第 39 卷 第 4 期	中国岩溶	Vol. 39 No. 4
2020年8月	CARSOLOGICA SINICA	Aug. 2020

周爱红,牛鹏飞,袁颖,等.基于PCA-PSO-SVM的凡口铅锌矿地区岩溶地表塌陷危险性预测[J].中国岩溶.2020,39(4): 622-628.

DOI:10.11932/karst2020y30

基于PCA-PSO-SVM的凡口铅锌矿 地区岩溶地表塌陷危险性预测

周爱红^{1,2},牛鹏飞¹,袁颖^{1,2},黄虎城³

(1. 河北地质大学城市地质与工程学院, 石家庄 050031; 2. 河北省高校生态环境地质应用技 术研发中心,石家庄 050031;3.山西省地质调查院,太原 030006)

摘 要: 岩溶地表塌陷是由多个影响因素共同作用导致地面形成塌陷坑(洞)的一种动力地质现象, 具有隐蔽性和突发性的特点,常规简单数学模型难以对地表塌陷危险性准确预测。文章先通过主成 分分析法(PCA)对选取的地下水位、地下水位波动幅度、给水度等11个影响因素提取5个主成分,对 导致地表塌陷危险性的主成分进行全新的解释,同时引入粒子群算法(PSO)优化的支持向量机 (SVM)方法,建立PCA-PSO-SVM 岩溶地表塌陷危险性预测模型,并结合凡口铅锌矿地区工程实 例,将预测结果与单一的SVM模型预测结果进行对比,表明PCA-PSO-SVM 危险性预测模型精度 更高,可以更好地为岩溶地表塌陷防治工作提供依据。

关键词:岩溶地表塌陷;主成分分析;粒子群;支持向量机;危险性预测

中图分类号:P642.25 文献标识码:A

文章编号:1001-4810(2020)04-0622-07 开放科学(资源服务)标识码(OSID):



0 引 言

地表塌陷是我国常见的一种地质灾害,可分为 采空区塌陷和岩溶塌陷[1-2]。岩溶地表塌陷存在于多 数岩溶地质环境的矿山,具有隐蔽性和突发性的特 点,较难预测塌陷的危险性,严重影响矿区生态环 境,对企业安全生产构成威胁[3-4]。

岩溶地表塌陷致灾机理具有非线性动力学的特 点,影响地表塌陷危险性的因素众多,各因素与塌陷 稳定性之间的关系是未知且非线性的,一种简单的 数学模型难以准确地反映出各因素之间的不确定、 且复杂的关系^[5]。长期以来,针对岩溶地表塌陷危险 性预测研究,国内外专家和学者们提出了众多预测 方法。目前,用于地表塌陷危险性预测的方法主要 有模糊综合评判法^[6-7]、灰色关联度法^[8]、聚类分析 法^[9-10]、神经网络法^[11-13],上述方法在地表塌陷危险性 预测中取得了一定的效果,但也存在模糊评判结果 出现超模糊现象;灰色关联度法处理数据波动加大, 降低模型精度;聚类分析缺乏可比性;神经网络法易 陷入局部最优等问题。

基于以上分析,本文综合利用主成分分析法 (Principle Component Analysis, PCA)的特征提取能 力^[14-15], 粒子群算法(Particle Swarm Optimization, PSO)对支持向量机(Support Vector Machine, SVM)参 数的全局寻优效果16,以及支持向量机方法在解决 小样本及非线性等实际问题中的优势^[17],建立 PCA-PSO-SVM的岩溶地表塌陷危险性预测模型。通过对 文献^[18]中凡口铅锌矿地区的地下水位、地下水位波

基金项目:国家自然科学基金资助项目(41807231);河北省自然科学基金项目(D201903182);河北省教育厅青年基金项目(QN2019196);河北省 教育厅在读研究生创新能力培养资助项目(CXZZSS2019115);山西省国土资源厅省级地质勘查项目(SXZDF20170820)

