



引用格式:万 宇,齐金平,张 儒,等. 基于信息增益优化支持向量机模型的煤矿瓦斯爆炸风险预测[J]. 科学技术与工程, 2021, 21(9): 3544-3549.

Wan Yu, Qi Jinping, Zhang Ru, et al. Risk prediction of coal mine gas explosion based on information gain-support vector machine (IG-SVM) model[J]. Science Technology and Engineering, 2021, 21(9): 3544-3549.

基于信息增益优化支持向量机模型的煤矿瓦斯爆炸风险预测

万 宇, 齐金平*, 张 儒, 闫 森
(兰州交通大学机电技术研究所, 兰州 730070)

摘要 为了探索基于样本数据的煤矿瓦斯爆炸风险预测,依据本质安全理念构建了预测瓦斯爆炸风险的指标集,结合机器学习与特征优化算法提出了信息增益(information gain, IG)与支持向量机(support vector machine, SVM)的组合模型,通过对优化后的14种特征信息的分类学习,完成对风险未知样本的预测任务。以全国100家煤矿企业为研究对象,使用不同模型分别预测瓦斯爆炸风险并全面分析和比较,实验结果表明,经过IG优化后的SVM模型预测正确率达到了95.45%,相对于单一SVM模型提高了9.09%,同时高于其他预测模型,证明了该组合模型在瓦斯爆炸风险预测领域的优越性。

关键词 瓦斯爆炸风险; 本质安全; 支持向量机; 信息增益

中图法分类号 TD76; 文献标志码 A

Risk Prediction of Coal Mine Gas Explosion Based on Information Gain-Support Vector Machine (IG-SVM) Model

WAN Yu, QI Jin-ping*, ZHANG Ru, YAN Sen
(Mechatronics T&R Institute, Lanzhou Jiaotong University, Lanzhou 730070, China)

[Abstract] In order to explore the risk prediction of coal mine gas explosion based on the sample data, an index set for predicting gas explosion risk based on intrinsic safety concept was established. A combined model of information gain(IG) support vector machine (SVM) combined with machine learning and feature optimization algorithm was proposed, the prediction task of unknown risk samples was completed through the classification learning of 14 optimized feature information. 100 coal mining companies across the country were selected as the research objects. Different models were used to predict the risk of gas explosion, and then comprehensively analyzed and compared. The experimental results show that the accuracy of SVM model after IG optimization reaches 95.45%. Compared with the single SVM model, it increases by 9.09%, higher than other prediction models. It proves the superiority of the combined model in the field of gas explosion risk prediction.

[Keywords] gas explosion risk; intrinsic safety; support vector machine(SVM); information gain(IG)

目前,随着社会经济的蓬勃发展,中国已成为世界上最大的煤炭生产、消费国,安全生产已经成为一个重要现实问题^[1]。煤炭行业是典型的高危行业,其事故主要包括瓦斯、顶板、底板、放炮、机电、火灾、水害、运输以及其他事故,其中瓦斯事故一般被认为是威胁性最大的灾害事故。中国高瓦斯矿井数占到了接近总量的一半,每年因瓦斯事故带来的伤亡人数不计其数,而其中瓦斯和煤尘爆炸事故占大多数^[2]。煤炭的安全开采是经济增长的重要保障,“十三五”规划中国家也对能源安全和绿色生产提出了明确的要求。因此,为了响应国家号

召,应将未来的工作重心从事后响应转移到预控预防,做到从根源上遏制安全事故的发生,减少人员伤亡率。

瓦斯爆炸是瓦斯事故中典型的一种,传统预测方法主要依赖专家判断,具有较强的主观性,预测精度不能达到分析任务的要求,近年来随着计算机技术的进步与发展,机器学习算法的引入极大地推动了风险预测领域的进步,田水承等^[3]利用反向传播神经网络(back propagation neural network, BPNN)对掘进面瓦斯爆炸危险进行了安全评价,李润求等^[4]构建了基于区间层次分析法和功效系数

收稿日期: 2020-07-01; 修订日期: 2020-10-24

基金项目: 国家自然科学基金(71861021); 甘肃省高等学校科研项目(2018A-026); 甘肃省重点研发项目(17YF1FA122)

