

袁颖, 张天亮. 基于主成分分析和支持向量机的苏锡常地区地裂缝危险性预测[J]. 灾害学, 2019, 34(4): 57-63.
[YUAN Ying and ZHANG Tianliang. Prediction of ground fissures risk in suxichang area based on principal component analysis and support vector machine[J]. Journal of Catastrophology, 2019, 34(4): 57-63. doi: 10.3969/j.issn.1000-811X.2019.04.011.]

基于主成分分析和支持向量机的苏锡常地区 地裂缝危险性预测^{*}

袁颖^{1,2}, 张天亮¹

(1. 河北地质大学 勘查技术与工程学院, 河北 石家庄 050031;
2. 河北省高校生态环境地质应用技术研发中心, 河北 石家庄 050031)

摘 要: 地裂缝地质灾害是由多种复杂因素共同导致的岩层、土体开裂现象, 常规数学模型难以对地裂缝的危险性做出准确预测。该文首先采用主成分分析(PCA)方法对选取的导水系数、水位、粘性土层厚度、基岩起伏度、基岩埋深5个影响因素提取3个主成分, 对导致地裂缝发生的主成分进行了全新的解释, 同时引入支持向量机的方法(SVM)建立了基于主成分分析和支持向量机的苏锡常地区地裂缝危险性预测模型。并结合工程实例将预测结果与SVM模型预测结果进行比较分析。结果表明: 基于主成分分析和支持向量机的地裂缝危险性预测模型精度较高, 具有一定的参考价值, 可为预测苏锡常地区地裂缝危险性提供有效依据。

关键词: 地裂缝; 主成分分析; 支持向量机; 苏锡常; 危险性预测

中图分类号: P542; X43; X915.5 **文献标志码:** A **文章编号:** 1000-811X(2019)04-0057-07
doi: 10.3969/j.issn.1000-811X.2019.04.011

地裂缝是一种常见的地区性环境地质灾害, 造成的危害和损失巨大。因此, 地裂缝危险性的准确预测是保障人民生命财产安全, 进行地区开发建设与灾害防治的必要工作^[1]。由于地裂缝发生的影响因素众多, 且这些因素之间往往具有很强的随机性和不确定性, 难以用一种简单的数学模型准确的反映出各影响因素之间的复杂关系^[2]。

长期以来, 针对地裂缝危险性预测研究, 专家和学者们提出了众多预测方法。目前, 用于地裂缝危险性的预测方法主要有神经网络^[3-6]、灰色关联度法^[7-8]、模糊综合评判法^[9-10]、信息量法^[11-12]。以上方法在地裂缝危险性预测中都取得了一定的效果, 但也存在不足之处。神经网络的收敛速度比较慢, 容易导致局部最优, 使得预测结果误差较大; 灰色关联度理论在处理数据波动较大的情况时, 容易降低模型的预测精度; 模糊综合评判模型在选择何种隶属函数时仍有较大难度; 信息量方法不适用于非构造成因的地裂缝危险性预测。

基于以上分析, 本文在结合主成分分析法(Principle Component Analysis, PCA)较强提取因子能力的同时^[13], 充分发挥支持向量机(Support

Vector Machine, SVM)在解决小样本数据以及非线性等实际问题中的优势^[14], 提出了基于主成分分析和支持向量机的地裂缝危险性预测模型。本文通过主成分分析法对文献^[15]中选取的导水系数(m^2/d)、水位(m)、粘性土层厚度(m)、基岩起伏度(m)、基岩埋深(m)这5个评价指标进行分析, 通过剔除这5个评价指标之间的关联性, 重新进行线性组合提取出这些评价指标的主成分, 并对其进行分析讨论。最后通过支持向量机用于苏锡常地区地裂缝危险性实测数据的训练并建立相应的预测模型。有效解决了复杂影响因素情况下模型难以确定的问题, 为地裂缝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

较多的影响因素之间往往存在一定的关联性, 存在信息量的重叠现象, 在造成复杂计算程度的同时, 还将直接影响计算的效率以及准确度^[16]。主成分分析法能很好的解决这一问题。主成分分

