 算法步骤如下:

Step 1: 将矿压监测数据分成训练集和测试集并进行归一化处理;

Step 2: 确定嵌入维数 m 和延迟时间 τ , 建立用于预测的相空间学习样本;

Step 3: 初始化模型参数, 样本权重, 平均权重赋予所有训练集样本;

Step 4: 进行 m 次迭代, 每次迭代采用分类算法进行分类, 并计算错误率;

Step 5: 计算目标函数, 根据误差更新权重 w ;

Step 6: 判断是否满足终止条件, 如满足进行

Step 7, 否则返回Step 5;

Step 7: 输出优化参数: 树深度, 树的数量;

Step 8: 完成迭代, 构建完成XGBoost预测模型。

2 试验结果分析

2.1 相似材料模型试验

试验数据来自于千秋煤矿首采工作面上覆岩层, 其上覆岩层及煤层共805.32 m, 209.5 m处有巨厚砾层, 该岩层厚410.2 m, 根据工作面地质特征, 选用粉煤灰、石膏、碳酸钙等制作采场覆岩三维立体模型(图1), 模型尺寸为3 000 mm × 2 000 mm × 2 000 mm, 并采用NBX-6055分布式光纤应力分析仪作为测量仪器监测(图2)。



图1 三维立体模型

Fig. 1 Three-dimensional mode

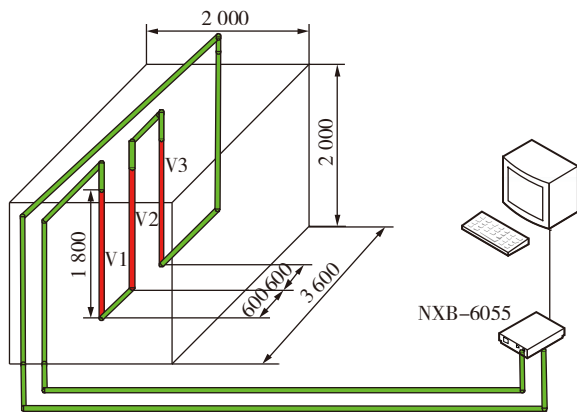


图2 光纤监测系统

Fig. 2 Fiber optic monitoring system

试验过程中选择50 mm空间分辨率和10 mm采样间隔作为测试参数。在模型内部分别布置3条垂直光纤V1, V2, V3。3条光纤之间的间距均为600 mm。主煤层为2号煤层, 厚为16.8 ~ 26.78 mm, 几何相似比1:400, 容重相似比1:1.6。模拟煤层厚度为60 mm, 上覆岩层厚度为174 mm, 地板厚度为200 mm。通过堆积沙子加载, 均匀布置载荷为

3.4 kPa, 从左至右进行开挖, 工作面推进步距40 mm, 共计完成60次开挖, 推进2 400 mm。

2.2 光纤监测来压数据预处理及测试结果

矿压监测数据属于一维监测数据, 使用Min-max标准化, 再对数据进行重构相空间。模型共开挖60次, 这里的延迟时间 τ 可以看作开挖后频移变化的观测间隔, 由互信息法得到最佳延迟时间, 如图3所示, 在延迟时间约为1时出现局部最小值, 即 $\tau=1$, 又被称为采样时间间隔。嵌入维数 m 的选取, 本文采用最终预报误差准则, 以模型误差最小时来确定 m , 经计算当模型中 $m=3$ 时, 模型的误差最小。

