

基于可见光-近红外光谱的煤岩识别方法实验研究

徐良骥^{1,2}, 孟雪莹², 韦任², 张坤²

1. 深部煤矿采动响应与灾害防控国家重点实验室, 安徽淮南 232001
2. 安徽理工大学空间信息与测绘工程学院, 安徽淮南 232001

摘要 以淮南矿区谢桥矿和潘二矿的煤和岩石样本为研究对象, 通过地物光谱仪采集样本反射率光谱曲线, 同时检测样本氧化物含量、水分、灰分及挥发分含量, 将样本的反射率光谱曲线和样本成分含量分别作为自变量, 样本类别“煤”和“岩石”两种矿物类型作为因变量, 建立煤和岩石识别模型对煤和岩石进行二分类。该研究主要采用三种模型, 分别为主成分分析结合支持向量机(PCA-SVM)、主成分分析结合BP神经网络(PCA-BP)模型和核主成分分析结合支持向量机(KPCA-SVM)模型。结果表明, 基于可见光近红外光谱的三个模型中, 核主成分分析结合支持向量机模型的识别精度最高, 建模平均精度为95.5%, 验证平均精度约为90.56%; 基于样本成分的三个模型中, 核主成分分析结合支持向量机模型的识别精度最高, 建模平均精度为98.5%, 验证平均精度约为95%。

关键词 可见光-近红外光谱; 主成分分析; 核主成分分析; 支持向量机; BP神经网络

中图分类号: TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)07-2135-08

引言

煤岩识别是实现智能化选煤和综采工作面无人化的前提。近年来随着计算机技术的迅速发展, 煤矿生产自动化水平显著提高, 也使得综合开采工作面的无人化操作和选煤厂智能化选煤的实现成为可能^[1]。智能化技术的实现在很大程度上提高煤炭生产的安全性^[2]。煤岩识别技术的实现对煤炭绿色开采、精准开采具有重大的实际意义^[3]。

近年来, 遥感技术在矿区煤炭勘探开采、煤和岩石的性质测定、土壤和水体的污染监测等方面得到广泛应用^[4,5]。在煤炭资源勘探开发各阶段遥感技术主要应用于资源调查评价、矿区卫星测图、矿产资源开发及矿区环境监测等领域^[6], 在煤炭开采过程中, 可以利用遥感图像对矿区地层分布特征进行提取。对于土壤与水体污染监测方面, 遥感技术更是可以直接利用地物光谱反射率进行地物提取与监测。近年来, 井下无人开采成为国内外研究的热点, 而煤岩识别正是井下无人开采的关键, 在传统煤岩识别过程中, 人工识别效率低下, 因此如何准确的实现煤岩识别成为该研究领域亟待解决的问题。宋亮等^[7]研究了基于可见光-近红外和热红外联合分析煤和岩石分类方法; 王卫东等^[8]研究了基于激光三维扫描与动态称重的煤岩石光电分选系统; Yang等^[9]利

用可见光和近红外(VIS-NIR)反射光谱进行煤岩石识别; Wang等^[10]使用太赫兹时域光谱法对煤和岩石进行表征和分类。本文基于可见光-近红外光谱和样本成分含量研究煤岩的识别方法。

1 实验部分

1.1 研究区概况

实验样本共29份, 其中煤样15份, 岩石样本14份, 样本来源如下:

从谢桥矿采集煤和岩石样本共11份, 其中煤样7份, 岩石样本4份, 7份煤样按顺序命名为XQM1, ..., XQM7, 岩石样本按顺序命名为XQG1, ..., XQG4。煤的类型主要为烟煤, 岩石类型主要为砂岩、和泥岩。

从潘二矿采集煤和岩石样本共18份, 其中煤样8份, 岩石样本10份, 8份煤样按顺序命名为P2M1, ..., P2M8, 10份岩石样本按顺序命名为P2G1, ..., P2G10。

谢桥矿位于安徽省颍上县东北部, 距颍上县城约20 km, 现生产能力超过800万吨/年、配套800万吨选煤厂的特大型现代化矿井。潘二煤是淮南矿业集团所属的一座年设计生产能力为300万吨的大型现代化矿井。该矿位于安徽省淮南市西北部约30 km的潘集区境内。矿井地处淮河以北属