第一作者: 万 宇(1996—),男,汉族,安徽合肥人,硕士研究生。研究方向: 检测技术与自动化装置。E-mail: 1421121975@qq.com。

*通信作者: 齐金平(1978—),男,汉族,山东诸城人,博士,副教授。研究方向: 载运工具运用工程。E-mail: qijinping@mail.lzjtu.cn。

法结合(interval analytic hierarchy process-efficacy coefficient method, IAHP-ECM)的瓦斯爆炸灾害风险评估模型,邵良彬等^[5]针对数据缺失问题将随机森林算法引入对瓦斯灾害进行预测,李爽等^[6]将极限学习机与贝叶斯网络结合对风险进行预测。上述方法从不同角度对瓦斯爆炸风险预测进行了研究,机器学习的应用很大程度上提升了预测的精准度,除了各种算法上的进步,改变优化特征向量的输入也是提升速度精度的一大方向。

现将本质安全理念引入评价指标集的构建,从人、机、管理、环境四个角度构建了28个指标,并针对煤矿瓦斯爆炸灾害的特点,选取支持向量机(support vector machine, SVM)模型来进行分类预测,在此基础上选择信息增益法(information gain, IG)根据熵值优化输入变量,组成了IG-SVM的组合模型,以期能够实现对瓦斯爆炸风险的快速、高精预测。

1 分类预测模型的构建

1.1 支持向量机

支持向量机是一种基于VC维(Vapnik-Chervonenkis dimension)理论和结构风险最小化准则的监督学习算法^[7],最早起源于分类、回归领域,Vapnik^[8]引入核理论,将原始低维数据映射到高维空间,在维数足够高的空间中利用超平面来进行分类,在数学上可归结为求解一个二次规划问题。SVM在面对非线性、小样本等问题时表现出色,计算复杂度取决于支持向量的数目,从而避免了“维数灾难”,并且具有良好的鲁棒性和泛化性能^[9]。目前已广泛应用于计算机视觉、时间序列预测、人工智能等领域。

设样本数据集 $(\mathbf{x}_i, y_i) (i = 1, 2, \dots, l), \mathbf{x}_i \in \mathbf{R}^n, y_i \in \{-1, +1\}$, l 为训练样本总数, n 为空间的维数, \mathbf{x}_i 为待分类数据, y_i 为标记类别,对线性可分的样本,存在一个超平面 H 能够将不同类的样本分开,用法向量表示其方向, H_1, H_2 分别表示过两类样本点中与超平面 H 距离最近的平行平面,其间距离称为分类间隔,在保证能够正确分类的前提下分类间隔最大的平面称为最优超平面。超平面方程为 $\mathbf{w}\mathbf{x} + b = 0$, d 维空间中的线性判别函数为 $g(\mathbf{x}) = \mathbf{w}\mathbf{x} + b$,其中 $\mathbf{w} \in \mathbf{R}^n, \mathbf{w}$ 为参数向量,即超平面的法向量, $b \in \mathbf{R}$ 为分类阈值。

归一化使 $|g(\mathbf{x})| \geq 1$,离分类面最近的样本的 $|g(\mathbf{x})| = 1$,此时分类间隔为 $2/\|\mathbf{w}\|$,要求分类间隔最大,即要求 $\|\mathbf{w}\|$ 最小,要求所有样本分类正确,则需要满足:

$$y_i[(\mathbf{w}\mathbf{x}_i + b)] \geq 1, \quad i = 1, 2, \dots, l \quad (1)$$

上述条件可以转化为下面带约束条件的优化问题进行求解:

$$\begin{cases} \min \Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s. t. } y_i[(\mathbf{w}\mathbf{x}_i + b)] \geq 1, \quad i = 1, 2, \dots, l \end{cases} \quad (2)$$

样本线性可分的情况下,求解即可得到最优分类超平面,对于线性不可分的情况,存在一定的训练误差,一部分无法满足式(1)的样本数据将被视为噪声,为了给这样的噪声数据引入容错性需要在表达式中增加一个松弛变量 $\xi_i (\xi_i \geq 0)$,则式(1)变成

$$y_i[(\mathbf{w}\mathbf{x}_i + b)] \geq 1 - \xi_i, \quad i = 1, 2, \dots, l \quad (3)$$