* 收稿日期: 2019-04-09 修回日期: 2019-06-19

基金项目: 国家自然科学基金资助项目(41807231); 河北省教育厅重点资助项目(ZD2016038); 河北省自然科学基金项目(D2019403182); 河北省教育厅青年基金项目(QN2019196)

第一作者简介: 袁颖(1976-), 男, 江西景德镇人, 教授, 博士, 主要研究方向为环境和灾害地质、工程地质的教学和科研工作。

E-mail: yuanyingson@163.com

析法主要是通过降维的思想构造原始因素之间的线性组合,从而达到减少维数且产生的主成分之间互不相关的目的,这些产生的主成分能够反映原始因素所提供的大部分信息^[17]。具体步骤如下:

(1)原始数据进行标准化。假设观测样本数据矩阵为:

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

标准化处理原始数据:

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

式中:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \text{Var}(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2, j=1, 2, \cdots, p. \quad (3)$$

(2)计算样本相关系数矩阵。原始数据经过标准化处理以后,我们假定其仍然用 X 表示,则可得数据相关系数:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (4)$$

式中:

$$r_{ij} = \frac{\text{Cov}(x_i, x_j)}{\sqrt{\text{Var}(x_i)} \sqrt{\text{Var}(x_j)}} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}, n > 1. \quad (5)$$

(3)计算样本相关系数矩阵 R 的特征值($\lambda_1, \lambda_2, \cdots, \lambda_p$)以及相应的特征向量:

$$a_i = (a_{i1}, a_{i2}, \cdots, a_{ip}), i=1, 2, \cdots, p. \quad (6)$$

(4)选取重要主成分。主成分分析后可得到 p 个主成分,值得注意的是主成分的方差是呈递减形式的,导致其所承载的信息量也呈递减形式。因此我们在进行主成分分析时会依据各个主成分累计贡献率的大小来选取前 k 个主成分。贡献率 = $\frac{\lambda_i}{\sum_{i=1}^p \lambda_i}$, 是指某一个主成分的方差占全部方差的比重,也就是这个主成分特征值占全部特征值的比重。因此贡献率越大,表明该主成分所承载的信息量越多。在进行主成分分析时,利用各主成分累计贡献率(一般达到 85% 以上)确定系统综合得分,从而确定主成分的个数 k 并对系统做出科学评价。

1.2 支持向量机

支持向量机作为一种有监督的学习方法,已知学习样本的类别,通过学习样本和类别之间的对应关系,预测新的训练点所对应的类别。在学习过程中, SVM 算法构建了一个分割两类的超平面(图 1),当对新的样本分类时,基于学习所得的超平面,旨在使两类之间的分割达到最大化,即使预测错误的概率最小化。

习过程中, SVM 算法构建了一个分割两类的超平面(图 1),当对新的样本分类时,基于学习所得的超平面,旨在使两类之间的分割达到最大化,即使预测错误的概率最小化。

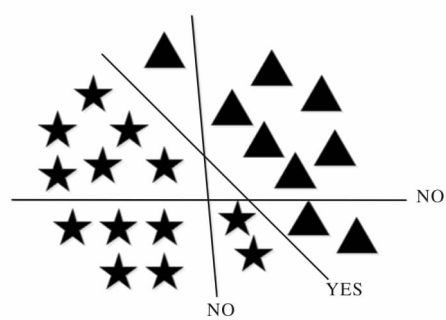


图 1 支持向量机划分算法示意图

假设容量为 n 的训练样本集 $\{(x_i, y_i), i=1, 2, \cdots, n\}$ 由两个类别构成,假设 x_i 属于第一类,则 $y_i = 1$; x_i 属于第二类,则 $y_i = -1$ 。在学习过程中,若存在分类超平面: $w^T x + b = 0$ (w 为可调的权值矢量; b 为偏置值)即具有相同类别的样本能够落在分类超平面的同一侧,则满足:

$$\begin{cases} w^T x_i + b \geq 1, y_i = 1; \\ w^T x_i + b \leq -1, y_i = -1. \end{cases} \quad (7)$$