第一作者简介:周爱红(1976一),女,博士,教授,研究领域为岩土工程、地质灾害。E-mail:sensiblecall@163.com。

通信作者:牛鹏飞(1995-),男,硕士研究生,研究领域为地质灾害防治。E-mail:1843591926@qq.com。。 收稿日期:2019-09-24

动幅度、给水度、渗透系数包括横向渗透系数和纵向 渗透系数、贮水系数、覆盖层厚度、人工抽水强度、河 流和湖泊、降雨量和井下涌水点含沙率11个影响因 素进行主成分分析,去除这11个因素之间的冗余属 性,重新线性组合出这些影响因素的主成分,作为支 持向量机模型的学习样本和预测样本,有效地解决 了影响因素之间的复杂关系,为岩溶地表塌陷危险 性的预测提供了一种新思路。

1 基本原理

1.1 主成分分析

PCA是一种数学降维的方法,主要是设法将原来 众多具有一定相关性的变量,重新组合成一组新的 线性无关综合变量来代替原来变量^[19-20]。具体步骤:

①假设样本观测数据为:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$
(1)

按照如下方法对原始数据进行标准化处理:

$$x_{ij}^* = \frac{x_{ij} - x_j}{\sqrt{Var(x_j)}} \quad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, n) \quad (2)$$

式中:
$$\overline{x_j} = \frac{1}{n} \sum_{i=1}^n x_{ij}$$
, $Var(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \overline{x_j})^2$ (j =

 $1, 2, \dots, p)_{\circ}$

②数据标准化处理后的相关系数为:

$$R = X^* T X^* \tag{3}$$

式中:X*为标准化后数据矩阵。

③计算相关系数矩阵R的特征值 $(\lambda_1, \lambda_2, \dots, \lambda_p)$ 和相应的特征向量。

④选择重要的主成分。根据各个主成分累计贡 献率的大小取前 k 个主成分,对于 k 的选取,一般要求 累计贡献率达到 85% 以上,这样才能保证线性组合 后变量能包括绝大多数原始变量。

1.2 支持向量机

1.2.1 核心思想

把数据非线性映射到高维核空间,在核空间构造具有低VC维的最优分类超平面^[21-22]。

假定训练集为 $T=\left\{ \left(x_{1},y_{1}\right) ,\cdots \cdots \left(x_{1},y_{1}\right) \right\} \in \left(X\times \right.$

Y)^{*l*},其中 $x_i \in X = R^l, y_i \in Y = \{1, -1\}, i = 1, ...l; T 为$ 整个数据样本集,*l* $为整个数据样本总数,<math>(X \times Y)^l$ 为 X 和 Y 组成的*l*维样本空间,*X*为属性样本数据集,*R^l* 为*X*所在的*l*维空间,*x_i*为第*i*个属性样本向量数据,*Y* 为分类样本数据集,*y_i*为第*i*个分类样本数据。

选择合适的核函数 $K(x_i, x_j)$,构造函数,然后求解最优化问题:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{l} \sum_{i=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{j=1}^{l} \alpha_j \qquad (4)$$

$$s.t. \quad \sum_{i=1}^{l} y_i \alpha_i = 0 \qquad (5)$$

$$t. \quad \sum_{i=1}^{n} y_i \alpha_i = 0 \tag{5}$$

式中: x_i, x_j 分别为核函数中的不同模式;C为合适的 惩罚因子参数 $(0 \le \alpha_i \le C, i = 1, \dots, l)_{\circ}$

求得最优解

$$\boldsymbol{\alpha}^* = \left(\boldsymbol{\alpha}_1^*, \cdots \boldsymbol{\alpha}_l^*\right)^T \tag{6}$$

式中: α_i 为[0, C]区间内的第i个常量; α^* 为一组 α_i 的最优解向量。

选择 α^* 的一个正分量 $0 < \alpha_j^* < C$,并根据此计算阈值:

$$b^* = y_j - \sum_{i=1}^{l} y_i \alpha_i K(x_i, x_j)$$
(7)

构造决策函数:

$$f(x) = \operatorname{sgn}\left[\sum_{i=1}^{l} y_i \alpha_i^* K(x_i, x_j) + b^*\right]$$
(8)

式中:sgn为符号函数;x_i,x_j表示样本数据中地下水 位、地下水位波动幅度、给水度等11个岩溶地表塌陷 危险性的评价指标。

2 预测模型的建立

2.1 评价指标的确定

以文献[18]中的凡口铅锌矿地区岩溶地表塌陷 危险性实测数据为例,选取地下水位 (x_1) 、地下水位 波动幅度 (x_2) 、给水度 (x_3) 、横向渗透系数 (x_4) 、纵向 渗透系数 $(x_5)、贮水系数(x_6)、覆盖层厚度<math>(x_7)$ 、人工 抽水强度 $(x_8)、河流和湖泊(x_9)、降雨量<math>(x_{10})$ 和井下 涌水点含沙率 (x_{11}) 11个影响因素作为地表塌陷危险 性的评价指标。