收稿日期: 2021-01-11, 修订日期: 2021-11-05

基金项目: 国家自然科学基金项目(41472323, 41372369), 安徽省重点研发计划项目(201904b11020015)资助

作者简介: 徐良骥, 1978年生, 安徽理工大学空间信息与测绘工程学院教授 e-mail: ljxu@aust.edu.cn

江淮平原,西南部与潘一矿井接壤,西北部与潘北矿井毗邻。



图 1 部分煤和岩石样本

Fig. 1 Some coal and rock samples

1.2 煤岩光谱差异机理分析

图 2 所示为煤和岩石样本的反射光谱,由图可知,煤的整体反射率较低,上升平缓,而岩石的反射率偏高,且有吸收谷。在 1 450 nm 附近,由于水分子 O—H 官能基伸缩振动的第一倍频,岩石在此处有较强吸收谷。在 1 900 nm 附近,由于岩石中的二价 Fe 离子和煤样中的 Al_2O_3 , 岩石存在更强的吸收谷。在 2 130~2 250 nm 波段内煤与岩石存在较大差异,这是由于 Al 元素在煤中主要以 Al_2O_3 的形式存在,而在岩石中则主要以 $\text{Al}(\text{OH})_3$ 形式存在, $\text{Al}(\text{OH})_3$ 的 Al—OH 晶格振动使得其在 2 210 nm 附近具有强吸收峰^[11]。

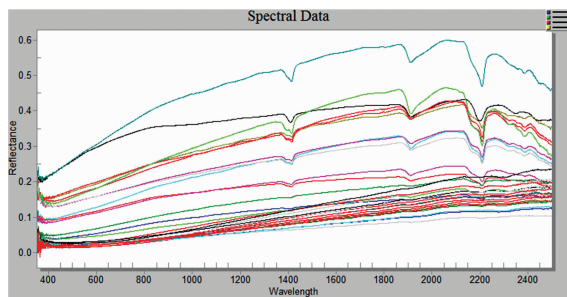


图 2 原始光谱反射率曲线

Fig. 2 Original reflectance spectra

近红外光谱的主要吸收带是含氢基团 C—H, O—H, N—H 等的一级倍频和 C—O, C—N, C—C 等的多级倍频。煤中的有机物主要包括碳(C)、氢(H)、氧(O)、氮(N)等,主要指工业分析指标的挥发分和固定碳;无机物包括水和碳物质,主要指工业指标的水分和灰分。根据中国国家标准

GB/T 5751《中国煤炭分类》和国际标准 ISO 11760 Classification of coals 对煤炭的定义标准,煤炭是由植物遗体经煤化作用转化而成的富含碳的固体可燃有机沉积岩,含有一定量的矿物质,其灰分产率小于或等于 50%^[12]。据此,灰分和挥发分的含量差异可用于进行煤岩识别分析。

1.3 光谱数据的采集

美国 ASD 公司生产的地物光谱仪 FieldSpec 4 的光谱范围是 350~2 500 nm,有两种采样间隔分别为 1.4 nm(350~1 000 nm)和 2 nm(1 000~2 500 nm);该实验的重采样间隔为 1 nm。数据采集在暗室中进行,选择 50 W 的卤素灯为光源、25°裸光纤镜头接收反射波段。测量时将样本放入直径为 100 mm,高 2 mm 的透明玻璃培养皿中,光源距离样品 40 cm,光线与样品成 45°角,探头距样本 10 cm 位于光源对面,探头光纤末端位于煤、岩石样本正上方。在对样本进行光谱测量前需要进行白板校正,每个样本采集 30 条曲线,对获取的煤和岩石样本反射光谱曲线采用 ViewSpecPro 软件进行预处理(剔除异常数据、断点修复,光滑处理等),最后将各组曲线的算术平均值作为样本的原始光谱反射率值,如图 2 所示。

1.4 样品成分含量检测

采用 GB212—2008 煤的工业分析方法对煤和岩石样本的水分、灰分、挥发分进行测定。测定结果如表 1 所示,具体测定步骤如下:

(1) 样本水分的质量分数(Mad)采用空气干燥法^[13]。

称取粒度粒径为 0.15 mm 煤样、岩石样本(1±0.1) g,称准至 0.000 2 g,平摊在称量瓶中。每个样本设置三个对比样,将样品置于 105~110 °C 鼓风干燥箱内,于气流中干燥到质量恒定。根据煤样的质量损失计算出水分的质量分数(Mad)。