式中: $w^T x_i + b = 1$ 和 $w^T x_i + b = -1$ 表示两个平行的边界超平面,这两个边界超平面的距离是 $\frac{2}{\|w\|}$ 。为了使两个平面之间的距离最大化,则需满足其倒数最小化,即:

$$\min \frac{\|w\|}{2} = \frac{1}{2} \sqrt{w^T w}. \quad (8)$$

由此可得以下目标规划问题:

$$\begin{cases} \min \frac{\|w\|}{2} = \frac{1}{2} \sqrt{w^T w}; \\ \text{s. t. } y_i(w^T x_i + b) \geq 1, i=1, 2, \cdots, n \end{cases} \quad (9)$$

式(9)为一个凸优化问题,目标函数是二次的,约束条件则为线性约束,可根据拉格朗日对偶理论转化为对偶问题,如式(10)所示。

$$\begin{cases} \max L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i^T x_j); \\ \text{s. t. } \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0. \end{cases} \quad (10)$$

式中: α_i 和 α_j 为拉格朗日乘值,这里设所得的最优解为 $\alpha^* = [\alpha_1^*, \alpha_2^*, \cdots, \alpha_n^*]^T$, 则可得到最优的 w^* 和 b^* 为:

$$\begin{cases} w^* = \sum_{i=1}^n \alpha_i^* x_i y_i; \\ b^* = -\frac{1}{2} w^{*T} (x_r + x_s). \end{cases} \quad (11)$$

式中: x_r 和 x_s 为两个类别中的任意一对支持向量。得到最优分类函数为:

$$f(x) = \text{sgn} \left[\sum_{i=1}^n \alpha_i^* y_i (x^T x_i) + b^* \right]. \quad (12)$$

式中: x_i 表示样本数据中导水系数、水位、粘性土

层厚度、基岩起伏度、基岩埋深这 5 个地裂缝危险性的评价指标。

2 预测模型的建立

2.1 评价指标的确定

以文献[15]中的苏锡常地区地裂缝危险性实测数据为例, 以导水系数(m^2/d)、水位(m)、粘性土层厚度(m)、基岩起伏度(m)、基岩埋深(m)这 5 个因素作为地裂缝危险性的评价指标。对苏锡常地区地裂缝调查表明, 地面沉降梯度越大, 地裂缝灾害发生的可能性越大, 但由于地面沉降存在不均一性, 因此需对地面沉降梯度进行量化处理^[18]。当地面沉降梯度 $> 0.01\%$ 时, 量化值为 0.9; 当地面沉降梯度处于 $0.005\% \sim 0.01\%$ 时, 量化值为 0.75; 当地面沉降梯度处于 $0.001\% \sim 0.005\%$ 时, 量化值为 0.25; 当地面沉降梯度 $< 0.001\%$ 时, 量化值为 0.1。将样本数据归一化处

理后从 34 组样本数据中随机选出 29 组作为训练样本(表 1), 剩余的 5 组作为测试样本(表 2)。表中 0.9 表示地裂缝危险区、0.75 表示较危险区、0.25 表示次安全区、0.1 表示安全区。

2.2 地裂缝危险性评价指标的主成分提取

为了更好的表示各影响因素之间的相关性大小, 对表 1 和表 2 中的影响因素进行主成分分析, 得到相关系数矩阵如表 3 所示。其中, 相关系数越接近 1 则反映两种影响因素的相关性越大。由表 3 相关系数矩阵可以看出, 导水系数与水位、粘性土层厚度、基岩埋深均有较强的相关性; 粘性土层厚度与基岩埋深有较强的相关性; 如果对这些具有相关性的因素直接进行分析, 可能会导致严重的共线性现象。采用主成分分析方法可以提取适当数量的主成分, 并对各评价指标重新进行线性组合, 可以进一步清楚地阐述各因素之间的关系。

