

王帅伟,于少将,李绍康,等.基于 RS-PCA-GA-SVM 的砂土液化预测方法研究[J].地震工程学报,2019,41(2):445-453.doi:10.3969/j.issn.1000-0844.2019.02.445

WANG Shuaiwei, YU Shaojiang, LI Shaokang, et al. A Method of Predicting Sand Liquefaction Based on RS-PCA-GA-SVM[J]. China Earthquake Engineering Journal, 2019, 41(2): 445-453. doi: 10.3969/j.issn.1000-0844.2019.02.445

基于 RS-PCA-GA-SVM 的砂土液化预测方法研究

王帅伟¹, 于少将², 李绍康³, 袁颖²

(1. 中国地质科学院水文地质环境地质研究所, 河北 石家庄 050061; 2. 河北地质大学 勘查技术与工程学院, 河北 石家庄 050031; 3. 中国环境科学研究院, 北京 100012)

摘要:砂土液化是一种危害性比较大的自然灾害,对砂土液化进行判定预测在地质灾害防治领域中有重要的研究意义。通过粗糙集理论(Rough Set,RS)对影响砂土液化的6个初始评价指标(包括震级、土深、震中距、地下水位、标贯击数和地震持续时间)进行属性约简,去掉冗余或干扰信息,得到基于4个核心预测指标的数据集。通过主成分分析法(Principal Component Analysis,PCA)从核心评价指标中提取出主成分,采用支持向量机(Support Vector Machine,SVM)对数据集进行训练,用遗传算法(Genetic Algorithm,GA)优化参数,建立砂土液化的RS-PCA-GA-SVM预测模型。并结合砂土液化实际数据将预测结果与基于Levenberg-Marquardt算法改进的BP神经网络模型(LM-BP)的预测结果做比较。实例计算表明:基于RS-PCA-GA-SVM模型得到的砂土液化预测结果精度较LM-BP神经网络有很大的提高,判别结果与实际情况比较吻合,可在实际工程中应用。

关键词:砂土液化;粗糙集;遗传算法;主成分分析;支持向量机;预测模型

中图分类号: TU43

文献标志码: A

文章编号: 1000-0844(2019)02-0445-09

DOI: 10.3969/j.issn.1000-0844.2019.02.445

A Method of Predicting Sand Liquefaction Based on RS-PCA-GA-SVM

WANG Shuaiwei¹, YU Shaojiang³, LI Shaokang³, YUAN Ying²

(1. Institute of Hydrogeology and Environmental Geology, Chinese Academy of Geological Sciences Shijiazhang 050061, Hebei, China;

2. School of Prospecting Technology & Engineering, Hebei GEO University, Shijiazhuang 050031, Hebei, China;

3. Chinese Research Academy of Environmental Sciences, Beijing 100012, China)

Abstract: Sand liquefaction is a harmful natural disaster, and it is of great importance to evaluate and predict sand liquefaction in the field of geological disaster prevention and control. In this paper, the rough set theory (RS) was used to perform attribute reduction on six initial evaluation indices, including magnitude, depth of soil, epicentral distance, groundwater level, standard penetration test blow count, and earthquake duration, all of which affect sand liquefaction. After removing redundant or interference information, we obtained a data set based on four core predictors. The principal component analysis (PCA) method was then used to extract the principal component from the four-core evaluation indices. The support vector machine (SVM) was used to

收稿日期: 2017-10-21

基金项目: 国家自然科学基金(41301015); 河北省教育厅重点项目(ZD2015073, ZD2016038)

第一作者简介: 王帅伟,男,博士研究生,研究方向为地质灾害防治设计、工程地质和水文地质。E-mail: 626792005@qq.com。

通信作者: 袁颖(1970-),男,教授,研究方向为地质灾害防治设计和工程地质。E-mail: yuanyingson@163.com。

train the data set, and the genetic algorithm (GA) was used to optimize the parameters. Finally, the RS-PCA-GA-SVM prediction model for sand liquefaction was established. Combined with the actual data of sand liquefaction, the predicted result of the proposed model was compared with that of the back propagation (BP) neural network model based on the improved Levenberg-Marquardt algorithm (LM-BP). The calculated results showed that the accuracy of sand liquefaction prediction results based on a RS-PCA-GA-SVM model are much better than those of the LM-BP neural network. The discriminant results were in good agreement with the actual situation and can be applied in practical engineering.