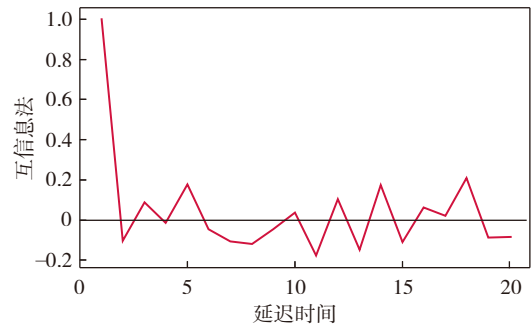


图3 互信息法确定延迟时间

Fig. 3 Mutual information method to determine delay time

三维立体模型工作面开挖过程中, 根据式(1), 将V1, V2, V3光纤的布里渊频移求和, 再除以累采样点数, 可得到工作面光纤频移变化度, 在开挖过程中共出现14次布里渊频移变化度“尖峰”, 在工作面开挖680 mm时出现首次光纤频移变化度突增, 即初次来压。随后开挖到960, 1 080, 1 200, 1 320, 1 440, 1 560, 1 680, 1 840, 1 960, 2 080, 2 200, 2 280, 2 400 mm, 光纤频移变化度发生阶跃点, 此时阶跃点为来压位置。通过光纤测得数据为样本集, 具体见表1。

3 数据分析对比

3.1 预测矿压数据

模型数据采用5折交叉验证提高泛化性能, 即将数据分5份, 取前4份共48次开挖为训练数据, 剩下一份共12次开挖为测试数据, 进行5次试验。试验准确率的平均值作为最终的准确率。

(1) SVM模型

采用网格寻优算法对SVM模型参数寻优, 结果表明, SVM核函数选取高斯径向基核函数对模型的

表 1 光纤频移变化度及周期来压
Table 1 Fiber frequency shift change and pressure

序号	工作面开挖步距/mm	光纤频移变化度/MHa	是否有周期来压	序号	工作面开挖步距/mm	光纤频移变化度/MHa	是否有周期来压
1	40	0.79	否	31	1 240	127.70	否
2	80	0.23	否	32	1 280	129.19	否
3	120	0.23	否	33	1 320	141.12	是
4	160	0.23	否	34	1 360	122.79	否
5	200	0.23	否	35	1 400	122.32	否
6	240	0.25	否	36	1 440	149.95	是
7	280	0.22	否	37	1 480	118.57	否
8	320	0.21	否	38	1 520	123.58	否
9	360	0.23	否	39	1 560	153.26	是
10	400	0.23	否	40	1 600	127.46	否
11	440	0.93	否	41	1 640	128.35	否
12	480	0.79	否	42	1 680	151.68	是
13	520	0.60	否	43	1 720	132.63	否
14	560	2.62	否	44	1 760	121.97	否
15	600	5.51	否	45	1 800	123.01	否
16	640	6.07	否	46	1 840	179.27	是
17	680	19.28	是	47	1 880	124.88	否
18	720	19.25	否	48	1 920	123.29	否
19	760	21.04	否	49	1 960	182.71	是
20	800	21.55	否	50	2 000	123.35	否
21	840	23.52	否	51	2 040	121.13	否
22	880	24.27	否	52	2 080	184.80	是
23	920	26.83	否	53	2 120	121.97	否
24	960	55.64	是	54	2 160	122.60	否
25	1 000	44.43	否	55	2 200	184.69	是
26	1 040	44.80	否	56	2 240	133.67	否
27	1 080	86.92	是	57	2 280	184.76	是
28	1 120	72.86	否	58	2 320	142.85	否
29	1 160	71.93	否	59	2 360	135.03	否
30	1 200	138.40	是	60	2 400	182.60	是

预测结果较好,模型中的3个参数,惩罚函数系数 C 的取值范围为 $[0, 100, 10]$,表示为从0到100之间每间隔10取一个值,核函数的超参数 g 的取值范围为 $[0.1, 1, 0.01]$,表示从0.1到1之间每间隔0.01取一个值,结果为 $C=100, g=0.1$,松弛系数 ε 采用python自带学习库sklearn中的默认数值0.1。

(2) BPNN模型

对于建立的多层神经网络回归模型BPNN,输入层神经元个数为混沌时间序列嵌入维数3,输出层的神经元个数为预测的步数取1,模型泛化能力关键在于隐含层的层数及隐含层神经元的个数,模型采用两层隐含层BP神经模型,神经元个数为(5,