表 1 训练样本归一化处理

序号	导水系数/ (m^2/d)	水位/m	粘性土层厚度/m	基岩起伏度/m	基岩埋深/m	期望输出
1	0.300	0.379	0.711	0.064	0.602	0.9
2	0.400	0.314	0.766	0.066	0.640	0.75
3	0.400	0.068	0.731	0.544	0.654	0.9
4	0.400	0.038	0.669	0.313	0.611	0.9
5	0.100	0.711	0.065	0.267	0.065	0.1
6	0.100	0.727	0.721	0.246	0.705	0.75
7	0.300	0	0.295	1.000	0.306	0.1
8	0.400	0.059	0.152	0.943	0.243	0.1
9	0.400	0.130	0	0.307	0	0.9
10	0.500	0.171	0.648	0.477	0.596	0.9
11	0.050	0.320	0.057	0.696	0.099	0.9
12	0.050	0.351	0.155	0.730	0.129	0.9
13	0.400	0.413	0.673	0.165	0.574	0.9
14	0.100	0.370	0.334	0.623	0.334	0.9
15	0.400	0.024	0.625	0.437	0.541	0.9
16	0.925	0.979	0.407	0.106	0.637	0.1
17	0.500	0.940	0.363	0.179	0.780	0.1
18	1.000	1.000	1.000	0.094	1.000	0.1
19	0.925	0.957	0.842	0.615	0.888	0.1
20	0.300	0.857	0.739	0.126	0.854	0.1
21	1.000	0.916	0.509	0.039	0.787	0.1
22	1.000	0.947	0.551	0.250	0.855	0.1
23	0.925	0.967	0.445	0.022	0.741	0.1
24	0.100	0.669	0.600	0.119	0.445	0.1
25	1.000	0.891	0.711	0.149	0.756	0.1
26	0.925	0.949	0.160	0.691	0.384	0.75
27	0.100	0.399	0.759	0.087	0.616	0.1
28	0.050	0.586	0.758	0	0.732	0.1
29	0.500	0.096	0.574	0.129	0.582	0.9

表 2 测试样本归一化处理

序号	导水系数/(m ² /d)	水位/m	粘性土层厚度/m	基岩起伏度/m	基岩埋深/m	期望输出
1	0.400	0.591	0.740	0.547	0.622	0.1
2	0.400	0.039	0.688	0.273	0.624	0.9
3	0.100	0.730	0.728	0.113	0.749	0.9
4	0.050	0.774	0.611	0.054	0.685	0.1
5	0	0.025	0	0.466	0	0.9

表 3 相关系数矩阵

影响因素	导水系数/(m ² /d)	水位/m	粘性土层厚度/m	基岩起伏度/m	基岩埋深/m
导水系数/(m ² /d)	1.000	0.505	0.147	-0.109	0.470
水位/m	0.505	1.000	0.190	-0.467	0.533
粘性土层厚度/m	0.147	0.190	1.000	-0.502	0.845
基岩起伏度/m	-0.109	-0.467	-0.502	1.000	-0.555
基岩埋深/m	0.470	0.533	0.845	-0.555	1.000

表 4 公因子方差比

影响因素	原始值	公因子方差比
导水系数/(m ² /d)	1.000	0.916
水位/m	1.000	0.885
粘性土层厚度/m	1.000	0.971
基岩起伏度/m	1.000	0.891
基岩埋深/m	1.000	0.955

表 5 主成分特征值及贡献率

主成分	起始特征值			提取的主成分		
	特征值	方差/%	累计方差/%	特征值	方差/%	累计方差/%
1	2.791	55.827	55.827	2.791	55.827	55.827
2	1.108	22.165	77.992	1.108	22.165	77.992
3	0.718	14.356	92.348	0.718	14.356	92.348
4	0.321	6.427	98.775			
5	0.061	1.225	100.000			

表 4 为各评价指标的公因子方差比。可知，其中 3 个评价指标的信息被提取得较充分，而水位有 11.5% 以及基岩起伏度有 10.9% 的信息未被提取。

对 5 个评价指标进行主成分分析，得到各主成分特征值及贡献率如表 5 所示，经过分析可以得到 5 个主成分，按照各主成分特征值从大到小排序为第 1、第 2、……、第 5 主成分。第 1 主成分的特征值为 2.791，其解释方差的贡献率为 55.827%，第 2 主成分的特征值为 1.108，贡献率为 22.165%，第 3 主成分的特征值为 0.718，贡献率为 14.356%。由此前 3 个主成分的累计贡献率为 92.348%，同时结合图 2 的碎石图得出，从第 4 个主成分开始，后面的特征值都相对较低，表明对原始数据进行分析时用前 3 个主成分即可。