当 $0 < \xi_i < 1$ 时,样本点正确分类;当 $\xi_i \geq 1$ 时,样本点 \mathbf{x}_i 将被错分。为了解决此问题需要在目标函数中加入惩罚项 $C \sum_{i=1}^l \xi_i$,即

$$\Phi(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (4)$$

由此可以得到线性不可分时的最优分类超平面,称为广义分类超平面,可以表示为

$$\begin{cases} \min \Phi(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s. t. } y_i[(\mathbf{w}\mathbf{x}_i + b)] \geq 1 - \xi_i, \\ \quad i = 1, 2, \dots, l \end{cases} \quad (5)$$

式中: $\sum_{i=1}^l \xi_i$ 表示错分样本数量的上限,用来衡量忽略的噪声数据带来的损失大小,即离群点的偏离程度; $C (C > 0)$ 为惩罚因子,是一个需要提前设定的量,它的大小决定了对上述损失大小的重视程度。 C 取值越小,对目标函数的损失就越小,从而能够获得更大的分类间隔,提升了推广性能;反之 C 取值越大,对目标函数的损失就越大,但因为重视了离群点对分类的影响,能够获得较小的经验风险。通过引入拉格朗日对偶定理将上述最优分类问题转化为

$$\begin{cases} \max L(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \\ \text{s. t. } 0 \leq \alpha \leq c, \quad i, j = 1, 2, \dots, l, \\ \quad \sum_{i=1}^l y_i \alpha_i = 0 \end{cases} \quad (6)$$

式(6)中: α_i 为拉格朗日乘子。在处理非线性问题时,低维空间中样本数据的离群点数量非常多,引入松弛变量也无法取得较好的分类效果。根据泛函理论,只要一种核函数 $K = (\mathbf{x}_i, \mathbf{x}_j)$ 满足Mercer条件,就可以代替原空间中的内积。本文拟采用的核函数有双曲正切(Sigmoid)核函数、径向基核函数(radial basis function, RBF)、线性核函数(linear ker-

nel function, LKF) 以及多项式核函数 (polynomial function, PF)。利用核函数技巧后的最优分类函数为

$$f(x) = \text{sgn} \left[\sum_{j=1}^l a_j y_j K(\mathbf{x}_j, \mathbf{x}) + b \right] \quad (7)$$

1.2 信息增益

信息增益是一种优化特征选取的方法^[10]。信息量是对信息的度量, 信息的大小与随机事件的概率有关, 概率越小产生的信息量越大, 设离散随机变量 X 的概率分布为 $P(X = x_i) = p_i$ ($i = 1, 2, \dots, n$)。信息熵代表所有可能发生事件信息量的期望, 也可以理解为离散随机变量的复杂度, 表达式为

$$H(X) = - \sum_{i=1}^n p_i \lg p_i \quad (8)$$

设有随机变量 (X, Y) , 其联合概率分布为 $P(X = x_i, Y = y_i) = P_{ij}$ ($i = 1, 2, \dots, n, j = 1, 2, \dots, m$)。条件熵 $H(Y|X)$ 代表在已知随机变量 X 的条件下随机变量 Y 的不确定性, 表达式为

$$H(Y|X) = \sum_{i=1}^n p_i H(Y|X = x_i) \quad (9)$$

信息增益代表了在 X 已知的条件下, 信息复杂度减少的程度, 即信息熵与条件熵的差值, 特征 T 对训练数据集 D 所带来的信息增益为

$$g(D, T) = H(D) - H(D|T) \quad (10)$$

某个特征的信息增益值越大, 也就是说根据这个特征所做决定的不确定度减少越多。用这种方法可以考量特征信息对整个系统的贡献, 贡献越大说明这个特征对分类决策的影响越深。

1.3 IG-SVM 模型

IG-SVM 模型流程如图 1 所示。

模型运行步骤如下:

步骤 1 基于本质安全理论从人、机器、管理、环境 4 个方面构建瓦斯爆炸事故风险评价指标体系。

步骤 2 通过实际调查煤矿监控系统以及问卷调查法获取大量现场数据(包括实时数据与非实时数据), 组成原始数据集 $\{D_1, D_2, \dots, D_m\}$ 。

步骤 3 将**步骤 2**所得的原始数据集进行信息增益处理, 可以获得各指标的增益值排序, 选择最优指标组成新的数据集 $\{I_1, I_2, \dots, I_m\}$ 。