凡口铅锌矿位于广东省韶关市南部仁化县,是 中金岭南有色金属股份有限公司下属的重要金属矿 山,典型的岩溶地质大水矿山。从凡口铅锌矿地区 实测的21组样本数据(地表塌陷非常易陷区11个, 易陷区5个,临界区2个,基本稳定区2个,稳定区1 个)中随机抽取16组作为学习样本(表1),剩下5组 作为预测样本(表2),检验建立的PCA-PSO-SVM地 表塌陷危险性预测模型的准确性。表2中输出列"1" 表示地表塌陷非常易陷区、"0.75"表示易陷区、 "0.5"表示临界区、"0.25"表示基本稳定区、"0"表示 稳定区。

表1 学习样本 Table 1 Learning samples

序号	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	输出
1	96.33	0.44	0.0510	0.045	0.023	0.006 12	19.840	10 303	0.70	26.8	0.036	0.5
2	99.37	4.11	0.0470	0.030	0.010	0.00574	11.203	9 565	0.30	18.7	0.103	1
3	-16.36	0.08	0.1189	9.400	3.060	0.00500	22.130	10 303	0.75	26.8	0.036	0.5
4	3.17	2.39	0.0513	7.400	0.860	0.00500	16.890	10 303	0.80	26.8	0.036	0.25
5	97.28	3.03	0.0505	0.600	0.720	0.00500	14.780	12 403	1.00	160.5	0.503	1
6	94.60	13.83	0.0370	0.290	0.090	0.00365	2.840	12 403	0.90	0.8	0.637	1
7	-3.44	13.74	0.0469	0.900	0.048	0.00500	28.900	9 565	0.70	18.7	0.103	1
8	45.13	4.02	0.0606	0.640	0.200	0.00507	16.700	12 403	0	160.5	0.503	1
9	58.15	6.77	0.0500	0.120	0.055	0.00501	30.690	10 303	0	26.8	0.036	0
10	76.93	2.02	0.0421	0.230	0.300	0.00501	17.030	12 403	0	160.5	0.503	1
11	85.37	0.95	0.0530	0.910	0.720	0.004 21	4.910	12 403	0.70	0.8	0.637	1
12	95.26	0.69	0.0330	0.230	0.110	0.004 32	17.600	10 303	0.90	26.8	0.036	0.25
13	91.50	5.27	0.0430	0.820	0.720	0.00574	5.920	12 403	1.00	160.5	0.503	1
14	72.09	0.33	0.0504	0.010	0.400	0.000 50	23.300	9 565	0.75	18.7	0.103	1
15	86.80	0.04	0.0640	0.210	0.130	0.00632	1.890	12 403	0.85	0.8	0.637	1
16	76.59	0.57	0.0390	2.040	1.720	0.004 35	14.930	10 303	0.40	26.8	0.036	0.75

表2 预测样本

Г	abl	le	2	Pre	edic	etin	g	sam	pl	e
---	-----	----	---	-----	------	------	---	-----	----	---

序号	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	输出
17	2.23	2.60	0.0501	0.22	0.09	0.00500	21.800	10 303	0.7	26.8	0.036	0.75
18	81.23	6.25	0.0502	0.51	0.40	0.00500	12.670	10 303	0.8	26.8	0.036	0.75
19	99.37	4.11	0.0470	0.03	0.01	0.00574	16.239	12 403	0.3	160.5	0.503	1
20	85.91	2.47	0.038	0.64	0.46	0.00532	18.74	10 303	0.2	26.8	0.036	0.75
21	83.01	0.11	0.024	0.41	0.31	0.004 21	15.56	10 303	0.1	26.8	0.036	0.75

2.2 地表塌陷危险性评价指标的主成分提取

表3为各评价指标的公因子方差比。由表3可知,除了给水度(x₃)有21.8%的信息、河流和湖泊(x₉)有26.2%的信息、降雨量(x₁₀)有20.1%的信息未被提取外,其它8个评价指标的信息被提取较为充分。