表 1 两矿煤样工业指标测试结果

Table 1 Test results of industrial indicators of coal samples from two mines

样本编号	Mad/%	Vad/%	Aad/%
XQM1	1.60	28.16	33.74
XQM2	1.71	27.16	29.60
XQM3	1.00	28.37	33.66
XQM4	1.50	28.07	27.08
XQM5	1.10	27.63	31.38
XQM6	1.20	28.57	31.96
XQM7	1.20	27.50	31.51
XQG1	1.40	10.84	85.05
XQG2	1.00	10.66	86.11
XQG3	1.40	12.39	86.31
XQG4	1.20	10.83	86.44
P2M1	0.99	23.67	40.80
P2M2	0.85	33.15	5.86
P2M3	0.79	32.09	6.61
P2M4	1.06	33.74	7.05
P2M5	0.92	34.25	11.24
P2M6	1.32	26.85	22.71
P2M7	0.90	29.11	17.00

续表 1

P2M8	1.38	25.59	32.60
P2G1	0.40	2.71	97.00
P2G2	0.80	9.26	87.10
P2G3	1.80	9.64	88.46
P2G4	0.70	7.27	85.43
P2G5	1.70	8.96	89.78
P2G6	2.10	12.31	82.57
P2G7	0.80	10.75	84.12
P2G8	0.70	10.97	88.08
P2G9	1.03	6.93	92.19
P2G10	1.11	9.20	88.31

(2) 样本的灰分(Aad)采用快速灰化法^[13]。

①在预先灼烧至质量恒定的灰皿中，称取粒度小于 0.2 mm 的一般分析实验煤样(1±0.1) g，称准至 0.000 2 g，均匀地摊平在灰皿中，使其每平方厘米的质量不超过 0.15 g。将盛有煤样的灰皿预先分排放在耐热瓷板或石棉板上。

②将马弗炉加热到 850 ℃，打开炉门，将放有灰皿的耐热瓷板缓慢地推入马弗炉中，待 5~10 min 后煤样不再冒烟时，以每分钟不大于 2 cm 的速度把其余各排灰皿顺序推入炉内炽热部分(若煤样着火爆燃，停止该实验、作废)。

③关上炉门并使炉门留有 15 mm 左右的缝隙，在(815±10) ℃温度下灼烧 40 min。

④将灰皿从炉中取出，放在空气中冷却 5 min，然后移入干燥器中冷却至室温(约 20 min)后，称量。

⑤进行检查性灼烧，温度为(815±10) ℃，每次 20 min，直到连续两次灼烧后的质量变化不超过 0.001 0 g 为止。以最后一次灼烧后的质量为计算依据。

(3) 样本的挥发分(Vad)测定^[13]：

①在预先于 900 ℃温度下灼烧至质量恒定的带盖瓷坩锅中，称取粒度小于 0.2 mm 的一般分析实验煤样(1±0.1) g，称准至 0.000 2 g，然后轻轻振动坩锅，使样品摊平，盖上盖，放在坩锅架上。

②将马弗炉预先加热至 920 ℃左右，打开炉门，迅速将



图 3 煤和岩石样本灰分实验图

Fig. 3 Experimental chart of ash content of coal and rock samples



图 4 煤和岩石样本挥发分实验炉

Fig. 4 Set up for determining the volatile matter in coal and gangue

放有坩锅的坩锅架送入恒温区，立即关上炉门并计时，准确加热 7 min；坩锅及坩锅架放入后，要求炉温在 3 min 内恢复至(900±10) ℃，此后保持在(900±10) ℃，否则此次实验作废，加热时间包括温度恢复时间。