步骤 4 用 mapminmax 函数对数据进行归一化处理, 将原始数据映射到 $[0, 1]$ 区间内, 这样可以避免量纲不同对分类结果的影响, 得到的新数据集为 $\{G_1, G_2, \dots, G_m\}$, 用新实验数据集训练 SVM 模型, 重复训练优化参数后可以得到分类预测模型 $f(I_i)$ 。

步骤 5 将未知风险的数据集导入模型训练,

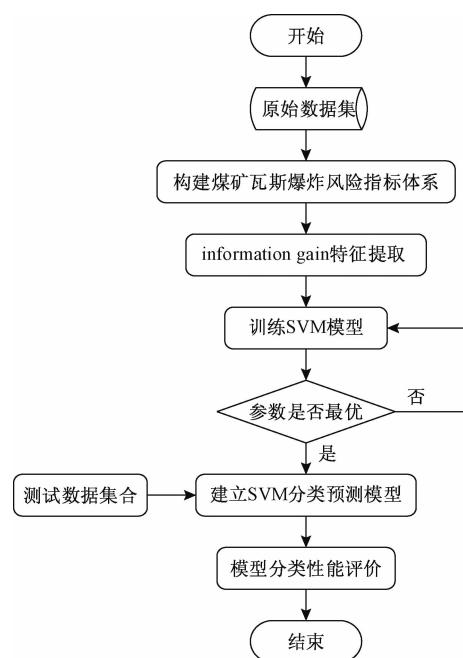


图 1 IG-SVM 模型流程示意图

Fig. 1 Flow diagram of IG-SVM model

获得预测结果。

2 实验分析

2.1 样本的采集

瓦斯爆炸事故是煤矿特有的极其严重的一种灾害, 利用事故发生与否在指标上的差异化表现可以对瓦斯爆炸风险进行预测, 对事故的预控预防具有指导意义。以甘肃、山西、内蒙古自治区等地区 100 家中小煤矿企业为研究对象, 收集他们的瓦斯爆炸事故数据, 定性指标以问卷调查法获取, 调查问卷向目标企业相关部门的管理人员发放。此次研究共发放问卷 170 份, 收回 141 份, 有效问卷为 110 份。对数据进行统计整理, 共获得 110 个样本点, 包括 20 个有风险样本(事故样本), 90 个无风险样本。在此基础上, 还需要对用于训练和测试的样本数据集进行划分, 本文选取训练和测试的样本数量比例为 8:2, 分别在有风险和无风险的样本中按比例随机抽取共 88 个作为预测模型的训练样本, 剩下 22 个作为检验模型预测效果的测试样本, 如表 1 所示。

表 1 实验样本分布情况

Table 1 Distribution of experimental samples

样本	有风险样本数	无风险样本数	样本总数
训练样本	16	72	88
测试样本	4	18	22
样本总数	20	90	110

2.2 风险评价指标

本质安全是一种以事故致因论为基础的科学全面的安全理论,完全符合国家安全生产的法律法规以及“十三五”发展规划,其核心理念是从人、机器、管理、环境四个要素的角度,做到各方面协调统一,消除不安全因素,建立本质安全型企业。本文将本质安全理念结合煤矿生产的实际情况,建立了一套较为全面的指标体系,如表2所示。为达到简化SVM模型提高预测正确率的目的,通过IG模型分析输入变量,提取对瓦斯爆炸风险分类结果有显著影响的指标,影响程度以信息增益值表示,信息增益的排序结果如表3所示,本文选择信息增益值大于0.015所对应的14个特征变量组成新数据集合。

2.3 参数设置及核函数的选取

本文建立的预测模型由Windows系统下的MATLAB19.0实现,采用MATLAB中的mapminmax对原始数据进行归一化,得到的新数据集与原数据集具有高度紧密性。SVM中核函数的选取对特征空间起着决定性的作用,目前还没有算法用于选取