2)。

(3) XGBR

对于XGBoost算法,树的深度和数量的选取影响着模型预测精度和计算时间。如图4所示,在相同样本集下模型训练过程中,由树的数量与模型评分之间的关系,可以看到,当树的数量在60时,模型的评分开始收敛到一定的范围内。由图5可知,随着树的深度增加,在0~6之间模型出现不稳定,在0~3之间模型评分随着树的深度增加评分不断下降直至模型评分达到最低,随后在3~6之间模型评分再不断增加并出现“尖峰”,直至树的深度到达6时模型开始收敛,不再发生变化。因此,模型树的

深度和数量分别为6和60时,能使模型达到最佳,又可以提高模型的计算速度。

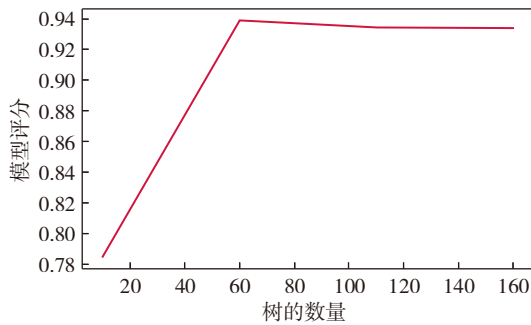


图4 模型评分与树的数量的关系

Fig. 4 Relationship between model score and number of tree

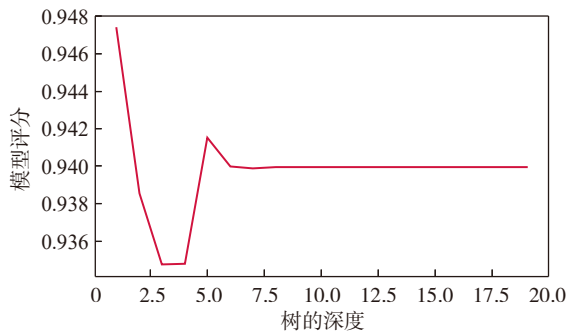


图5 模型评分与树的深度的关系

Fig. 5 Relationship between model score and tree depth

上述3种模型训练和预测结果如图6所示,可知,测试集取实测值最后12次开挖,共出现5次布里渊频移变化度尖峰,分别是在1 960, 2 080, 2 200, 2 280, 2 400 mm出现尖峰,尖峰反应即周期来压。对于矿压时间序列数据的开挖预测,可以看出XGBoost预测性能较好,12次开挖过程中的5次周期来压都被预测出来。SVM在前8次开挖过程中预测的精度较高,后4次开挖的预测结果相对较差。BPNN在12次开挖过程中正确预测出3次周期来压,

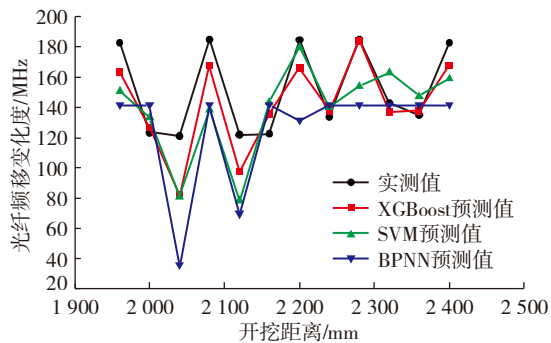


图6 光纤频移变化度实测值和预测值关系

Fig. 6 Optical fiber frequency shift change degree measured value and predicted value relationship diagram

预测效果相对较差,12次开挖过程仅正确预测2次周期来压,最后6次开挖误差较大。因此,结果说明XGBoost算法的预测效果好于SVM和BPNN。

为了更清楚地对比3种模型的性能,表2分别从均方误差、决定系数 R^2 、计算时间进行对比。保持训练样本和测试样本不变的条件下,模型的性能评估采用回归模型的评估指标,均方误差(MSE)、决定系数 R^2 , $f(x_i)$ 为预测值, y_i 为真实值,其定义分别为

$$\text{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2} \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f(x_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

