

基于 GA-SVM 耦合模型的煤与瓦斯突出危险性预测

薛峰, 李希建*, 徐恩宇

(贵州大学 矿业学院, 贵州 贵阳 550025; 贵州大学 复杂地质矿山开采安全技术工程中心, 贵州 贵阳 550025;
贵州大学 瓦斯灾害防治与煤层气开发研究所, 贵州 贵阳 550025)

摘要: 为了更加科学地预测煤与瓦斯突出危险性, 提出一种基于遗传算法(GA)的支持向量机(SVM)参数选择方法, 选取瓦斯含量、开采深度、瓦斯压力、瓦斯放散初速度、煤坚固性系数、煤体破坏类型这 6 个主要影响因素研究煤与瓦斯突出危险性, 建立 GA-SVM 瓦斯突出危险性预测模型, 并结合工程实例进行验证. 研究表明: 样本突出危险性预测准确度达到 91.67%, 预测结果与实际情况相符. 采用遗传算法与支持向量机模型预测, 有效降低人为因素带来的误差, 对提高煤与瓦斯突出危险性预测精度具有一定的参考意义.

关键词: 煤与瓦斯突出; 遗传算法(GA); 支持向量机(SVM); 突出预测; 预测模型

中图分类号: TD77+1 **文献标志码:** A **文章编号:** 1672-9102(2022)03-0039-06

Application of GA-SVM Coupling Model in Prediction Coal and Gas Outburst

XUE Feng, LI Xijian, XU Enyu

(Mining College, Guizhou University, Guiyang 550025, China; Engineering Center for Safe Mining Technology Under Complex Geologic Condition, Guizhou University, Guiyang 550025, China; Institute of Gas Disaster Prevention and Coalbed Methane Development, Guizhou University, Guiyang 550025, China)

Abstract: In order to predict the risk of coal and gas outburst more scientifically, a parameter selection method of support vector machine (SVM) based on genetic algorithm (GA) is proposed. Six main influencing factors, including gas content, mining depth, gas pressure, initial velocity of gas emission, coal firmness coefficient and coal failure type are selected to study the risk of coal and gas outburst, and the GA-SVM gas outburst risk prediction model is established, which is verified by examples. Results show that the outstanding prediction accuracy of the prediction sample reaches 91.67%, and the prediction results are consistent with the actual situation. Through the prediction of genetic algorithm and support vector machine model, the error caused by human factors is effectively reduced. It can provide some references for improving the accuracy of coal and gas outburst risk prediction.

Keywords: coal and gas outburst; genetic algorithm (GA); support vector machine (SVM); outburst forecast; prediction model

煤与瓦斯突出是煤矿安全生产最严重的动力灾害之一^[1]. 随着开采深度增加, 地应力增大, 瓦斯突出

收稿日期: 2021-10-07

基金项目: 国家自然科学基金资助项目(51874107; 52164015); 贵州省省级科技计划资助项目(黔科合平台人才[2018]5781号; 黔科合支撑[2022]一般231)

* 通信作者, E-mail: 575914635@qq.com

灾害特征更为明显.煤与瓦斯突出受多种影响因素控制,且机理复杂,给煤与瓦斯突出预测带来了极大困难^[2].因此,研究煤与瓦斯突出与其影响因素的内在联系,对煤与瓦斯突出预测具有重要意义.

目前,已有众多学者对煤与瓦斯突出做了大量研究.匡亮等^[3]基于瓦斯突出综合作用假说机理,选取突出预测评价指标,构建隧道煤与瓦斯突出预测网络模型,得到了较好的预测结果,验证了该模型在隧道应用的可行性;张友谊等^[4]通过运用多类别距离判别法和网络分析法(Aalytic Network Process, ANP),对灰色关联模型进行改进,构建多指标耦合预测模型进行预测;李成武等^[5]构建突出能量预测模型并应用于实例,验证了该模型可行性;王刚等^[6]通过建立煤与瓦斯三维结构模型,结合实际案例分析,验证模型实用性;李长兴等^[7]构建 Fisher 逐步判别分析模型进行煤与瓦斯突出预测;朱宝合等^[8-9]基于非线性支持向量机理论,构建支持向量机(Support Vector Machine, SVM)预测模型,对煤与瓦斯突出进行预测.综上所述,前人多考虑煤与瓦斯突出单因素或简单综合因素,对多因素之间的非线性关系少有研究,且在分析时,常人为设定参数进行预测,导致预测模型精准度不高,可靠性较低.