表2 瓦斯爆炸风险指标体系

Table 2 Risk index system of gas explosion

分组归类	具体指标	单位
人的影响因素	受教育年限	a
	受教育水平	{1,2,3,4,5}
	年龄	a
	工龄	a
	农民工比例	%
	操作行为分数	[0,10]
机器的影响因素	三违率	%
	电器失爆等级	{0,1,2,3,4}
	设备老化程度	[0,10]
	通风系统可靠率	%
	排水系统可靠率	%
	监控系统可靠率	%
	运输设备可靠率	%
	消防设施完善率	%
管理的影响因素	安全培训参与率	%
	管理人员不在岗时间	min
	劳动纪律分数	[0,10]
	违章指挥分数	[0,10]
	火工品管理分数	[0,10]
	应急制度完善率	%
环境的影响因素	信息化程度	[0,10]
	火源情况	{0,1}
	瓦斯浓度	%
	摩擦撞击情况	{0,1}
	瓦斯涌出	m^3/min
	煤层倾角	(°)
	涌水量	m^3/h
	粉尘浓度	mg/m^3

最适合核函数的方法,一般采取试凑法,本文选取最常用的4种核函数分别是:Sigmoid核函数、LKF、PF、RBF,对分类结果进行对比,选取最优的核函数,其中LKF没有需要专门设置的参数,其余参数如表4所示。

瓦斯爆炸事故风险实际上是一个二元分类问题,即将煤矿开采区域分为两类有风险、无风险,采用混淆矩阵来评价分类模型的性能,如表5所示。

表3 特征指标信息增益排序表

Table 3 Feature index information gain ordering

指标	信息增益值	排序
受教育年限 x_1	0.008 7	24
受教育水平 x_2	0.013 9	17
年龄 x_3	0.015 6	13
工龄 x_4	0.006 6	27
农民工比例 x_5	0.009 0	23
操作行为分数 x_6	0.022 2	10
三违率 x_7	0.037 7	2
电器失爆等级 x_8	0.007 4	26
设备老化程度 x_9	0.028 0	4
通风系统可靠率 x_{10}	0.026 2	6
排水系统可靠率 x_{11}	0.012 1	18
监控系统可靠率 x_{12}	0.015 6	14
运输设备可靠率 x_{13}	0.006 3	28
消防设施完善率 x_{14}	0.025 0	8
安全培训参与率 x_{15}	0.025 4	7
管理人员不在岗时间 x_{16}	0.021 1	11
劳动纪律分数 x_{17}	0.007 7	25
违章指挥分数 x_{18}	0.009 6	21
火工品管理分数 x_{19}	0.026 8	5
应急制度完善率 x_{20}	0.013 9	16
信息化程度 x_{21}	0.011 8	19
火源情况 x_{22}	0.014 0	15
瓦斯浓度 x_{23}	0.030 4	3
摩擦撞击情况 x_{24}	0.020 7	12
瓦斯涌出 x_{25}	0.009 1	22
煤层倾角 x_{26}	0.024 4	9
涌水量 x_{27}	0.010 3	20
粉尘浓度 x_{28}	0.037 9	1

表4 核函数参数情况

Table 4 Parameters of kernel function

参数	SVM		
核函数	Sigmoid	PF	RBF
消耗	1 024	265	0.25
Gamma	4	0.25	0.062 5

表5 混淆矩阵

Table 5 Confusion matrix

真实	预测	
	true positive(TP)	true negative(TN)
false positive(FP)		
false negative(FN)		

P/N (positive/negative) 表示预测对象的所属类别, T/F (true/false) 表示分类的正确与否, 例如 TP 即将实际有风险的煤矿预测为有风险的样本数量, 利用混淆矩阵可以计算出一些评价模型性能的指标如: 正确率 $Acc = (TP + TN) / (TP + TN + FP + FN)$ 、准确率 $Precision = TP / (TP + FP)$ 、召回率 $Recall = TP / (TP + FN)$ 等^[11]。以上指标虽然能够一定程度上的评价模型性能, 但是应用于本文的分类器模型评价中大致有两方面的问题: 一是煤矿瓦斯事故属于低概率事件, 那么在收集数据时极大可能发生正负样本数量悬殊, 例如事故样本数量占总量 1% 时, 分类器只需将所有样本判定为无风险, 则准确率也可达到 99%, 这显然是不合理的。其次风险的发生虽然是一个二元分类问题, 但是所选对象归属类别的程度不同, 简单地用 0 或 1 概括会导致一些信息丢失。