Keywords: sand liquefaction; rough set; genetic algorithm; principal component analysis; support vector machine; forecast model

0 引言

砂土液化是指饱和状态砂土在一定强度的动荷载作用下表现出类似液体的性质,完全丧失强度和刚度的现象^[1]。砂土液化容易引起地基承载力降低,形成地面沉陷、高速滑坡、侧移、冒水喷砂等地质灾害,给人民生命和财产带来重大的损失。因此判定砂土液化产生的影响因素,建立合理的预测模型,对于城市规划、建筑场地选择以及制定液化区建筑物防护措施有着重要的现实意义。经过多年的理论与试验研究,国内外学者分别采用了灰色综合评判法^[2]、Fisher 判别法^[3]、模糊综合评判法^[4]和人工神经网络^[5]等对砂土液化进行判别^[6]。但这些方法均存在一定的局限性:采用灰色理论方法预测得到的结果容易被原始数据序列波动性与信息分散程度干扰;模糊综合评判法中指标权重的确定具有较大的主观性;人工神经网络存在学习率不稳定,容易陷入局部最小的缺点。选取判别方法时应充分考虑砂土液化的可变性、不确定性、没有原型、数据有限和影响因素复杂多样等特点。

支持向量机(SVM)是 Vapnik 等于 1995 年首次提出的一种新的机器学习方法,该方法可以很好地处理小样本、非线性、高维数和局部极小点等问题。近几年不少学者采用 SVM 方法对砂土液化进行预测,均取得了较好的成果^[7-11]。本文以搜集到的 40 个实际地震现场勘探资料为样本,以砂土液化的评价指标为依据,先后引入粗糙集理论(RS)和主成分分析理论(PCA)进行数据处理,并应用遗传算法(GA)确定支持向量机参数,建立基于 RS-PCA-GA-SVM 的砂土液化预测模型。用该模型对预测样本的液化分类进行预测,以期将预测结果与 LM-BP 神经网络模型的预测结果对比,以期为预测砂土液化分类提供新的思路。

1 模型原理

1.1 粗糙集理论

粗糙集理论是 Pawlak 教授于 1982 年提出的一种处理不确定性知识的数学理论。其主要思想是用已知系统中的信息来(近似)表示不精确或不确定因素的信息或数据。该理论不仅能有效地分析处理不确定因素和不完备信息,还可以通过分析和推理数据,找到信息间隐含的知识和潜在的规律^[12-14]。其具体内容如下:

设知识系统 $S = (U, A, V, f)$, 其中 U 是对象的非空有限集合,即论域; A 是属性的非空有限集合; $V = \bigcup_{a \in A} V_a$, V_a 是属性 a 的值域; $f: U \times A \rightarrow V$ 代表信息函,为每个对象指定属性值。对于信息系统 S ,若 $A = C \cup D$ (C 为条件属性, D 为决策属性),且 $C \cap D = \phi$, 则该知识系统为决策系统或决策表。决策属性支持度为:

$$k = \gamma_c(D) = \frac{|POS_c(D)|}{|U|}, (0 \leq k) \quad (1)$$

决策属性 D 依赖于条件属性 C , $POS_c(D)$ 为 D 关于 C 的正域, $\gamma_c(D)$ 为 D 对 C 的依赖度。当 $k = 1$ 时, D 完全依赖于 C ; 当 $0 < k < 1$ 时, D 粗糙依赖于 C ; 当 $k = 0$ 时, D 独立于 C 。