因此,为了更合理地煤与瓦斯突出危险性进行预测,选取瓦斯含量、开采深度、瓦斯压力、瓦斯放散初速度、煤坚固性系数、煤体破坏类型这6个影响煤与瓦斯突出的主要因素,基于遗传算法(Genetic Algorithm, GA)和支持向量机(SVM),建立煤与瓦斯突出危险性预测模型,分析煤与瓦斯突出的各影响因素之间的非线性映射关系,以期对煤与瓦斯突出危险性预测提供一定参考.

1 GA-SVM 煤与瓦斯突出预测模型

1.1 遗传算法

遗传算法是一种模拟达尔文进化学说中优胜劣汰、适者生存的计算模型^[10].算法计算步骤如下:

- 1) 随机产生初始种群,通过初始种群模拟达尔文进化理论,得出问题假设最优解.
- 2) 选择合适的编码方式对参数进行编码^[11],得出离散变量.
- 3) 设置个体适应度值,计算个体被选择的概率,概率计算公式如式(1)所示.

$$p(x_i) = \frac{f(x_i)}{\sum_{j=1}^n f(x_j)}. \quad (1)$$

式中: $p(x_i)$ 为每个个体被选择的概率; $f(x_i)$ 为设置的个体适应度值; $i = 1, 2, 3, \dots, n$, n 为个体总数.

- 4) 计算个体累加适应度值 $F(x_i)$ 与个体累加被选择概率 $P(x_i)$,如式(2)所示.

$$\begin{cases} F(x_i) = \sum_{j=1}^i f(x_j); \\ P(x_i) = \sum_{j=1}^i p(x_j). \end{cases} \quad (2)$$

5) 计算交叉概率.在繁衍下一代个体时,为产生新个体,需要交换两个不同个体中相同位置的基因序列,这种交换行为称之为交叉.若交叉概率过大,则产生新个体越快,同时使具有高适应度值的个体被遗忘掉;若交叉概率过小,则搜索全局最优过程缓慢.因此,提出一种交叉概率随适应度值变化而变化的全局搜索方法.交叉概率 P_c 为

$$P_c = \begin{cases} \frac{k_1(f_{\max} - f)}{f_{\max} - f_{\text{avg}}}, f \geq f_{\text{avg}}; \\ k_2, f < f_{\text{avg}}. \end{cases} \quad (3)$$

式中: f_{\max} 为群体中的最大适应度值; f_{avg} 为群体平均适应度值; f 为需要交叉的两个个体中较大的适应度值; k_1 和 k_2 为常数.

- 6) 计算变异概率 P_m ,如式(4)所示.

$$P_m = \begin{cases} k_3(f_{\max} - v), v \geq f_{\text{avg}}; \\ f_{\max} - f_{\text{avg}}, v < f_{\text{avg}}. \end{cases} \quad (4)$$

式中: v 为将要变异个体的适应度值; k_3 和 k_4 为常数.

1.2 支持向量机原理

支持向量机(SVM)是一种既可用于线性分类,又可用于非线性分类的数学模型^[12].通过输入训练样本和学习目标,使 SVM 二类分类器进行学习并应用于新数据,建立分类超曲面作为决策曲面,将学习目标按正类和负类分开,并使两个平行分类超平面间隔边缘距离最大化^[12].