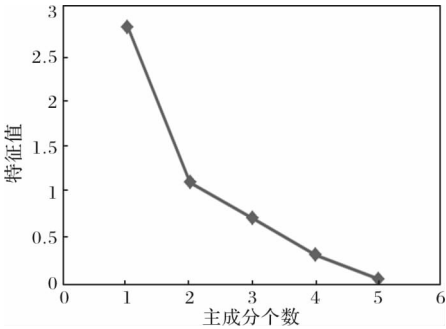


图 2 主成分分析碎石图

通过主成分分析中最大方差法进行因子旋转，

使因子载荷效果更加充分，得到因子载荷矩阵如表 6 所示。通过观察前 3 个主成分在各评价指标影响下的权重系数可知，在第 1 主成分中，系数绝对值较大的影响因素为粘性土层厚度、基岩埋深，表明第 1 主成分与这两个影响因素的关系比较紧密且第 1 主成分主要反映覆盖层厚度以及基岩面的深浅对地裂缝的影响；同理可知第 2 主成分主要反映岩土体的渗透性对地裂缝的影响；第 3 主成分主要反映地下水的赋存状态和基岩面的形态对地裂缝的影响。

表 6 因子载荷矩阵

影响因素	主成分		
	1	2	3
粘性土层厚度/m	0.972	0.011	0.159
基岩埋深/m	0.831	0.395	0.329
导水系数/(m ² /d)	0.155	0.944	0.018
基岩起伏度/m	-0.405	0.068	-0.850
水位/m	0.037	0.614	0.712

通过主成分分析得到的因子得分系数矩阵如表 7 所示。各评价指标可以通过该系数矩阵对各主成分重新进行线性组合，例如 $Y_1 = 0.11x_1 - 0.322x_2 + 0.666x_3 + 0.030x_4 + 0.451x_5$ ，式中 Y_1 表示第 1 主成分， x_1, x_2, x_3, x_4, x_5 分别表示导水系数、水位、粘性土层厚度、基岩起伏度、基岩埋深。同理可写出第 2、第 3 主成分的表达式见式 (13)。

表 7 因子得分系数矩阵

影响因素	主成分		
	1	2	3
导水系数/(m ² /d)	0.011	0.764	-0.289
水位/m	-0.322	0.321	0.588
粘性土层厚度/m	0.666	-0.134	-0.223
基岩起伏度/m	0.030	0.320	-0.765
基岩埋深/m	0.451	0.156	-0.084

$$\begin{cases} Y_1 = 0.011x_1 - 0.322x_2 + 0.666x_3 + 0.030x_4 + 0.451x_5; \\ Y_2 = 0.764x_1 + 0.321x_2 - 0.134x_3 + 0.320x_4 + 0.156x_5; \\ Y_3 = 0.289x_1 + 0.588x_2 - 0.223x_3 - 0.765x_4 - 0.084x_5. \end{cases} \quad (13)$$

使用主成分分析方法对上述样本数据进行分析, 将提取出的 3 个线性无关的主成分代替原来的 5 个影响因素, 并将其作为输入变量以此建立支持向量机预测模型, 使得冗余数据被剔除、由于量纲不同带来的影响被消除, 有效降低了变量的维数和数据之间的相关性, 极大提高了模型的运算效率, 从而对数据有一个更加合理的解释。