去掉条件属性集对决策属性集的支持度为:

$$\gamma_{C-C_i}(D) = \frac{|POS_{C-C_i}(D)|}{|U|} \quad (2)$$

属性子集 $C_i \subseteq C$ 关于 D 的重要程度为:

$$\sigma_{CD}(C_i) = \gamma_C(D) - \gamma_{C-C_i}(D) \quad (3)$$

其中: $\sigma_{CD}(C_i)$ 为在条件属性中,属性子集 C_i 被去掉后对于决策属性集的影响程度。 $\sigma_{CD}(C_i)$ 越大,说明该属性子集 C_i 对决策属性越重要。

对砂土液化预测决策表约简的主要思想是在确保对砂土液化预测模型正确分类的前提下,剔除于

决策属性无关紧要的冗余属性,并构造一个具有最优属性集的决策表。本文在砂土液化预测中,将收集到的 40 个砂土液化数据集合作为论域 $U = \{U_1, U_2, \dots, U_{40}\}$,砂土液化的影响指标作为条件属性 $C = \{X_1, X_2, \dots, X_{40}\}$,决策属性 $D = \{d_1, d_2\}$,其中, $d_1 =$ 砂土液化, $d_2 =$ 砂土未液化。

1.2 主成分分析

主成分分析方法的基本思想是设法将原有变量重新组合成一组新的线性无关的几个综合变量,同时根据实际需要,按照累计贡献率大小从中取出几个综合变量,使其尽可能多地反映原来变量的信息^[15]。其具体步骤如下:

设 (X_1, X_2, \dots, X_n) 是总体 $X_{m \times n}$ 的 n 个样本,每个样本为 m 维变量,则数据矩阵可记为:

$$X_{m \times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (4)$$

(1) 首先为了避免不同指标间量纲的差异性影响,将原始数据进行标准化。

$$zx_{ij} = \frac{x_{ij} - \bar{x}_{ij}}{\sqrt{\text{var}(x_j)}} \quad (5)$$

$$\text{var}(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (6)$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (7)$$

(2) 然后计算该矩阵的协方差阵,并根据公式 $\rho_{ij} = \text{cov}(X_i, X_j) / \sqrt{\sigma_{ii}\sigma_{jj}}$ 将协方差阵化其对应的相关系数阵。

(3) 求出相关系数阵的特征值,并按特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ 由大到小的顺序,计算出对应的正交单位化特征向量 e_1, e_2, \dots, e_m 。

(4) 确定 m 维变量的第 i 个主成分为 $Y_i = e_i^T X = e_{1i}x_1 + e_{2i}x_2 + \dots + e_{mi}x_m$,该主成分的贡献率为 λ_i/P ,其中 $i=1, 2, 3, \dots, m, P = \sum_{i=1}^m \lambda_i$,前 n 个主成分的累计贡献率为 $\sum_{i=1}^q \lambda_i/P$ 。

(5) 在实际应用中常略去那些贡献率小的主成份,若前 n 个主成份的累计贡献率已超过 85%,则只需取前 n 个主成分个数作为 SVM 模型的输入个数。

1.3 支持向量机

支持向量机是一种基于结构风险最小化原则,

以构造最优超平面为目标的统计学习方法^[16]。其基本思想为:通过非线性映射,把输入样本映射到高维的特征空间中,使得在原来的样本空间中非线性可分转化为在特征空间中的线性可分。该方法在处理小样本、非线性和高维数据的统计学习问题上有很好的推广能力。