假定带标签样本集 $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$, 其中 \mathbf{x}_i 为给定输出数据的特征向量, $\mathbf{x}_i \in \mathbf{R}^d$; y_i 为二分类标签, $y_i \in \{-1, 1\}$, $y_i = 1$ 表示正类, $y_i = -1$ 表示负类.则存在一个决策边界的分类超曲面,使得正类和负类样本分布在超曲面两侧,满足式(5).

$$\begin{cases} y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1; \\ \mathbf{w}^T \mathbf{x} + b = 0. \end{cases} \quad (5)$$

式中: \mathbf{w} 为超平面法向量; b 为超平面截距; \mathbf{x} 为超平面样本特征向量.

为求解最优分类面,使分类超平面间隔距离最大化,优化计算问题,需满足式(6)的条件.

$$\begin{cases} \max \frac{2}{\sqrt{\mathbf{w}\mathbf{w}^T}} \Rightarrow \min \frac{1}{2} \mathbf{w}^T \mathbf{w}, \mathbf{w} \in \mathbf{R}^d; \\ \text{s.t.} (\mathbf{w}^T \mathbf{x}_i + b) y_i \geq 1. \end{cases} \quad (6)$$

为得到较满意的分类超曲面,引入松弛变量 ξ_i .当松弛变量为 0 时,所有样本点都能被正确分类;当松弛变量为无穷大时,任意样本都满足.为限制 ξ_i 的大小,引入惩罚系数 C 求解分类超曲面,满足式(7)的条件.

$$\begin{cases} \min \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_i \xi_i; \\ y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i. \end{cases} \quad (7)$$

最后,选择高斯径向基核函数构建最优化目标函数,找寻最佳决策面,将初步优化问题转化为对偶问题.如式(8)所示.

$$\begin{cases} K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{2\sigma^2}\right); \\ \max \sum_i a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) a_i a_j; \\ \sum_{i=1}^n a_i y_i = 0. \end{cases} \quad (8)$$

式中: $K(\cdot)$ 为高斯径向基核函数; σ 为高斯径向基核函数宽度; a_i 为拉格朗日乘子.

1.3 核函数和参数确定

核函数是内积的一般化形式^[13].高斯核函数(RBF)具有较好的性能,且相比其他核函数所需参数少.因此,选择高斯核函数进行预测,具有良好的映射能力^[14].高斯核函数半径 g 与惩罚系数 C 由遗传算法进行确定.

1.4 GA-SVM 耦合

- 1) 为保证训练样本包含充足的数据信息,将数据尽可能按 3 : 1 的比例分成训练样本集和预测样本集.
- 2) 为提高模型准确度和收敛速度,将数据进行归一化预处理,以消除各因素之间由于量纲和单位所带来的误差影响.
- 3) 为得到离散型变量及定理分析,选择二进制编码方式对 SVM 参数进行编码,并确定惩罚系数 C 和

高斯核函数相关参数的变化区域.

- 4) 输入 GA 初始参数, 并随即生成初始种群.
- 5) 选择 SVM 核函数对小样本数据进行分析, 本文选择高斯核函数进行突出强度预测.
- 6) 对 SVM 进行训练, 并计算分类准确度.
- 7) 计算适应度函数值, 根据交叉验证搜索凸优化处理后的惩罚系数 C 和高斯核函数半径 g .
- 8) 分类准确度若达到遗传算法适应度函数值, 则进行解码操作; 若未达到, 则进行选择、变异、交叉, 产生新种群, 程序转到步骤 6) 继续执行, 直至跳出循环.

上述 GA-SVM 耦合流程如图 1 所示.