表4为11个评价指标计算标准化后的数据特征 值及累计方差贡献率。由表4可以看出,前5个主成 分的累计贡献率为86.776%>85%,取前5个主成分 代表原来11个变量所携带的信息。

表5为通过PCA中最大方差法得到的因子负荷 矩阵,从表5中可以看出,第1主成分主要反映地下 水位 (x_1) 、给水度 (x_3) 、横向渗透系数 (x_4) 和纵向渗透 系数 (x_5) 4个影响因素的信息;第2主成分主要反映 人工抽水强度 (x_8) 、降雨量 (x_{10}) 和井下涌水点含沙率 (x_{11}) 3个因素的信息;第3主成分主要反映覆盖层厚 度 (x_7) 、河流和湖泊 (x_9) 2个因素的信息;第4主成分 主要反映地下水位波动幅度 (x_2) 的信息;第5主成分 主要反映贮水系数 (x_6) 的信息。

表6为通过主成分分析得到的因子得分系数矩阵。11个评价指标通过该系数矩阵重新线性组合成 5个主成分,例如,

 $F_{1} = -0.259x_{1} - 0.009x_{2} + 0.310x_{3} + 0.303x_{4} + 0.287x_{5} - 0.004x_{6} + 0.065x_{7} + 0.059x_{8} + 0.065x_{9} + 0.086x_{10} + 0.039x_{11} (9)$

表3 各评价指标公因子方差比

Table 3 Variance ratios of common factors

影响因素	原始值	公因子方差比
x_1	1.000	0.849
x_2	1.000	0.899
x_3	1.000	0.782
x_4	1.000	0.871
x_5	1.000	0.844
x_6	1.000	0.989
x_7	1.000	0.890
x_8	1.000	0.954
x_9	1.000	0.738
x_{10}	1.000	0.799
x_{11}	1.000	0.933

表4 主成分特征值及贡献率

Table 4 Eigenvalues and contribution rates of principle compo-

nent

十古		初始特征	值	损	取的主成	访分
土成	特征	方	累积方		方	累积方
177	值	差/%	差/%	百月	差/%	差/%
1	3.606	32.783	32.783	3.606	32.783	32.783
2	2.426	22.059	54.842	2.426	22.059	54.842
3	1.437	13.060	67.902	1.437	13.060	67.902
4	1.189	10.808	78.709	1.189	10.808	78.709
5	0.887	8.066	86.776	0.887	8.066	86.776
6	0.531	4.826	91.602			
7	0.403	3.667	95.268			
8	0.294	2.670	97.939			
9	0.108	0.981	98.920			
10	0.095	0.860	99.779			
11	0.024	0.221	100.000			

式(9)中,*F*₁表示第一主成分。同理,可以写出 第二、第三、第四和第五主成分,依此可以得到重新 线性组合成的5个主成分。

表7为重新计算21组原始数据样本的主成分。 把5个线性无关的主成分作为支持向量机模型输入 变量,既可以降低评价指标的维数,也能够提高支持 向量机模型的运算效率。

2.3 PSO-SVM模型的建立

以确定的5个主成分作为输入变量,岩溶地表塌 陷状态作为输出变量。选择径向基函数(其形式为

		l'able 5 Mat	rix of factor	load	
影响			主成分		
因素	1	2	3	4	5
x_4	0.912	-0.155	0.090	-0.072	0.041
x_3	0.873	0.056	0.047	-0.043	0.106
x_5	0.851	-0.021	0.099	-0.312	-0.111
x_1	-0.732	0.331	0.198	-0.405	0.022
x_8	-0.068	0.959	0.114	-0.012	0.127
x_{11}	-0.133	0.922	0.235	0.097	0.023
x_{10}	0.005	0.738	-0.498	-0.004	0.080
x_9	0.211	0.021	0.809	0.156	-0.119
x_7	0.212	-0.541	-0.682	0.239	-0.175
x_2	-0.173	0.065	0.086	0.926	0.021
x_6	0.051	0.144	-0.059	0.022	0.981