ROC (receiver operating characteristic) 曲线也称受试者工作特性曲线, 其横纵坐标分别是假阳性率 (false positive rate, FPR) 和真阳性率 (true positive rate, TPR), 坐标 (0, 1) 对应的是理想的分类模型, 用曲线下的面积 (area under curve, AUC) 来表示分类器的性能, AUC 的大小一般在 0.5 ~ 1, 越靠近 1 表示越接近理想分类模型。选择 ROC 曲线不仅能够解决前文提出的两方面问题, 而且能直观地反映不同核函数模型分类情况的优劣。

将数据代入不同核函数的模型, ROC 曲线如图 2 所示, ROC 曲线越靠近左上角表示分类效果越好, 为了更加准确地描述可以计算曲线下面积, 即 AUC 值, 如表 6 所示。

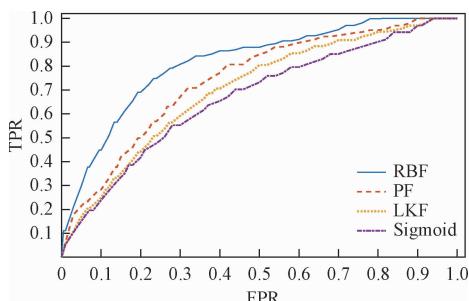


图 2 模型输出的 ROC 曲线

Fig. 2 ROC curve of model output

表 6 不同核函数 AUC 比较

Table 6 Comparison of AUC values of different kernel functions

核函数	RBF	Sigmoid	LKF	PF
AUC	0.880	0.711	0.746	0.783

由表 6 可知使用 RBF 核函数构造的模型输出的 AUC 最高, 而使用 Sigmoid 核函数、PF、LKF 的构造的模型输出的 AUC 都在 0.7 ~ 0.8, 效果不甚理想且相互之间差别不大。LKF 作为 RBF 的一种特殊形式, 在实验过程中发现两种核函数差距较大, 推测可能是由于煤矿瓦斯爆炸风险并不是线性可分的。Sigmoid 核函数可能在某些参数下是无效的, 因此给参数选取带来困难。PF 的支持向量较多分布于边界处, 可能对分类预测产生干扰。此外, RBF 有很强的映射复杂非线性关系的能力, 且学习规则简单、调试参数迅速, 鲁棒性强。综上所述, 本文选用 RBF 作为核函数建立预测模型。

2.4 结果与讨论

为了体现本文选用的 IG-SVM 模型的效能, 将该模型的预测结果与一些常用的分类器在同样的样本条件下进行比较, 为了简化比较结果, 仅对预测结果作二元分类 (+1 表示有风险, -1 表示无风险), 结果以样本的分类正确率表示, 如表 7 所示。

表 7 各模型分类预测结果对比

Table 7 Comparison of prediction results of different models

模型	训练样本	测试样本	有风险样本	无风险样本
SVM	0.852 3	0.863 6	0.900 0	0.911 1
KPCA-SVM	0.875 0	0.909 1	0.900 0	0.922 2
IG-SVM	0.977 3	0.954 5	0.950 0	0.988 9
SGD	0.863 6	0.681 8	0.650 0	0.811 1
KPCA-SGD	0.886 4	0.727 3	0.800 0	0.833 3
IG-SGD	0.909 1	0.772 7	0.850 0	0.866 7
DT	0.875 0	0.727 3	0.700 0	0.822 2
KPCA-DT	0.897 7	0.818 2	0.750 0	0.844 4
IG-DT	0.920 5	0.818 2	0.850 0	0.877 8
BP	0.670 5	0.545 5	0.550 0	0.700 0
KPCA-BP	0.715 9	0.590 9	0.600 0	0.733 3
IG-BP	0.750 0	0.636 4	0.650 0	0.777 8

实验结果表明: 在 4 种单一分类模型中, SVM 对训练样本的分类正确率为 85.23%, 略低于决策树 (decision tree, DT) 模型和随机梯度下降 (stochastic gradient descent, SGD) 模型, 而对测试样本的分类正确率最高, 达到了 86.36%, 而本文建模分类的目的是预测瓦斯爆炸风险, 显然测试样本的结果更加重要, 在有风险样本和无风险样本中, SVM 的表现要优于 3 种其他模型。比较使用了 IG 优化后的各种模型, 4 种分类正确率均有所提高, 例如 IG-SVM 模型相对于 SVM 模型 4 种正确率分别提高了 12.5%、9.09%、5%、7.78%, 说明了经过 IG 特征提取后的模型分类性能有所提高。此外, 实验还使用了核主成分分析法 (kernel principal component analysis, KPCA) 进行特征选取优化作为对比, 可以看出, 经过 KPCA 优化后的模型, 虽然整体水平也有所