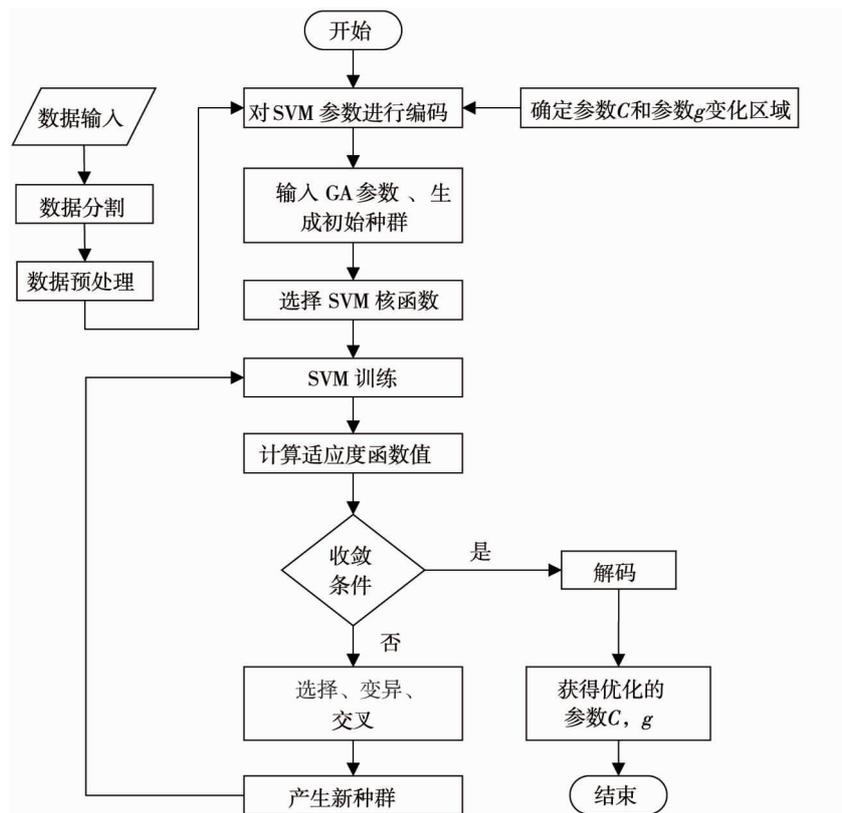


图 1 GA-SVM 耦合流程

2 工程应用

2.1 判别指标的选取

国内外学者普遍认同煤与瓦斯突出是地应力、煤的物理性质、煤层瓦斯等因素共同作用的结果^[15-16]. 煤与瓦斯突出是在一定压力作用下涌出大量煤和瓦斯, 突出灾害通常是由于地应力扰动破坏了煤层内部平衡, 地应力影响范围的重要因素是开采深度; 瓦斯放散初速度是衡量煤层受动力扰动下释放瓦斯能力指标之一; 煤的坚固性系数是在突出发展过程中突出能量所受阻力大小的重要指标. 综上, 收集文献[17-19]的数据并分析, 选取开采深度、瓦斯含量、瓦斯压力、瓦斯放散初速度、煤的坚固性系数、煤体破坏类型这 6 个主要因素作为 GA-SVM 耦合模型的输入数据

根据煤与瓦斯突出的煤量将突出强度分为 4 类 (见表 1), 突出强度作为预测模型的输出结果.

表 1 突出强度分类

突出煤量/瓦斯量	无	<50 t	50~100 t	>100 t
突出强度	无突出	小型突出	中型突出	大中型突出
类别	I	II	III	IV

2.2 数据处理

依据上述理论分析构建 GA-SVM 煤与瓦斯突出预测模型.选取国内 49 个典型的煤与瓦斯突出矿井的数据作为数据样本^[17-19].采用 MATLAB R2016b 中 mapminmax 函数对数据进行归一化预处理,以消除各因素之间由于量纲和单位不同所带来的误差,避免影响预测结果的精度^[20].

2.3 应用分析

根据 1.4 节所建的 GA-SVM 耦合模型,将样本中 37 个突出矿井的数据作为训练样本进行学习,并将训练所得预测模型对剩余的 12 个测试样本进行突出预测,结果如表 2 所示.