设:训练样本 $\{X_i, y_i\}_{i=1}^m$,其中 $X_i \in R^n, y_i \in R, R$ 为实数域。以 $\phi(X_i)$ 表示映射到高维空间中的样本,以 ϕ 表示对应的非线性映射,则构造得到的函数表达式为:

$$y_i = W \cdot \phi(X_i) + b \quad (8)$$

其中: W 称为可调的权值向量, b 为偏置值,且 W 和 $\phi(X_i)$ 均为 n 维向量。寻找最优的分类超平面亦即寻找最优的 w 和 b 。考虑到允许存在拟合误差,因此引入松弛变量 ξ 和 ξ^* ,依据结构风险最小化准则,采用 ϵ -SVR 模型建立带有约束条件的模型优化函数:

$$\begin{aligned} \min_{w, b, \xi, \xi^*} & \frac{1}{2} W^T W + C \sum_{i=1}^m (\xi_i + \xi_i^*) \\ \text{s.t.} & \begin{cases} y_i - W^T \phi(X_i) - b \leq \epsilon + \xi_i^* \\ W^T \phi(X_i) + b - y_i \leq \epsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (9)$$

$$(i=1, \dots, m)$$

其中式(9)是一个凸二次优化问题,采用 Lagrange 乘子法进行求解,建立 Lagrange 函数:

$$\begin{aligned} L(W, b, \xi, \xi^*, a, a^*, r, r^*) &= \frac{1}{2} W^T W + \\ & C \sum_{i=1}^m (\xi_i + \xi_i^*) - \sum_{i=1}^m a_i^* [y_i - W^T \cdot \phi(X_i) - \\ & b - \epsilon - \xi_i^*] - \sum_{i=1}^m a_i [W^T \cdot \phi(X_i) + \\ & b - y_i - \epsilon - \xi_i] - \sum_{i=1}^m r_i - \sum_{i=1}^m r_i^* \end{aligned} \quad (10)$$

式中: C 为惩罚系数; a_i, a_i^*, r_i 和 r_i^* 都为 Lagrange 乘子。

将式(10)分别对 W 和 b 求偏导并置零,反代回式(10)中就可以得到上述问题的对偶问题,找到对应的 KKT 条件,并使用二次规划优化算法作为训练算法依次计算,得到参数 a_i 和 a_i^* 对应的最优乘子 a_i^{new} ,同时构造得到预测函数:

$$f(X_i) = \sum_{i=1}^m (a_i^* - a_i) \phi(X_i)^T \phi(X_i) + b, \quad (11)$$

式中: $f(X_i)$ 为本文预测的砂土液化情况; a_i^* 与 a_i 为 Lagrange 乘子; $\phi(X_i)^T \phi(X_i)$ 为核函数; X_i 为

评价指标(震级、土深、震中距、地下水位、标贯击数和地震持续时间等6个影响因素)。

2 基于RS-PCA-GA-SVM砂土液化预测

2.1 砂土液化评价指标的确定

砂土液化的影响因素很多,归纳起来包括动荷条件、埋藏条件和土性条件三个类别^[17]。其中动荷条件主要指地震的强度与持续时间;埋藏条件指的

是埋深与地下水位;土性条件指的是砂土层自身因素,如土的种类、颗粒组成、密实度等。本文依据Seed提供的13次大地震宏观调查资料^[18],选取震级、土深、震中距、地下水位、标贯击数和地震持续时间共6个影响因素作为砂土液化的评价指标,并选择34组数据作为支持向量机预测模型的训练样本(表1),6组数据作为支持向量机的测试样本(表2),其中1代表砂土液化,2代表砂土未液化。

表1 砂土液化训练样本

Table 1 Training samples of sand liquefaction

序号	震级 M	土深/m	震中距/km	标贯击数/次	地下水位/m	持续时间/s	液化分类
1	8.4	9	51.5	10	2	75	1
2	8.4	4	51.5	10	2.5	75	1
3	6.3	7.5	17.7	3	4.5	15	1
4	7	4.5	12.9	9	4.5	30	1
5	7	7.5	12.8	4	6	30	1
.....							
.....							
32	7.8	3.5	277	14	1	45	2
33	6.1	6	75.7	6	1	12	2
34	7.5	6	83.7	12	1	40	2