③从炉中取出坩锅，放在空气中冷却 5 min 左右，然后移入干燥器中冷却至室温(约 20 min)后称量。

煤和岩石氧化物含量采用 XRF 检测，检测结果如表 2 所示。

表 2 两矿煤和岩石样本氧化物百分比含量

Table 2 Oxygen percentage contents of coal and rock samples from two mines

样品编号	Na ₂ O	MgO	Al ₂ O ₃	SiO ₂	P ₂ O ₅	SO ₃	K ₂ O	CaO	TiO ₂	Fe ₂ O ₃
XQM1	0.37	0.51	30.00	57.10	0.24	2.37	0.00	1.93	1.53	5.97
XQM2	0.34	0.53	27.93	57.74	0.16	4.29	1.13	0.19	2.08	5.62
XQM3	0.22	0.52	28.71	58.42	0.21	2.97	0.13	1.19	1.86	5.76
XQM4	0.36	0.55	27.16	57.34	0.18	4.66	1.16	0.67	2.11	5.82
XQM5	0.33	0.52	28.68	57.92	0.16	3.95	0.82	0.09	1.97	5.56
XQM6	0.39	0.51	28.66	57.08	0.17	3.36	0.50	1.66	1.93	5.76
XQM7	0.30	0.51	28.37	57.88	0.17	3.77	1.05	0.50	1.98	5.48
XQG1	0.17	0.33	33.15	60.68	0.10	0.16	0.44	0.00	1.44	3.53
XQG2	0.36	0.31	32.37	61.22	0.09	0.17	0.87	0.00	1.20	3.42
XQG3	0.44	0.31	35.67	58.63	0.10	0.09	0.16	0.00	1.08	3.53
XQG4	0.28	0.32	32.58	60.48	0.10	0.16	1.08	0.00	1.46	3.55
P2M1	0.33	0.43	28.04	60.24	0.18	2.09	0.97	1.41	1.31	4.99

续表 2

P2M2	1.49	0.90	13.64	47.82	5.63	7.47	0.56	11.73	1.80	8.97
P2M3	1.53	0.99	17.76	57.14	1.15	9.36	0.00	0.66	2.80	8.60
P2M4	1.24	0.99	15.99	55.32	1.70	10.35	0.00	1.78	3.04	9.58
P2M5	0.65	2.57	9.71	38.28	0.23	5.71	0.00	33.71	0.85	8.30
P2M6	0.66	0.56	23.46	52.59	0.17	8.65	2.24	3.65	1.86	6.17
P2M7	0.90	0.62	24.23	51.89	0.19	6.48	0.98	6.71	1.91	6.10
P2M8	0.33	0.52	27.18	56.61	0.26	3.53	1.06	3.00	1.56	5.97
P2G1	0.05	0.32	19.99	68.38	0.10	0.13	6.01	0.77	1.11	3.15
P2G2	0.19	0.31	29.66	61.64	0.10	0.23	3.51	0.00	1.30	3.08
P2G3	0.20	0.31	30.13	62.91	0.10	0.12	1.55	0.00	1.26	3.42
P2G4	0.21	0.31	30.23	62.87	0.10	0.12	1.52	0.00	1.23	3.42
P2G5	0.31	0.30	28.81	62.82	0.09	0.24	2.78	0.07	1.25	3.34
P2G6	0.51	0.31	34.54	58.91	0.10	0.15	0.73	0.00	1.51	3.26
P2G7	0.36	0.30	29.80	61.57	0.09	0.54	2.91	0.00	1.26	3.18
P2G8	0.13	0.30	27.82	63.56	0.10	0.18	1.61	0.56	1.06	4.69
P2G9	0.37	0.29	29.40	61.34	0.09	0.10	4.28	0.00	1.00	3.15
P2G10	0.28	0.29	28.29	63.86	0.09	0.40	2.34	0.01	0.92	3.52

1.5 建模方法及参数选择分析

主成分分析是目前较常用的一种光谱特征信息提取方法,在处理线性问题时能够取得很好的效果。核成分分析则将数据映射到高维特征空间,利用主成分分析实现非线性的特征提取,改善主成分分析在非线性数据分布情况下分析结果不理想的状态。与其他机器学习算法相比,支持向量机(SVM)算法更适用于本实验的小训练样本、多维度成分因素、非线性关系问题^[14]。因此选择 PCA-SVM、PCA-BP 和 KPCA-SVM 共三种方法进行煤岩识别的建模。其中,惩罚因子由网格搜索法获得,经过多次建模试验确定最优迭代率和学习率。模型复杂度低,未出现过拟合现象。

2 结果与讨论

2.1 基于可见光-近红外光谱的煤和岩石识别模型

(1)主成分分析结合支持向量机(PCA-SVM)模型

煤岩识别的本质是一个二分类的问题,并不需要区分煤或岩石具体的矿物类型。将煤作为识别向量标签“0”,岩石作为识别向量标签“1”,对煤岩的特征值进行训练,将煤岩的特征值作为自变量,煤岩的识别标签“0”和“1”作为因变量,最后利用训练的模型对验证集进行识别。其中,煤岩对应的特征值分别为它们在可见光-近红外波段的反射率和工业分析与 XRF 仪器测量出的成分含量。随机建立训练集 20 个样本和验证集 9 个样本,对样本反射率进行主成分分析,确定主成分得分值利用 5 倍交叉验证确定最优惩罚参数和最优方差参数,进行支持向量机模型建立,并将验证集代入模型验证。随机建立的 20 组主成分分析结合支持向量机识别模型的最优参数、建模精度、验证结果如表 3 所示。