表 2 GA-SVM 模型预测结果

样本序号	瓦斯含量/ ($\text{m}^3 \cdot \text{t}^{-1}$)	开采深度/m	瓦斯压力/ MPa	放散初速度/ ($\text{mL} \cdot \text{s}^{-1}$)	坚固系数/ ($\text{kg} \cdot \text{cm}^{-2}$)	煤体破坏类型	实际突出强度类别	GA-SVM 预测突出强度类别
1	9.40	463	2.11	6.90	0.49	未受破坏	I	I
2	9.50	460	2.00	7.00	0.48	未受破坏	I	I
3	10.25	425	2.80	8.00	0.59	破坏成块	II	II
4	9.86	510	2.16	14.00	0.34	破坏成粒	II	II
5	10.10	520	2.26	12.00	0.35	破坏成粒	II	II
6	9.48	540	1.21	9.38	0.26	破坏成粉	II	II
7	12.25	737	3.12	17.51	0.24	破坏成块	III	III
8	29.59	489	3.02	26.52	0.31	破坏成块	IV	IV
9	8.19	550	3.87	14.38	0.23	破坏成块	IV	IV
10	26.44	451	2.58	20.12	0.25	轻微破坏	IV	III
11	10.10	836	1.19	12.54	0.27	破坏成块	IV	IV
12	13.20	445	0.95	6.00	0.24	破坏成块	II	III

通过 MATLAB R2016b 中支持向量机 LIBSVM 开源开发包^[21],编写 GA-SVM 耦合模型预测程序,对数据采用二进制编码,经过大量测试,设定遗传算法最大进化代数为 50、种群大小为 40、交叉概率为 0.7、变异概率为 0.03 的条件下得出的结果更为可靠.惩罚系数 C 的变化范围为 $[0.1, 100]$,高斯核函数半径 g 的变化范围为 $[0.1, 10]$.运行该程序读取训练样本和预测样本,进行全局寻优迭代,得到最佳适应度函数值和平均适应度函数值,结果如图 2 所示.

通过软件解算,应用交叉理论验证 GA-SVM 预测突出强度准确率为 90.97%,算法全局搜索最优惩罚系数 $C=6.7656$,高斯核函数半径 $g=2.8581$.

采用 GA-SVM 预测模型对 12 组测试样本数据进行预测,结果如图 3 所示.从表 2 和图 3 可以看出:采用 GA-SVM 预测模型预测 12 组样本中有 1 组出错,准确率为 91.67%,基本符合实际情况.

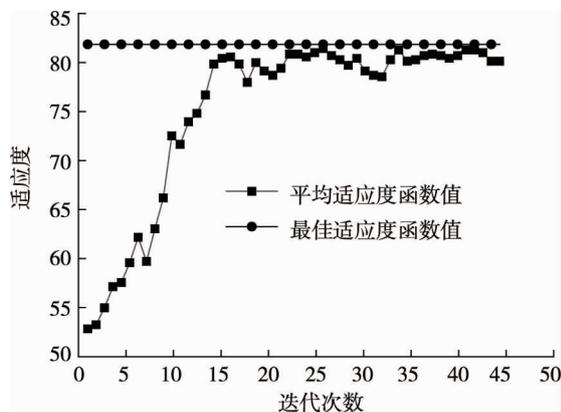


图 2 最佳适应度与平均适应度曲线

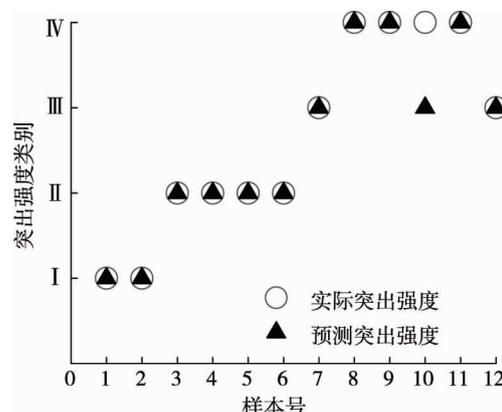


图 3 GA-SVM 预测与实际结果对比

为进一步验证该模型性能,采用 SVM 对上述样本数据进行训练预测,得到分类准确率为 83.33%.结

果表明,基于 GA 优化后的 SVM 预测模型较传统 SVM 准确率更高,更适用于煤与瓦斯突出危险性预测。

3 结论

1)采用遗传算法(GA)可将支持向量机(SVM)中惩罚系数 C 和高斯核函数半径 g 进行全局寻优搜索,既保留了 SVM 解决非线性分类的优点,同时又有效降低了人为因素对参数设定的误差,提高了突出危险性预测的准确率。