2.3 模型的建立

为了建立苏锡常地区地裂缝危险性预测模型, 引入 SVM 方法从样本数据中随机抽取 29 组样本作为训练样本, 余下的 5 组样本作为预测样本。为了使预测模型满足精度要求, 论文选择径向基函数 (RBF) 作为 SVM 的核函数, 并利用 PSO 算法对 SVM 模型的惩罚参数 c 和核函数 g 进行优化取值。粒子群算法的初始参数为: 加速度 $C1 = 1.5$, $C2 = 1.7$, 种群数量 $N = 20$, 终止代数 = 200, 其参数的粒子群算法流程图和寻优过程如图 3 和图 4 所示, 经过多次迭代寻优, 最终得到 SVM 参数最优值 $c = 12.6934$, $g = 14.6204$ 。同时以式 (13) 确定的 3 个主成分作为输入变量, 用该 c 和 g 参数确定的 SVM 模型对训练样本进行回归仿真训练, 对训练样本原始数据与预测结果进行对比, 如图 5 所示, 该模型预测结果的相对误差见图 6。结合图 5 和图 6 可以看出, 基于 PCA-SVM 的预测模型训练样本的预测结果与原始数据极为接近, 相对误差较小, 精度能够满足工程要求。

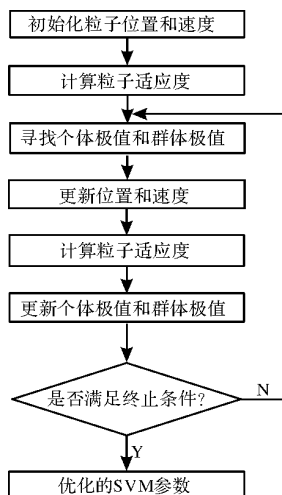


图 3 粒子群优化算法流程图

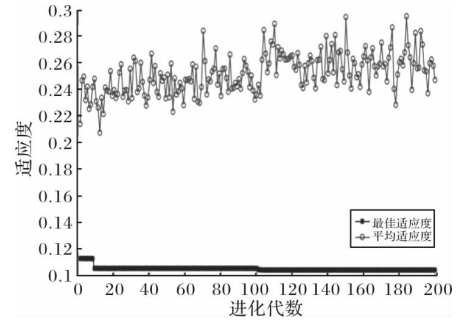


图 4 粒子群算法参数寻优

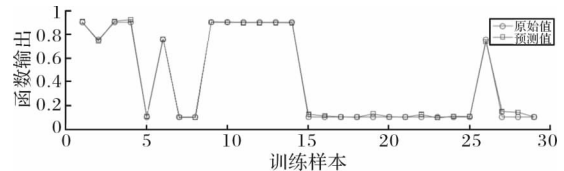


图 5 训练样本原始值和预测值对比图

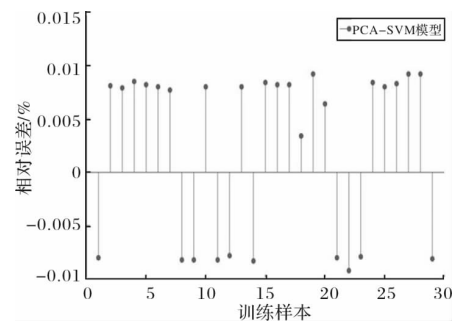


图 6 基于 PCA-SVM 模型的训练样本预测结果相对误差

2.4 预测结果与分析

通过基于 PCA-SVM 的地裂缝危险性预测模型对表 2 的测试样本进行预测, 并与原始数据进行对比, 如图 7 所示。同时可得到该预测模型预测结果与原始值之间的相关系数曲线见图 8。该模型预测结果误差与 SVM 模型预测结果误差进行比较见表 8, 以及预测结果相对误差对比见图 9。

由图 8 可知预测结果与原始值之间的相关系数达到 99.87%。由表 8 可知, 基于 PCA-SVM 的预测模型预测结果的最大绝对误差为 0.0082, 最大相对误差 0.0820%, 而 SVM 预测模型预测结果的最大绝对误差为 0.1267, 最大相对误差 1.2670%。结合图 7、图 8、图 9 以及表 8 可以看出此模型的预测结果与实际值吻合效果较好, 在预测精度上要优于 SVM 模型, 能够满足实际工程的需要。相对于文献^[15]的预测模型来说, 本文所采用的方法能够很好的处理小样本数据, 具有高效、简便的分类、预测过程, 有效解决了“过学习”和“维数灾难”的问题^[19], 应用于苏锡常地区地裂缝危险性实测样本数据的训练和预测中, 效果更佳。