表5 因子负荷矩阵

表6 因子得分系数矩阵

Table 6 Matrix of factor scor

影响			主成份		
因素	F_1	F_2	F_3	F_4	F_5
x_1	-0.259	0.026	0.117	-0.359	0.013
x_2	-0.009	0.039	0.075	0.766	0.008
x_3	0.310	0.082	0.007	0.013	0.069
x_4	0.303	0	0.055	-0.017	0.044
x_5	0.287	0.069	0.030	-0.212	-0.131
x_6	-0.004	-0.109	0.023	0.010	0.976
x_7	0.065	-0.088	-0.429	0.178	-0.148
x_8	0.059	0.371	-0.022	0.030	-0.037
x_9	0.065	-0.038	0.549	0.158	-0.07
x_{10}	0.086	0.364	-0.430	0.020	-0.098
x_{11}	0.039	0.358	0.061	0.118	-0.127

表7 原始数据样本的主成分

Table 7 Principal components of original data samples

			0		*
序号	F_1	F_2	F_3	F_4	F_5
1	-0.56	-0.90	0.16	-0.68	1.20
2	-0.85	-0.89	0.17	-0.38	1.14
3	3.78	-0.32	0.09	-5.03	0.06
				•••	•••
20	-0.57	-0.76	-0.55	-0.59	0.55
21	-0.91	-0.72	-0.63	-1.11	-0.32

 $K_g(|x_i - x_j|) = \exp(-g|x_i - x_j|^2)$ 作为支持向量机的 核函数^[23],运用MATLAB语言编写PSO优化SVM程 序确定核函数参数g和惩罚参数*C*,取PSO算法的加



速度参数C1=1.5,C2=1.7,种群数量N=20,终止代数

M=200。粒子群算法寻优过程如图1所示。

图1 粒子群算法寻优过程

Fig. 1 Optimization process of particle swarm algorithm

经过多次学习寻找最优解,最终得到SVM参数 最优值 g=8.1681, C=1.2226。图 2 为用优化好的 SVM模型对学习样本进行回归仿真训练,对比预测 结果与实测数据。由图2可以看出,基于 PCA-PSO-SVM模型对16组学习样本的预测值与实际值非常接 近,精度达到了工程要求,可以作为实际工作中预测 模型。

2.4 预测结果与分析

基于 PCA-PSO-SVM 的地表塌陷危险性预测模型对表 2 的预测样本进行预测,得到该模型预测值与实际值之间的相关系数曲线(图 3),预测样本预测值和实际值进行对比(图 4),表 8 为 PCA-PSO-SVM 模型预测结果误差以及 SVM 模型预测结果误差。



图2 学习样本预测值和实际值对比图

Fig. 2 Comparison between predicted values and actual values of learning samples



Fig. 3 Correlation coefficient curve of predicted and actual values





由图3可知数据预测值和实测值之间相关系数 达到96.41%。由表8可知,PCA-PSO-SVM预测模型 最大绝对误差为0.032,最大相对误差为4.27%,而 SVM预测模型最大绝对误差为0.075,最大相对误差 为10.00%。结合图3、图4和表8可知,PCA-PSO-SVM模型预测结果与实测值吻合较好,精度上优于 表8 误差对比

			Table 8 Err	ror comparison				
皮旦	在险州立厅昌业八区	PCA-PS	SO-SVM 预测	则模型	SVM预测模型			
厅与	厄应住头际重化力区	危险性预测结果	绝对误差	相对误差/%	危险性预测结果	绝对误差	相对误差/%	
1	易陷区	易陷区	0.015	2.00	易陷区	0.067	8.93	
2	易陷区	易陷区	0.023	3.07	易陷区	0.075	10.00	
3	非常易陷区	非常易陷区	-0.011	-1.10	非常易陷区	-0.034	-3.40	
4	易陷区	易陷区	0.001	0.13	易陷区	0.024	3.20	
5	易陷区	易陷区	0.032	4.27	易陷区	0.039	5.20	

SVM模型。相比于文献[18]中的BP神经网络预测模型,PCA-PSO-SVM预测模型能够较好地处理小样本数据,具有学习速度快,预测精度高的特点^[24],解决了各因素与塌陷稳定性之间未知且非线性的问题,能够很好地应用于岩溶地表塌陷地区的实际工程。