表2 砂土液化预测样本

Table 2 Predicting samples of sand liquefaction

序号	震级 M	土深/m	震中距/km	标贯击数/次	地下水位/m	持续时间/s	液化分类
1	7	6	12.9	1	1.5	30	1
2	7.2	3	9.7	3	1	30	1
3	8.4	5	51.5	4	1	75	1
4	7.5	5	83.7	6	1	40	1
5	6.1	6	75.7	12	1	12	2
6	7.8	6	277	15	1.5	45	2

2.2 基于粗糙集理论的信息分析

2.2.1 属性离散化

属性离散化是粗糙集进行属性约简前重要的数据处理过程。在粗糙集理论中,连续属性离散化的目的是在尽量减少决策系统信息损失的前提下简化、浓缩决策系统,并分析和获得决策所需要的知识^[19]。经比较分析,本文采用布尔逻辑离散化方法(boolean reasoning algorithm)对原始数据(表1与表2)进行离散化处理,得到各条件属性及其离散化值域(表3)。

2.2.2 属性约简

属性约简是粗糙集理论中一个重要的研究课题。一般来说,知识库中每个知识的重要程度是不一样的,冗余的知识属性不利于做出正确而简洁的决策。属性约简的主要思想是在确保决策分类准确性的前提下尽可能消除更多不相关的、冗余的属性指标,找到个数最少且最小的核心指标,达到最

优约简的效果,从而提高支持向量机的运算效率。Wong等^[20]曾经证明过完成某个决策信息系统的最优约简是一个NP-hard问题。但在同一个决策系统中,不同算法的约简结果是不同的。表4给出了Genetic Algorithm(遗传算法)和Johnson's Algorithm两种算法的约简结果。

两种算法的交集为核心指标,即{震级,震中距,标贯击数},这三个核心指标不能剔除,否则将大大减弱分类能力。通过比较由遗传算法得到的三组约简结果可知,剩余三个指标的重要性排序为:地下水位>土深=持续时间,因此本次属性约简剔除了冗余指标(土深、持续时间),得到震级、震中距、标贯击数、地下水位4个关键指标。

2.3 砂土液化的主成分分析

对属性约简后的影响因素进行主成分分析,变量 x_1, x_2, \dots, x_4 分别代表砂土液化的核心指标震级、震中距、标贯击数和地下水位,计算各变量之间

的相关系数,得到相关系数矩阵 $R = (r_{i,j})_{4 \times 4}$ (表 5)。

求解 R 的特征值 λ_i ,并根据 λ_i 计算得到贡献率与累计贡献率(表 6)。

表 3 砂土液化数据离散化决策表

Table 3 Discretization decision table of sand liquefaction data

序号	震级 M	土深/m	震中距/km	标贯击数/次	地下水位/m	持续时间/s	液化分类
1	[7.7, *)	[5.9, *)	[*, 59.2)	[*, 12)	[1.3, *)	[43, *)	1
2	[7.7, *)	[*, 5.9)	[*, 59.2)	[*, 12)	[1.3, *)	[43, *)	1
3	[*, 6.8)	[5.9, *)	[*, 59.2)	[*, 12)	[1.3, *)	[*, 43)	1
4	[6.8, 7.7)	[*, 5.9)	[*, 59.2)	[*, 12)	[1.3, *)	[*, 43)	1
5	[6.8, 7.7)	[5.9, *)	[*, 59.2)	[*, 12)	[1.3, *)	[*, 43)	1
.....							
32	[7.7, *)	[5.9, *)	[*, 59.2)	[14, *)	[1.3, *)	[43, *)	2
33	[6.8, 7.7)	[5.9, *)	[*, 59.2)	[14, *)	[*, 1.3)	[*, 43)	2
34	[7.7, *)	[5.9, *)	[59.2, *)	[14, *)	[1.3, *)	[43, *)	2
35	[7.7, *)	[5.9, *)	[59.2, *)	[14, *)	[*, 1.3)	[43, *)	2
36	[7.7, *)	[*, 5.9)	[59.2, *)	[14, *)	[*, 1.3)	[43, *)	2
37	[*, 6.8)	[5.9, *)	[59.2, *)	[*, 12)	[*, 1.3)	[*, 43)	2
38	[6.8, 7.7)	[5.9, *)	[59.2, *)	[12, 14)	[*, 1.3)	[*, 43)	2
39	[*, 6.8)	[5.9, *)	[59.2, *)	[12, 14)	[*, 1.3)	[*, 43)	2
40	[7.7, *)	[*, 5.9)	[59.2, *)	[14, *)	[1.3, *)	[43, *)	2