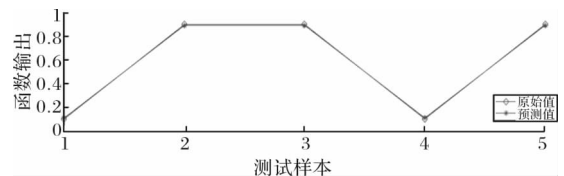


图 7 测试样本原始值和预测值对比图

表 8 误差对比

序号	危险性原始量化分区	基于 PCA 和 SVM 的预测模型			SVM 预测模型		
		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	安全区	安全区	0.0082	0.0820	安全区	0.0380	0.3800
2	危险区	危险区	-0.0080	-0.0089	危险区	-0.0472	-0.0524
3	危险区	危险区	0	0	危险区	-0.0253	-0.0281
4	安全区	安全区	0	0	安全区	0.0267	0.2670
5	危险区	危险区	-0.0081	0.0090	危险区	-0.0892	-0.0991

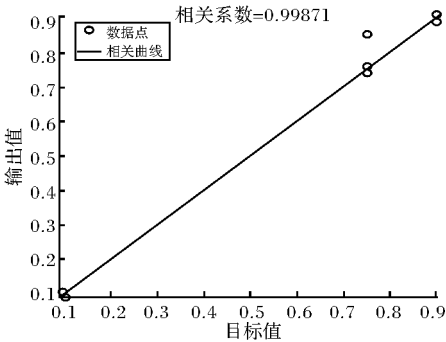


图 8 样本数据原始值和预测值相关系数曲线

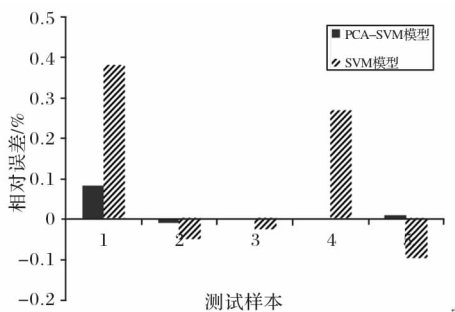


图 9 测试样本预测结果相对误差对比图

3 结论

(1) 地裂缝的发生受到众多因素的共同影响,难以用一种简单的数学模型准确的反映出各影响因素之间的复杂关系。本文通过 PCA 分析方法结合 SVM 模型较好地解决了这一问题,进一步清楚地阐述了各因素之间的关系。

(2) PCA 分析法采用降维的思想确定的 3 个主成分表达了 5 个变量所携带信息量的 92.348%,剔除了变量之间的相关性,提升了训练速度和预测精度。地裂缝发生的影响因素被解释的更加合理,为今后该地区地裂缝的探究提供全新的思路。

(3) 利用基于 PCA-SVM 的地裂缝危险性预测模型,对苏锡常地区 34 组实测数据进行训练并预测,预测结果显示苏锡常地区 34 组样本中地裂缝危险区有 14 个,较危险区有 3 个,次安全区有 0 个,安全区有 17 个,该结果与实际情况基本吻合,能够较为真实的反映该地区地裂缝危险性的实际

情况,具有一定的可行性。基于 PCA-SVM 这一数学方法对苏锡常地区地裂缝样本数据进行训练,建立的该地区地裂缝危险性预测模型,预测精度高,为地裂缝危险性预测提供了一个更实用、准确的方法。

参考文献:

[1] WANG G, YOU G, ZHU J, et al. Earth fissures in Su-Xi-Chang region, Jiangsu, China [J]. *Surveys in Geophysics*, 2016, 37 (6): 1095 - 1116.

[2] LI B, XU Y, ZHANG J, et al. Recent Advances in Computer Science and Information Engineering: Volume 3 [M]. Springer Berlin Heidelberg, 2012: 331 - 338.

[3] WU Q, YE S Y, WU X, et al. A Nonlinear modeling and forecasting system of earth fractures based on coupling of artificial neural network and geographical information system-exemplified by earth fractures in Yuci city, Shanxi, China [J]. *Environmental Geology*, 2003, 45 (1): 124 - 131.

[4] WU Q, DONG D L, WU X, et al. Study on simulation of earth-fissures development and prediction based on gis [J]. *Science in China*, 2001, 44 (1): 77 - 84.