表2 不同模型性能指标对比

Table 2 Comparison of performance indicators of different models

数据设置	回归模型	模型性能指标		计算时间/s
		MSE	R^2	
训练	XGBR	0.008 0	0.919 1	0.724 9
	SVM	0.040 0	0.879 0	2.339 9
	BPNN	0.080 7	0.878 9	4.822 7
测试	XGBR	0.011 5	0.910 2	
	SVM	0.049 7	0.871 8	
	BPNN	0.099 1	0.861 8	

表2中, R^2 是衡量模型对数据上的信息量的捕捉,反映了对信息的拟合度的好坏,其数值越接近1模型表现越好,均值误差(MSE)为关于预测值与真正值之间差异的平均误差,其数值越接近0模型表现越好。由表2可知,在测试阶段XGBR的 R^2 最接近1,MSE最接近0,且与训练阶段差别较小,具有较强的泛化性。其次SVM模型在测试样本和训练样本上表现差异较小,具有较好的泛化能力,由于模型较为复杂,其计算性能相对于XGBR有明显劣势。BPNN模型测试样本的MSE相对于训练样本增多, R^2 数值减小,说明BPNN模型存在“过拟合”现象,会引起测试样本误差增大。

3.2 工程对比分析

通过上述模型试验分析可以看到,XGBR的预测效果要高于传统的BP神经网络和支持向量机。千秋煤矿的周期来压可以由支架压力监测系统获得,以此数据来验证光纤监测的数据结合XGBR在工程应用上的准确性。由几何相似比为1:400可知,取现场开挖760~960 m的支架压力监测数据,

与相似模型同一开挖阶段XGBR预测的光纤频移变化度作对比,由图7可知,XGBR预测的光纤频移变化度表征的矿压趋势和支架平均支护强度基本一致,在工作面开挖760~960 m过程中,支架压力系统共监测到工作面5次周期来压,分别在开挖到784,832,880,928 m及960 m处出现周期来压,前3次和第5次周期来压对应了XGBR预测的光纤频移度峰值。第4次周期来压没有正确预测出来,且在开挖至900 m之后误差逐渐增大,但相对来说基本预测出整体趋势。周期来压时,光纤结合XGBR算法具有比超前于支架压力变化的特性,因此可以通过光纤监测结合XGBR算法来对工作面周期来压进行预测。

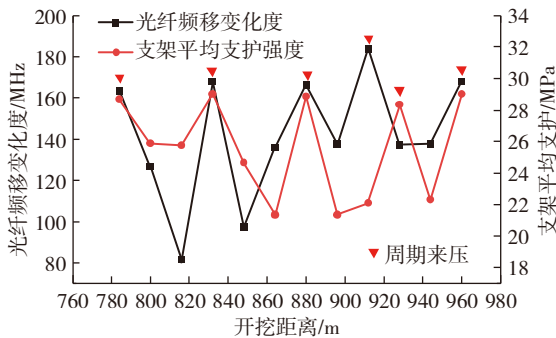


图7 XGBR预测值与支架支护强度的关系

Fig. 7 Relationship between predicted value of XGBR and support strength

4 结论

(1) 通过分布式光纤传感器技术结合XGBoost算法,构建基于相空间重构的矿压数据,并通过大型三维立体模型试验进行验证,建立光纤监测系统,对比分析表明,XGBoost在煤矿开采过程中预测矿压显现是一种可行、有效和准确的方法。

(2) 取相同样本数据分别用SVM和BPNN方法进行预测。试验结果表明,BPNN模型在其训练样本上表现良好,但在测试样本上易出现“过拟合”现象,会引起预测误差增大;SVM预测性能依赖于超参数的选择,网格搜索的速度相对较慢;XGBoost具有较强的泛化性,其拥有最佳的训练结果,最快的计算速度。

(3) 通过支架压力监测到的周期来压数据来验证算法的准确性,并得到较高的精度。因此XGBoost在矿压预测中具有较高的推广价值。

参考文献(References):