由表 3 可知,主成分分析结合支持向量机模型中,建模精度最高为 100%,最低精度为 55%,平均精度为 83.75%。验证集模型识别精度最高为 66.67%,最低为 22.22%,平均识别精度约为 45.97%。

表 3 基于可见光-近红外光谱的 PCA-SVM 模型的模型精度

Table 3 Model accuracy of PCA-SVM model based on visible-near infrared spectra

组别	最优惩罚参数 c	方差参数 g	建模精度 /%	验证精度 /%
1	0.378 93	6.964 4	85	44.44
2	0.5	1.148 7	75	75
3	1.071 8	36.758 3	100	66.67
4	1	2.143 5	75	55.56
5	0.000 976 56	0.000 976 56	65	22.22
6	1.624 5	0.378 93	75	55.56
7	13.928 8	4	90	33.33
8	0.933 03	12.125 7	90	22.22
9	1.231 1	27.857 6	95	66.67
10	19.698 3	1.741 1	95	66.67
11	675.588 1	0.435 28	95	66.67
12	0.000 976 56	0.000 976 56	55	33.33
13	1.148 7	315.173	100	33.33
14	27.857 6	2.639	90	44.44
15	0.000 976 56	0.000 976 56	55	44.44
16	0.000 976 56	36.758 3	85	44.44
17	3.249	0.203 06	65	22.22
18	1.319 5	48.502 9	90	44.44
19	1.231 1	84.448 5	95	44.44
20	73.516 7	6.062 9	100	33.33

(2)主成分分析结合 BP 神经网络(PCA-BP)模型

对样本光谱反射率数据进行主成分分析,得到 28 个主成分。首先对 BP 神经网络中的初始参数进行设置,构造神经网络结构,参数如下:迭代次数(epochs)=10 000,学习率(Ir)=0.05,训练目标误差(goal)=0.001。整个神经网络匹配系统分为四个部分:输入端、一个隐含层、一个输出层、输出端。神经网络结构如图 5 所示。

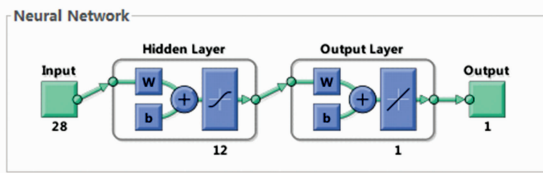


图 5 基于可见光-近红外光谱的 PCA-BP 神经网络模型结构

Fig. 5 Model structure of PCA-BP neural network based on visible-near infrared spectra

将主成分数据代入 BP 神经网络模型中进行煤和岩石识别分类，20 组数据的 PCA-BP 神经网络模型的预测精度如图 6 所示。

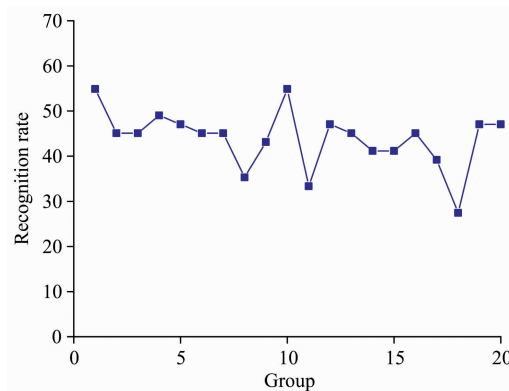


图 6 PCA-BP 神经网络模型识别率

Fig. 6 PCA-BP neural network model recognition rate

基于样本反射率光谱数据的 PCA-BP 神经网络模型识别率最高为 54.55%，最低为 26.25%，平均识别率为 43.33%。

(3)核主成分分析结合主成分分析(KPCA-SVM)模型

对样本反射率进行核主成分分析，确定主成分得分值利用 5 倍交叉验证确定最优惩罚参数和最优方差参数，进行支持向量机模型建立，并将验证集代入模型验证。20 组 KPCA-SVM 模型最优参数、训练集、验证精度如表 4 所示。

由表 4 可知，核主成分分析结合支持向量机模型中，建模精度最高为 100%，最低精度为 85%，平均精度为 95.5%。验证集模型识别精度最高为 100%，最低为 66.67%，平均识别精度约为 90.56%。