[5] ZHANG W, GAO L, JIAO X, et al. Occurrence assessment of earth fissure based on genetic algorithms and artificial neural networks in Su-Xi-Chang land subsidence area, China [J]. *Geosciences Journal*, 2014, 18 (4): 485 - 493.

[6] LEE S, CHOI JK. Spatial prediction of ground subsidence susceptibility using an artificial neural network [J]. *Environmental Management*, 2012, 49 (2): 347 - 358.

[7] 方冬, 张云, 于军, 等. 基于灰色关联度理论的苏锡常地裂缝危险性评价 [J]. *煤田地质与勘探*, 2018, 46 (2): 137 - 142.

[8] 刘德成, 靳小平, 周自梁. 灰色关联度分析法在北京通州区地裂缝灾害危险性评价中的应用 [J]. *中国地质灾害与防治学报*, 2010, 21 (3): 73 - 76.

[9] 张骏, 卢玉东, 程谦恭, 等. 岩溶区地裂缝灾害程度两级模糊综合评判 [J]. *中国地质灾害与防治学报*, 1996 (4): 1 - 7.

[10] 杨斌. 基于模糊层次综合评估法的西安地铁地裂缝灾害风险评估研究 [D]. 西安: 长安大学, 2017.

[11] 徐光黎, 张家明, 佟永贺, 等. 西安市地裂缝的时空预测预报 [J]. *西北地震学报*, 1992 (4): 75 - 81.

[12] 王兵虎, 马学军, 邵长庆, 等. 基于层次分析和信息量法的地裂缝危险性评价 [J]. *探矿工程 (岩土钻掘工程)*, 2018, 45 (10): 91 - 96.

- [13] 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.
- [14] YU H. Encyclopedia of database systems[M]. Springer New York, 2018: 3854–3857.
- [15] 焦珣. 基于 GA-ANN 技术的苏锡常地区地裂缝危险性评价[D]. 长春: 吉林大学, 2007.
- [16] 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.
- [17] TANG G, LU G. Block principle component analysis with lp-norm for robust and sparse modelling[J]. Journal of Shanghai Jiaotong University (science), 2018, 23(3): 398–403.
- [18] YUN W Z, QIN Y U, YUN Z. Forming process of earth fissure hazard in the Suzhou-Wuxi-Changzhou area[J]. Hydrogeology & Engineering Geology, 2003, 30(1): 67–72.
- [19] CHEN W, Pourghasemi HR, Naghibi SA. A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China[J]. Bulletin of Engineering Geology and the Environment, 2018, 77(2): 647–664.

Prediction of Ground Fissures Risk in Suxichang Area Based on Principal Component Analysis and Support Vector Machine

YUAN Ying^{1,2} and ZHANG Tianliang¹

- (1. Department of Exploration Technology and Engineering, Hebei GEO University, Shijiazhuang 050031, China;
2. Hebei Center for Ecological and Environmental Geology Research, Shijiazhuang 050031, China)

Abstract: Geological hazard of ground fissures is a phenomenon of rock and soil cracking caused by many complex factors. It is difficult to predict the risk of ground fissures accurately by conventional mathematical models. Firstly, the Principle Component Analysis (PCA) method is used to extract three principal components from five influencing factors, namely, transmissibility, water level, thickness of cohesive soil layer, fluctuation degree of bedrock and buried depth of bedrock. A new interpretation of the principal components leading to the occurrence of ground fissures is given. At the same time, the Support Vector Machine (SVM) method is introduced to establish the risk prediction model of ground fissures in Suzhou-Wuxi-Changzhou area based on Principle Component Analysis (PCA) and Support Vector Machine (SVM). Combined with practical work, the predicted results are compared with the predicted results of SVM model. The results show that the risk prediction model of ground fissures based on Principal Component Analysis and Support Vector Machine has high accuracy and certain reference value. It can provide effective basis for predicting the risk of ground fissures in Suzhou-Wuxi-Changzhou area.

Key words: ground fissures; principle component analysis; support vector machine; Suzhou-Wuxi-Changzhou; risk prediction