- [1] 柴敬,袁强,李毅,等. 采场覆岩变形的分布式光纤检测试验研究[J]. 岩石力学与工程学报,2016,35(S2):150-157.
CHAI Jing, YUAN Qiang, LI Yi, et al. Distributed optical fiber testing of overburden deformation in slope[J]. Journal of Rock Mechanics and Engineering, 2016, 35(S2): 150-157.
- [2] 柴敬,霍晓斌,钱云云,等. 采场覆岩变形和来压判别的分布式光纤监测模型试验[J]. 煤炭学报,2018,43(S1):42-49.
CHAI Jing, HUO Xiaobin, QIAN Yunyun, et al. Distributed optical fiber monitoring model test for deformation and pressure discrimination of overburden rock in slope[J]. Journal of China Coal Society, 2018, 43(S1): 42-49.
- [3] 陶慧,马小平,乔美英,等. 基于多变量混沌时间序列的冲击地压预测[J]. 煤炭学报,2012,37(10):1624-1629.
TAO Hui, MA Xiaoping, QIAO Meiyang, et al. Rockburst prediction based on multivariate chaotic time series[J]. Journal of China Coal Society, 2012, 37(10): 1624-1629.
- [4] 王雨虹,刘璐璐,付华,等. 基于改进BP神经网络的煤矿冲击地压预测方法研究[J]. 煤炭科学技术,2017,45(10):41-45.
WANG Yuhong, LIU Lulu, FU Hua, et al. Research on coal mine rock burst prediction method based on improved BP neural network[J]. Coal Science and Technology, 2017, 45(10): 41-45.
- [5] 谢丽蓉,路朋,王晋瑞,等. 基于LVQ-CPSO-BP算法的煤体瓦斯渗透率预测方法[J]. 采矿与安全工程学报,2017,34(3):398-404.
XIE Lirong, LU Peng, WANG Jinrui, et al. Coal gas permeability prediction method based on LVQ-CPSO-BP algorithm[J]. Journal of Mining and Safety Engineering, 2017, 34(3): 398-404.
- [6] 李慧民,李振雷,何荣军,等. 基于粒子群算法和BP神经网络的冲击危险性评估[J]. 采矿与安全工程学报,2014,31(2):203-207.
LI Huimin, LI Zhenlei, HE Rongjun, et al. Impact risk assessment based on particle swarm optimization and BP neural network[J]. Journal of Mining and Safety Engineering, 2014, 31(2): 203-207.
- [7] 李振华,许延春,李龙飞,等. 基于BP神经网络的导水裂隙带高度预测[J]. 采矿与安全工程学报,2015,32(6):905-910.
LI Zhenhua, XU Yanchun, LI Longfei, et al. Prediction of height of water-conducting fracture zone based on BP neural network[J]. Journal of Mining and Safety Engineering, 2015, 32(6): 905-910.
- [8] 赵洪波. 基于支持向量机的边坡可靠性分析[J]. 岩土工程学报,2007,29(6):819-823.
ZHAO Hongbo. Slope reliability analysis based on support vector machine[J]. Journal of Geotechnical Engineering, 2007, 29(6): 819-823.
- [9] 孙刚,周华平,孙克雷,等. 基于支持向量机的矿压监测数据预测研究[J]. 阜阳师范学院学报(自然科学版),2013,30(3):46-49.
SUN Gang, ZHOU Huaping, SUN Clay, et al. Prediction of mine pressure monitoring data based on support vector machine[J]. Journal of Fuyang Teachers College(Natural Science), 2013, 30(3): 46-49.
- [10] 董辉,傅鹤林,冷伍明,等. 支持向量机的时间序列回归与预测[J]. 系统仿真学报,2006,18(7):1785-1788.