2)所构建的 GA-SVM 模型利用 SVM 处理小样本、非线性、高维数据分类的优势,并采用 GA 算法进行参数寻优,避免了已有方法的局限性。

3)基于 GA-SVM 模型的煤与瓦斯突出危险性预测结果与实际情况相符,且优于传统 SVM。

参考文献:

- [1] 潘一山.煤与瓦斯突出、冲击地压复合动力灾害一体化研究[J].煤炭学报,2016,41(1):105-112.
- [2] 南存全,冯夏庭.基于 SVM 的煤与瓦斯突出区域预测研究[J].岩石力学与工程学报,2005,24(2):263-267.
- [3] 匡亮,赵万强,喻渝.BP 神经网络法预测隧道瓦斯突出的模型与实例[J].铁道工程学报,2018,35(2):56-61.
- [4] 张友谊,崔金雷,焦向东.煤与瓦斯突出多指标耦合预测模型研究及应用[J].工程科学学报,2018,40(11):1309-1316.
- [5] 李成武,付帅,解北京,等.煤与瓦斯突出能量预测模型及其在平煤矿区的应用[J].中国矿业大学学报,2018,47(2):231-239.
- [6] 王刚,武猛猛,程卫民,等.煤与瓦斯突出能量条件及突出强度影响因素分析[J].岩土力学,2015,36(10):2974-2982.
- [7] 李长兴,辛程鹏,李回贵,等.基于 Fisher-逐步判别法的煤与瓦斯突出预测[J].中国矿业,2020,29(7):106-111.
- [8] 朱宝合,郑邦友,戴亦军,等.基于非线性支持向量机的隧道煤与瓦斯突出危险性预测[J].现代隧道技术,2020,57(2):20-25.
- [9] 苏筱丽.基于主要成分降维的 SVM 回归模型在煤与瓦斯突出预测中的应用[J].工业计量,2020,30(1):74-77.
- [10] 李书全,孙雪,孙德辉,等.遗传算法中的交叉算子的述评[J].计算机工程与应用,2012,48(1):36-39.
- [11] 张超群,郑建国,钱洁.遗传算法编码方案比较[J].计算机应用研究,2011,28(3):819-822.
- [12] 王道明,鲁昌华,蒋薇薇,等.基于粒子群算法的决策树 SVM 多分类方法研究[J].电子测量与仪器学报,2015,29(4):611-615.
- [13] 袁泽恒,田润澜,张旭洲.改进加权 SVC 的雷达信号分选新方法[J].现代防御技术,2018,46(3):86-92.
- [14] 李幼军,钟宁,黄佳进,等.基于高斯核函数支持向量机的脑电信号时频特征情感多类识别[J].北京工业大学学报,2018,44(2):234-243.
- [15] 李希建,林柏泉.煤与瓦斯突出机理研究现状及分析[J].煤田地质与勘探,2010,38(1):7-13.
- [16] 彭守建,宋肖微,许江,等.基于综合作用假说煤与瓦斯突出强度影响因素试验[J].煤炭科学技术,2016,44(12):81-84.
- [17] 吴雅琴,李惠君,徐丹妮.基于 IPSO-Powell 优化 SVM 的煤与瓦斯突出预测算法[J].工矿自动化,2020,46(4):46-53.
- [18] 孙福成.基于优化量子门节点神经网络的煤与瓦斯突出预测模型研究[D].阜新:辽宁工程技术大学,2020.
- [19] 孙宇航.基于改进模糊支持向量机的煤与瓦斯突出预测[D].徐州:中国矿业大学,2019.
- [20] 谢国民,谢鸿,付华,等.煤与瓦斯突出预测的 NN-SVM 模型[J].传感技术学报,2016,29(5):733-738.
- [21] 李坤,刘鹏,吕雅洁,等.基于 Spark 的 LIBSVM 参数优选并行化算法[J].南京大学学报(自然科学),2016,52(2):343-352.