2.2 基于样本成分含量的煤和岩石识别模型

(1)主成分分析结合支持向量机(PCA-SVM)模型

对样本的 14 个成分含量进行主成分分析，确定主成分得分值利用 5 倍交叉验证确定最优惩罚参数和最优方差参数，进行支持向量机模型建立，并将验证集带入模型验证。基于样本成分的 PCA-SVM 模型最优参数、训练集、验证精度如表 5 所示。

由表 5 可知，主成分分析结合支持向量机模型中，建模精度最高为 100%，最低精度为 70%，平均精度为 83.75%。验证集模型识别精度最高为 100%，最低为 22.22%，平均识别精度约为 50.56%。

表 4 KPCA-SVM 模型的模型精度

Table 4 Accuracy of KPCA-SVM model

组别	最优惩罚参数 c	方差参数 g	建模精度 /%	验证精度 /%
1	0.319 5	2.639	100	100
2	0.176 78	0.659 75	90	88.89
3	0.267 94	0.5	90	100
4	0.574 35	0.435 28	100	77.78
5	0.435 28	0.353 55	95	77.78
6	0.574 35	0.757 86	95	88.89
7	0.203 06	0.659 75	95	77.78
8	34.296 8	4	100	100
9	0.233 26	0.757 86	85	100
10	0.000 976 56	2.462 3	100	100
11	2.828 4	1.414 2	100	100
12	0.267 94	3.249	95	88.89
13	0.287 17	1.148 7	100	66.67
14	0.233 26	0.812 25	95	88.89
15	0.25	0.615 57	95	77.78
16	3.482 2	0.574 35	90	100
17	0.307 79	3.031 4	95	88.89
18	548.748	1.866 1	100	100
19	0.5	2.639	95	88.89
20	0.203 06	0.574 35	95	100

表 5 基于样本成分的 PCA-SVM 模型的模型精度

Table 5 Accuracy of principal component analysis combined with support vector machine model

组别	最优惩罚参数 c	方差参数 g	建模精度 /%	验证精度 /%
1	0.574 35	5.278	75	44.44
2	10.556 1	0.217 64	70	66.67
3	0.757 86	111.430 5	90	44.44
4	6.062 9	6.964 4	85	33.33
5	256	1.515 7	95	22.22
6	13.928 8	0.066 986	80	66.67
7	0.757 86	207.936 6	100	33.33
8	1.231 1	630.345 9	100	22.22
9	137.187	0.094 732	70	66.67
10	0.000 976 56	2.639	70	66.67
11	6.062 9	0.066 986	75	66.67
12	0.267 94	0.870 55	100	100
13	0.757 86	34.296 8	95	33.33
14	6.062 9	2.828 4	90	66.67
15	0.707 11	64	90	44.44
16	0.870 55	55.715 2	85	44.44
17	25.992 1	1.414 2	80	66.67
18	0.615 57	51.984 2	80	44.44
19	0.870 55	2.462 3	75	44.44
20	0.615 57	1.741 1	70	33.33

(2)主成分分析结合 BP 神经网络(PCA-BP)模型

对样本光谱反射率数据进行主成分分析，得到 12 个主成分。首先对 BP 神经网络中的初始参数进行设置，构造神经网络结构，参数如下：迭代次数(epochs)=10 000，学习率

(Ir)=0.05, 训练目标误差(goal)=0.001。整个神经网络匹配系统分为四个部分:输入端、一个隐含层、一个输出层、输出端。神经网络结构如图 7 所示。

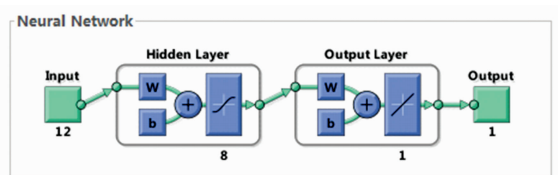


图 7 基于样本成分的 PCA-BP 神经网络结构

Fig. 7 Structure of PCA-BP neural network based on sample components

将主成分数据代入 BP 神经网络模型中进行煤和岩石识别分类, 20 组数据的 PCA-BP 神经网络模型的预测精度如图 8 所示。

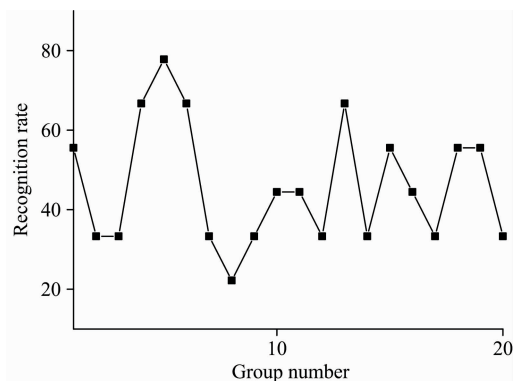


图 8 基于样本成分的 PCA-BP 神经网络模型识别率

Fig. 8 Recognition rate of PCA-BP neural network model based on sample components