表 4 Genetic Algorithm 与 Johnson 算法的属性约简

Table 4 Attribute reduction results of Genetic algorithm and Johnson algorithm

序号	算法	约简结果	Support	Length
1		{震级,震中距,标贯击数}	100	3
2	Genetic Algorithm	{震级,土深,标贯击数,地下水位}	100	4
3		{震中距,标贯击数,地下水位,持续时间}	100	4
4	Johnson's Algorithm	{震级,震中距,标贯击数}	100	3

表 5 砂土液化评价指标相关系数矩阵

Table 5 Correlation coefficient matrix of sand liquefaction's evaluation indexes

	震级	震中距	标贯击数	地下水位
震级	1.000	0.453	0.135	0.202
震中距	0.453	1.000	0.025	0.331
标贯击数	0.135	0.025	1.000	0.157
地下水位	0.202	0.331	0.157	1.000

表 6 特征值、贡献率和累计贡献率

Table 6 Eigenvalue, contribution rate, and cumulative contribution rate

主成分	特征值	贡献率/%	累计贡献率/%
F_1	1.67	41.72	41.72
F_2	1.13	28.32	70.05
F_3	0.68	16.95	87.00
F_4	0.52	13.00	100.00

由表 8 可以看出前三个因子的贡献率为 87.00% > 85%,即这 3 个因子能够代表原始 4 个核心指标 87.00% 的信息量,能够比较全面地反映砂土液化的所有信息。因此选择 F_1 、 F_2 和 F_3 为主成分进行 SVM 训练,且其与各评价指标的关系式为(5)~(7),其中 $z x_i$ 为砂土液化标准化后的数据:

$$F_1 = 0.588 \ 83z x_1 + 0.779 \ 69z x_2 + 0.043 \ 71z x_3 - 0.877 \ 22z x_4 \quad (12)$$

$$F_2 = 0.048 \ 62z x_1 + 0.007 \ 50z x_2 + 0.217 \ 46z x_3 + 0.133 \ 36z x_4 \quad (13)$$

$$F_3 = 0.053 \ 87z x_1 + 0.017 \ 33z x_2 - 0.047 \ 39z x_3 + 0.058 \ 35z x_4 \quad (14)$$

2.4 支持向量机参数优化的遗传算法

支持向量机砂土液化预测模型惩罚因子 C 和核函数参数 γ 是支持向量机的主要参数,核函数类型与核函数参数的不同将直接影响到支持向量机的学习效率与运算能力。RBF 核函数处理多元问题具有非线性映射能力强、效率高,且需要确定的参数少等优点。因此本文选用 RBF 核函数来对支持向量机模型进行训练。其公式如下:

$$K_g(x, y) = \exp(-g \|x - y\|^2) \quad (15)$$

核函数参数 g 和惩罚参数 C 的常用优化方法有网格搜索、粒子群算法和遗传算法等^[21-23]。本文采用遗传算法确定它们的取值。遗传算法是一种基于自然界的选择与遗传的机理来寻找最优解的进化算法。相比于其他传统方法,该算法具有很好的收敛