- DONG Hui, FU Helin, LENG Wuming, et al. Time series regression and prediction of support vector machine[J]. *Journal of Systems Simulation*, 2006, 18(7): 1785-1788.
- [11] 来兴平, 张海燕, 刘叶玲, 等. 支持向量机在岩石破裂失稳声发射定位实验中的应用[J]. *金属矿山*, 2006(12): 61-64.
- LAI Xingping, ZHANG Haiyan, LIU Yeling, et al. Support vector machine in acoustic emission location experiment of rock fracture instability[J]. *Metal Mine*, 2006(12): 61-64.
- [12] 单亚锋, 孙朋, 徐耀松, 等. 基于PSO-SVM的煤岩声发射源定位预测[J]. *传感技术学报*, 2013, 26(3): 402-406.
- SHAN Yafeng, SUN Peng, XU Yaosong, et al. Location prediction of coal and rock acoustic emission sources based on PSO-SVM[J]. *Journal of Sensing Technology*, 2013, 26(3): 402-406.
- [13] 邓军, 雷昌奎, 曹凯, 等. 煤自燃预测的支持向量回归方法[J]. *西安科技大学学报*, 2018, 38(2): 175-180.
- DENG Jun, LEI Changkui, CAO Kai, et al. Support vector regression method for coal spontaneous combustion prediction[J]. *Journal of Xi'an University of Science and Technology*, 2018, 38(2): 175-180.
- [14] 柴华彬, 张俊鹏, 严超, 等. 基于GA-SVR的采动覆岩导水裂隙带高度预测[J]. *采矿与安全工程学报*, 2018, 35(3): 359-365.
- CHAI Huabin, ZHANG Junpeng, YAN Chao, et al. Prediction of height of water-conducting fracture zone in mining overburden based on GA-SVR[J]. *Journal of Mining and Safety Engineering*, 2018, 35(3): 359-365.
- [15] CHEN T, GUESTRIN C. XGBoost: A scalable tree boosting system [A]. *The 22nd ACM SIGKDD International Conference*[C]. doi: 10.1145/2939672.2939785, 2016.
- [16] 张昊, 纪宏超, 张红宇, 等. XGBoost算法在电子商务商品推荐中的应用[J]. *物联网技术*, 2017(2): 108-110.
- ZHANG Hao, JI Hongchao, ZHANG Hongyu, et al. XGBoost algorithm in e-commerce commodity recommendation[J]. *Internet of Things Technology*, 2017(2): 108-110.
- [17] 张钰, 陈珺, 王晓峰, 等. XGBoost在滚动轴承故障诊断中的应用[J]. *噪声与振动控制*, 2017(4): 168-173.
- ZHANG Yu, CHEN Jun, WANG Xiaofeng, et al. Application of XGBoost in fault diagnosis of rolling bearings[J]. *Noise and Vibration Control*, 2017(4): 168-173.
- [18] 杨修德, 王金梅, 张丽娜, 等. 基于XGBoost的多维度超短期负荷预测研究[J]. *电气自动化*, 2019(1): 36-38.
- YANG Xiude, WANG Jinmei, ZHANG Lina, et al. Multidimensional ultra-short term load forecasting research based on XGBoost[J]. *Electrical Automation*, 2019(1): 36-38.
- [19] 杨修德, 王金梅, 张丽娜, 等. XGBoost在超短期负荷预测中的应用[J]. *电气传动自动化*, 2017(4): 21-25.
- YANG Xiude, WANG Jinmei, ZHANG Lina, et al. XGBoost in ultra-short term load forecasting[J]. *Electric Drive Automation*, 2017(4): 21-25.
- [20] 陶慧. 基于监测时间序列的冲击地压混沌特性分析及其智能预测研究[D]. 北京: 中国矿业大学(北京), 2013.
- TAO Hui. Analysis of chaotic characteristics of rock burst based on monitoring time series and its intelligent prediction[D]. Beijing: China University of Mining and Technology(Beijing), 2013.
- [21] 王海燕, 卢山. 非线性时间序列分析及其应用[M]. 北京: 科学出版社, 2006.
- WANG Haiyan, LU Shan. Nonlinear time series analysis and its application [M]. Beijing: Science Press, 2006.
- [22] 袁强. 采动覆岩变形的分布式光纤检测与表征模拟试验研究[D]. 西安: 西安科技大学, 2017.
- YUAN Qiang. Distributed optical fiber detection and characterization of overburden deformation caused by mining [D]. Xi'an: Xi'an University of Science and Technology, 2017.