基于样本成分的 PCA-BP 神经网络模型识别率最高为 77.78%, 最低为 22.22%, 平均识别率为 46.11%。

(3)核主成分分析结合主成分分析(KPCA-SVM)模型

对样本成分进行核主成分分析, 确定主成分得分值利用 5 倍交叉验证确定最优惩罚参数和最优方差参数, 进行支持向量机模型建立, 并将验证集带入模型验证。20 组 KPCA-SVM 模型最优参数、训练集、验证精度如表 6 所示。

由表 6 可知, 核主成分分析结合支持向量机模型中, 建模精度最高为 100%, 最低精度为 95%, 平均精度为 98.5%。验证集模型识别精度最高为 100%, 最低为 77.78%, 平均识别精度约为 95%。

3 结论

针对可见光-近红外光谱法煤岩石识别时易发生煤与岩石误分类的问题, 综合采用主成分分析结合支持向量机模型(PCA-SVM)、主成分分析结合 BP 神经网络模型、核主成分分析结合支持向量机模型(KPCA-SVM)进行了研究, 并将

最终的识别率、精度列在表 7 中。

表 6 基于样本成分的 KPCA-SVM 模型的模型精度

Table 6 Model accuracy of KPCA-SVM model based on sample components

组别	最优惩罚参数 c	方差参数 g	建模精度 / %	验证精度 / %
1	0.203 06	0.659 75	100	88.89
2	0.329 88	0.870 55	100	100
3	0.353 55	0.757 86	100	100
4	0.233 26	0.406 13	100	88.89
5	0.435 28	0.406 13	95	100
6	0.329 88	0.435 28	100	88.89
7	0.287 17	0.535 89	95	100
8	0.757 86	0.329 88	100	88.89
9	24.251 5	0.016 746	100	100
10	0.000 976 56	0.000 976 56	95	100
11	0.000 976 56	0.000 976 56	95	100
12	0.267 94	0.870 55	100	100
13	0.329 88	0.378 93	100	88.89
14	0.707 11	0.535 89	100	100
15	12.125 7	0.033 493	100	100
16	0.000 976 56	0.000 976 56	100	88.89
17	0.267 94	0.435 28	100	88.89
18	0.267 94	0.466 52	95	100
19	0.233 26	0.659 75	95	100
20	0.267 94	1	100	77.78

表 7 算法模型精度/识别率比较

Table 7 Algorithm model accuracy/recognition rate comparison

特征提取	算法模型	建模精度 / %	验证精度/识别率 / %
可见光-近红外光谱	PCA-SVM	83.75	66.67
	PCA-BP	/	46.11
	KPCA-SVM	95.5	90.56
样本成分含量	PCA-SVM	83.75	50.56
	PCA-BP	/	46.11
	KPCA-SVM	98.5	95

(1)基于可见光-近红外光谱的 KPCA-SVM 模型建模精度最高为 100%, 最低精度为 85%, 验证模型精度最高为 100%。基于样本成分含量的 KPCA-SVM 模型建模精度最高为 100%, 最低精度为 95%。验证模型精度最高为 100%。可见, 基于可见光-近红外光谱和煤岩成分含量进行煤岩识别是可行的, 在建模方法中 KPCA-SVM 表现较佳, 优于 PCA-SVM 和 PCA-BP 两种方法。

(2)基于样本成分含量对煤、岩石的分类过程中, KPCA-SVM 的建模精度高于 PCA-SVM, 平均建模精度为 98.5%; 同时 KPCA-SVM 的验证精度高于 PCA-SVM 和 PCA-BP 两种方法的验证精度, 平均精度为 95%。

(3) 提取了煤岩的光谱数据和成分含量两种特征, 结合 PCA-SVM, PCA-BP 和 KPCA-SVM 三种算法模型, 建立了六种煤岩识别的方法。由表 7 可知, 基于煤岩成分含量的