性,计算时间少,具有可扩展性,且容易与其他算法结合。其流程见图1。

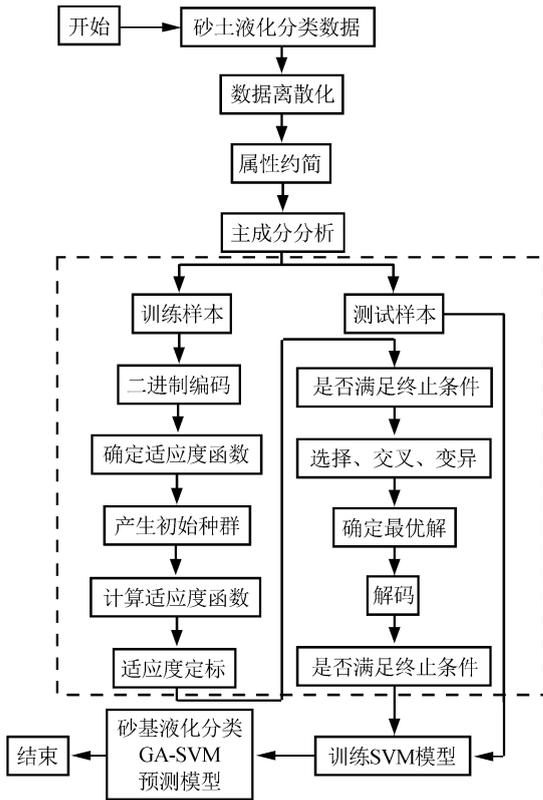


图1 利用遗传算法优化 SVM 参数的算法流程图
Fig.1 Flow chart for optimizing SVM parameters by genetic algorithm

选取表5和表6样本中粗糙集属性约简后的数据对砂土液化样本进行训练和分类预测。遗传算法的各参数设置为:种群数20、迭代数100、交叉率0.8和变异率0.2。通过多次遗传寻优迭代,得到最优的遗传算法的适应度曲线(图2)。SVM的最优参数确定为 $g = 4.6263$, $C = 3.0145$, 均方误差(MSE) = 0.0927。用该 g 和 C 参数确定的 RS-PCA-GA-SVM 模型对表5中的训练样本进行回归仿真,其结果如图3所示。