2.5 讨 论

(1)由主成分分析的原理可知:只需取累计贡献 率>85%的前几个主成分进行预测分析,就能很好地 抽取所有影响因素所携带的水文地质条件、地层厚 度、人类活动和降雨等信息,并且实现降维,进而降 低SVM模型的计算复杂度。但最终按照累计贡献率 来确定主成分达到理想的预测效果,还应结合实际 工程预测精度要求等来综合考虑;

(2)本文在岩溶地表塌陷危险性预测中,影响因 素达到了11个,经过主成分抽取后,5个主成分的累 计贡献率就超过了85%,降维效果明显。但对于影响 因素较少的情况,主成分分析降维的效果可能并不 明显,因此,实际应用时应考虑主成分分析的适用 性,并和不引入主成分分析的预测模型进行预测效 果对比,有效降低预测模型的复杂度。

3 结 论

(1)岩溶地表塌陷的发生受多个因素影响,常规 数学模型难以解释因素间未知且非线性的关系。本 文通过 PCA-PSO-SVM模型很好地解决了这一问题, 采用特征提取的5个重新组合的线性无关的主成分 代替了原来的11个影响因素,揭示了岩溶地表塌陷 与影响因素间的内在关系;

(2)采用PCA分析法的降维思想,用5个线性无关的主成分表达了11个影响因素所携带信息量的 86.776%,提高了模型的学习速度和预测精度。影响 岩溶地表塌陷因素被解释的更加透彻,为之后岩溶 地表塌陷的研究提供了新思路; (3)采用PCA-PSO-SVM预测模型对凡口铅锌矿 地区原始实测数据进行学习和预测的结果显示,凡 口铅锌矿地区21组数据中地表塌陷风险等级为非常 易陷区11个,易陷区5个,临界区2个,基本稳定区2 个,稳定区1个,该结果与当地实际情况勘查数据较 为吻合,表明该方法具有较强可行性。PCA-PSO-SVM地表塌陷危险性预测模型学习效果好,预测精 度较高,可为岩溶地表塌陷防治工作提供依据。

参考文献

- [1] 蒙彦,雷明堂.岩溶塌陷研究现状及趋势分析[J].中国岩溶, 2019,38(3):411-417.
- [2] 秦佳玉,梅钢,徐能雄.面向采空塌陷离散元模拟的地表沉降 确定方法[J].矿业研究与开发,2019,39(9):46-50.
- [3] Liu Z, Cui B, Liang Y, et al. Study on Foundation Deformation of Buildings in Mining Subsidence Area and Surface Subsidence Prediction [J]. Geotechnical and Geological Engineering, 2019, 37(3): 1755-1764.
- [4] 罗周全,徐海,杨彪,王益伟.矿区岩溶地表塌陷神经网络预 测研究[J].中国地质灾害与防治学报,2011,22(3):39-44.
- [5] 周泽.岩溶矿区采动裂隙发育及其地表塌陷规律研究[D].湘 潭:湖南科技大学,2017.
- [6] He K, Jia Y, Chen W, et al. Evaluation of Karst Collapse Risks Induced By Over—pumping and Karst Groundwater Resource Protection in Zaozhuang Region, China [J]. Environmental Earth Sciences, 2014, 71(8): 3443-3454.
- [7] Gan L, Zuo J, Wang Y, et al. Comprehensive Health Condition Assessment on Partial Sewers in a Southern Chinese City Based on Fuzzy Mathematic Methods [J]. Frontiers of Environmental Science & Engineering, 2014, 8(1): 144-150.
- [8] Gao C, Li S, Wang J, et al. The Risk Assessment of Tunnels Based on Grey Correlation and Entropy Weight Method [J]. Geotechnical and Geological Engineering, 2018, 36(3): 1621-1631.
- [9] Du J, He R, Sugumaran V. Clustering and Ontology-based Information Integration Framework for Surface Subsidence Risk Mitigation in Underground Tunnels[J]. Cluster Computing, 2016, 19(4): 2001-2014.
- [10] Sahu SP, Yadav M, Das AJ, et al. Multivariate Statistical Ap-

proach for Assessment of Subsidence in Jharia Coalfields, India [J]. Arabian Journal of Geosciences, 2017, 10(8): 191-201.