KPCA-SVM 的模型方法识别率达到了 95%, 高于其他五种方法的识别率, 故该模型最优。

References

- [1] QIAN Ming-gao, XU Jia-lin, WANG Jia-chen(钱鸣高, 许家林, 王家臣). Journal of China Coal Society(煤炭学报), 2018, 43(1): 1.
- [2] DONG Chang-shuang, YAO Ping-xi, LIU Zhi-he(董长双, 姚平喜, 刘志河). Coal Science and Technology(煤炭科学技术), 2007, 35(3): 54.
- [3] YUAN Liang(袁 亮). Journal of China Coal Society(煤炭学报), 2017, 42(1): 1.
- [4] Le B T, Xiao D, Okello D, et al. Spectroscopy Letters, 2017, 50(8): 440.
- [5] Shen Qiang, Xia Ke, Zhang Shiwen, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 222: 1.
- [6] XU Liang-ji, LIU Shu-guang, MENG Xue-ying, et al(徐良骥, 刘曙光, 孟雪莹, 等). Journal of China Coal Society(煤炭学报), 2019, 44(11): 3539.
- [7] SONG Liang, LIU Shan-jun, YU Mo-li, et al(宋 亮, 刘善军, 虞茉莉, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(2): 416.
- [8] WANG Wei-dong, ZHANG Chen, MA Zhong-liang, et al(王卫东, 张 晨, 马中良, 等). Industry and Mine Automation(工矿自动化), 2013, 39(12): 5.
- [9] Yang En, Ge Shirong, Wang Shibo. Journal of Spectroscopy, 2018, 2018(2): 1.
- [10] Wang X, Hu K, Zhang L, et al. Journal of Infrared Millimeter and Terahertz Waves, 2017, 38(2): 1.
- [11] YANG En, WANG Shi-bo, GE Shi-rong, et al(杨 恩, 王世博, 葛世荣, 等). Journal of China Coal Society(煤炭学报), 2018, 43(S2): 646.
- [12] YANG En, WANG Shi-bo, GE Shi-rong, et al(杨 恩, 王世博, 葛世荣, 等). Journal of China Coal Society(煤炭学报), 2019, 44(12): 3912.
- [13] General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China(中华人民共和国国家质量监督检验检疫总局). GB/T 212—2008 Proximate Analysis of Coal(GB/T 212—2008 煤的工业分析法), 2008.
- [14] LIANG Shun-lin, CHENG Jie, JIA Kun, et al(梁顺林, 程 洁, 贾 坤, 等). National Remote Sensing Bulletin(遥感学报), 2016, 20(5): 875.

Experimental Research on Coal-Rock Identification Method Based on Visible-Near Infrared Spectroscopy

XU Liang-ji^{1, 2}, MENG Xue-ying², WEI Ren², ZHANG Kun²

1. National Key Experiment of Mining Response and Disaster Prevention and Control in Deep Coal Mine, Huainan 232001, China
2. School of Spatial Information and Geomatics Engineering, Anhui University of Science and Technology, Huainan 232001, China

Abstract Taking the coal and rock samples retrieved from the Huainan Xieqiao Mine and the Paner II Mine as the research object, the sample reflectance spectrum curve was collected by a ground spectrometer, and the sample's oxide content, moisture, ash and volatile content were simultaneously detected to reflect the sample's reflection. The rate spectral curve and the sample component content are used as independent variables, and the sample type is used as the dependent variable to establish a coal and rock identification model to classify coal and rock. This paper mainly adopts three models, which are principal component analysis combined with support vector machine (PCA-SVM), principal component analysis combined with BP neural network (PCA-BP) model and kernel principal component analysis combined with support vector machine (KPCA-SVM) model. The results show that among the three models based on visible light near-infrared spectroscopy, nuclear principal component analysis combined with support vector machine model has the highest recognition accuracy, the average accuracy of modeling is 95.5%, and the average accuracy of verification is about 90.56%; three based on sample components. In the model, the kernel principal component analysis combined with the support vector machine model has the highest recognition accuracy, the average accuracy of modeling is 98.5%, and the average accuracy of verification is about 95%.

Keywords Visible-near infrared spectroscopy; Principal component analysis; Nuclear principal component analysis; Support vector machine; BP neural network

(Received Jan. 11, 2021; accepted Nov. 5, 2021)