提升,但普遍低于IG优化后的模型,充分说明了IG在特征选取优化方面的表现十分优异。

3 结论

依据本质安全理念建立了指标集,研究了IG-SVM在处理煤矿瓦斯爆炸风险预测上的表现,具体结论如下:

(1) 针对煤矿瓦斯爆炸风险问题,使用SVM进行分类预测取得了不错的效果,但是过于庞大的指标体系在一定程度上形成了特征信息的冗余,从而影响计算机仿真的速度与正确率。因此,优化评价指标体系是十分必要的。

(2) IG通过对信息增益值进行排序精简指标体系,删除了不必要的冗余信息,降低了模型整体的运算维度,从而优化了模型的速度与正确率,并通过实验证明IG与SVM的组合在解决本文问题时要优于其他模型,最终测试样本的正确率达到了95.45%,在防控灾害上有很好的应用价值。

参 考 文 献

- [1] 王双明. 对中国煤炭主体能源地位与绿色开采的思考[J]. 中国煤炭, 2020, 46(2): 11-16.
Wang Shuangming. Thinking on the main energy status and green mining of coal in China[J]. China Coal, 2020, 46(2): 11-16.
- [2] 杜恩蒙. 煤矿在用设备安全检测检验存在的问题及对策[J]. 内蒙古煤炭经济, 2019, 273(4): 69-70.
Du Enmeng. Problems and countermeasures in safety inspection of coal mine in use equipment[J]. Inner Mongolia Coal Economy, 2019, 273(4): 69-70.
- [3] 田水承, 李 华, 陈勇刚. 基于神经网络的掘进面瓦斯爆炸危险源安全评价[J]. 煤田地质与勘探, 2005(3): 19-21.
Tian Shuicheng, Li Hua, Chen Yonggang. Safety assessment of gas explosion hazard source in heading face based on neural network [J]. Coalfield Geology and Exploration, 2005(3): 19-21.
- [4] 李润求, 施式亮, 念其锋, 等. 基于IAHP-ECM的瓦斯爆炸灾害风险评估[J]. 中国安全科学学报, 2013, 23(3): 62-67.
Li Runqiu, Shi Shiliang, Nian Qifeng, et al. Risk assessment of gas explosion disaster based on IAHP-ECM[J]. Chinese Journal of safety Sciences, 2013, 23(3): 62-67.
- [5] 邵良彬, 詹小凡. 煤与瓦斯突出missForest-EGWO-SVM预测模型[J]. 辽宁工程技术大学学报(自然科学版), 2020, 39(3): 214-218.
Shao Liangshan, Zhan Xiaofan. MissForest-EGWO-SVM prediction model for coal and gas outburst[J]. Journal of Liaoning University of Technology (Natural Science Edition), 2020, 39(3): 214-218.
- [6] 李 爽, 李丁炜, 犹梦洁, 等. 基于BN-ELM的煤矿瓦斯安全态势预测方法[J]. 系统工程, 2020, 38(3): 132-140.
Li Shuang, Li Dingwei, You Mengjie, et al. Prediction method of coal mine gas safety situation based on BN-ELM[J]. Systems Engineering, 2020, 38(3): 132-140.
- [7] Wang Z Y. Research on coal mine gas prediction algorithm based on improved SVM[J]. Agro Food Industry Hi-Tech, 2017, 28(1): 1729-1733.
- [8] Vapnik V N. The nature of statistical learning theory[M]. Berlin: Springer Press, 1995.
- [9] Tan X, Yu F, Zhao X. Support vector machine algorithm for artificial intelligence optimization[J]. Cluster Computing, 2018, 22(6): 15015-15021.
- [10] 刘庆和, 梁正友. 一种基于信息增益的特征优化选择方法[J]. 计算机工程与应用, 2011, 47(12): 130-132, 136.
Liu Qinghe, Liang Zhengyou. A feature selection method based on information gain[J]. Computer Engineering and Applications, 2011, 47(12): 130-132, 136.
- [11] Stehman S V. Selecting and interpreting measures of thematic classification accuracy[J]. Remote Sensing of Environment, 1997, 62(1): 77-89.