- [11] Han D, Li X. The Surface Subsidence Prediction of Shield Construction Based on the Fuzzy Neural Network[C]. Springer Singapore, 2018: 190-197.
- [12] 管佳林,罗周全,杨彪,等.矿区岩溶地表塌陷神经网络预测 模型研究[J].中国安全科学学报,2011,21(9):28-33.
- [13] Salehi T, Shokrian M, Modirrousta A, et al. Estimation of the Collapse Potential of Loess Soils in Golestan Province Using Neural Networks and Neuro-fuzzy Systems [J]. Arabian Journal of Geosciences, 2015, 8(11): 9557-9567.
- [14] 杨斌,杨永军,张亚,等.基于主成分分析与神经网络复合模型的汽轮机排汽焓计算[J].中国电力,2018,51(1):126-132.
- [15] Mondal I, Bandyopadhyay J, Dhara S. Detecting Shoreline Changing Trends Using Principle Component Analysis in Sagar Island, West Bengal, India[J]. Spatial Information Research, 2017, 25(1): 67-73.
- [16] 郭天颈,张菊清,韩煜,等.基于粒子群优化支持向量机的延 长县滑坡易发性评价[J].地质科技情报,2019,38(3): 236-243.
- [17] Yu H. Encyclopedia of Database Systems [M]. Springer New York, 2018: 3854-3857.

- [18] 杨彪.矿山地下水害防治工程可视化及地表塌陷预测研究 [D].长沙:中南大学,2011.
- [19] 韩冉冉.建筑工程施工班组安全氛围测度研究[D].南京:东 南大学,2015.
- [20] Mulas M, Bonacini F, Petitta M, et al. Landslide Zoning Using the Principal Component Analysis on Monitoring Data: The Sauna Earth Slide-Earth Flow (Parma, Italy)[C]. Workshop on World Landslide Forum. Springer, Cham, 2017.
- [21] 薛云,戴塔根,杨自安,等.基于光谱和纹理的SVM矿化蚀变 信息提取研究[J].地质找矿论丛,2008,23(3):254-259.
- [22] Jiang X, Lu W, Zhao H, et al. Quantitative Evaluation of Mining Geo-environmental Quality in Northeast China: Comprehensive Index Method and Support Vector Machine Models [J]. Environmental Earth Sciences, 2015, 73 (12) : 7945-7955.
- [23] 周爱红, 倪莹莹, 尹超,等. 一种盾构施工引起的地面沉降预 测方法[J]. 测绘科学, 2018, 43(3):167-172.
- [24] De gregorio L, Callegari M, Mazzoli P, et al. Operational River Discharge Forecasting with Support Vector Regression Technique Applied to Alpine Catchments: Results, Advantages, Limits and Lesson Learned[J]. Water Resources Management, 2018, 32(1): 229-242.

Prediction of karst surface subsidence risk in the Fankou lead-zinc mine area based on PCA-PSO-SVM

ZHOU Aihong^{1,2}, NIU Pengfei¹, YUAN Ying^{1,2}, HUANG Hucheng³

(1. School of Urban Geology and Engineering, Hebei Geologic University, Shijiazhuang Hebei 050031, China; 2. Center of Applied Technology Development for Eco-environmental Geology, Hebei Geologic University, Shijiazhuang Hebei 050031, China; 3. Shanxi Institute of Geological Survey, Taiyuan Shanxi 030006, China)

Abstract Karst surface subsidence is a dynamic geological phenomenon with the characteristics of concealment and suddenness, which results from the joint effect many factors. Thus, it is difficult to accurately predict the risk of surface subsidence by the conventional simple mathematical model. In this paper, the Principal Component Analysis (PCA) is used to extract five principal components from 11 influencing factors, including groundwater level, fluctuation range of groundwater level and water supply, so as to make a new interpretation of the principal components leading to the risk of surface subsidence. Additionally, the Support Vector Machine (SVM) method optimized by Particle Swarm Optimization (PSO) is introduced to establish a PCA–PSO–SVM model for prediction of risk of karst surface subsidence. Finally, combined with the engineering example of the Fankou lead—zinc mine, the prediction results by the above proposed model are compared with those obtained by the single SVM model. The results show that the PCA–PSO–SVM risk prediction model has higher accuracy, which can provide a basis for prevention and control of karst surface subsidence.

Key words karst surface subsidence, principal component analysis, particle swarm optimization, support vector machine, risk prediction

(编辑 张玲)