由图3可以看出,基于RS-PCA-GA-SVM砂土

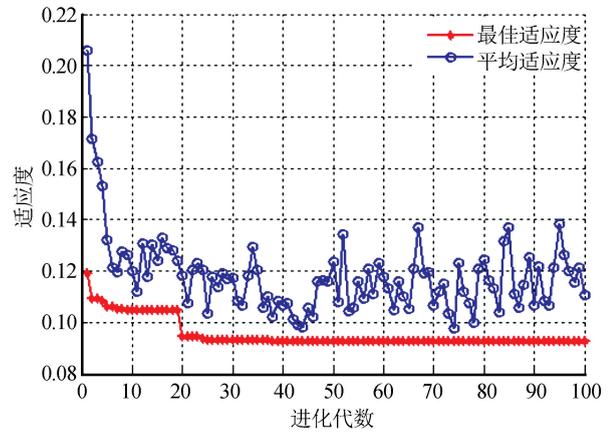


图2 遗传算法优化过程

Fig.2 Optimization process of genetic algorithm

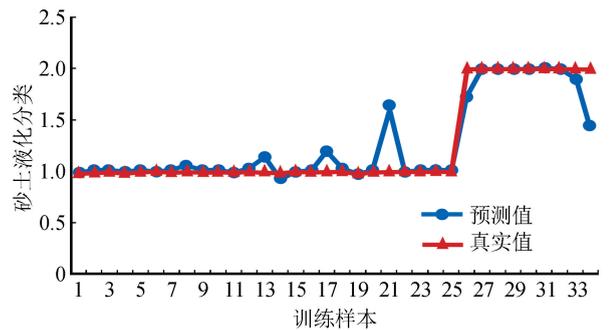


图3 训练样本回归预测结果

Fig.3 Regression prediction results of training samples

液化预测模型对训练样本的回归仿真结果与实际值极为接近,可以达到砂土液化预测精度要求。

3 结果及其分析

通过建立的 RS-PCA-GA-SVM 砂土液化预测模型对表6中的6个实例进行分类预测,得到砂土液化的分类结果。BP神经网络模型能够实现复杂的非线性映射功能,非常适合解决内部机制复杂的各类问题。因此本文对相同的训练样本建立基于 Levenberg-Marquardt 优化算法的 BP神经网络模型(LM-BP),对砂土液化测试样本进行预测,各模型预测结果见表7、8与图4、5。

表7 SVM模型预测结果及相对误差

Table 7 Prediction results and relative errors of SVM models

序号	真实值	GA-SVM		RS-GA-SVM		RS-PCA-GA-SVM	
		预测值	相对误差	预测值	相对误差	预测值	相对误差
1	1	0.877	12.33%	1.059	5.93%	1.053	5.34%
2	1	0.905	9.54%	1.004	0.40%	0.997	0.33%
3	1	0.925	7.55%	0.959	4.07%	0.972	2.84%
4	1	1.210	21.02%	1.056	5.56%	1.035	3.50%
5	2	2.068	3.40%	2.135	6.75%	2.078	3.90%
6	2	2.057	2.86%	2.117	5.85%	2.092	4.60%

表 8 LM-BP 模型预测结果及相对误差

Table 8 Prediction results and relative errors of LM-BP models

序号	真实值	LM-BP		RS-LM-BP		RS-PCA-LM-BP	
		预测值	相对误差	预测值	相对误差	预测值	相对误差
1	1	1.125	12.52%	1.040	4.02%	0.949	5.07%
2	1	1.207	20.70%	1.041	4.09%	0.962	3.75%
3	1	0.997	0.33%	0.818	18.17%	1.059	5.91%
4	1	1.239	23.88%	1.161	16.14%	1.137	13.72%
5	2	1.802	9.90%	2.164	8.18%	1.800	9.98%
6	2	1.690	15.48%	2.141	7.07%	1.771	11.46%

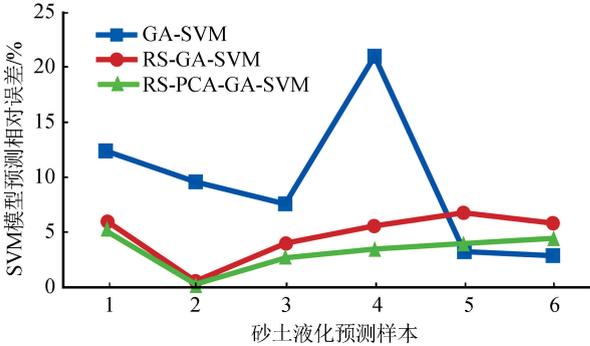


图 4 SVM 模型预测结果相对误差

Fig.4 Relative errors of SVM models' prediction results

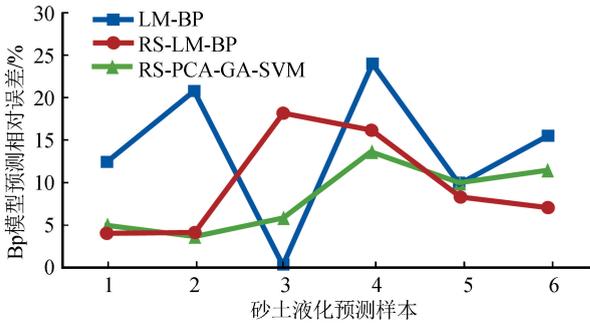


图 5 LM-BP 模型预测结果相对误差

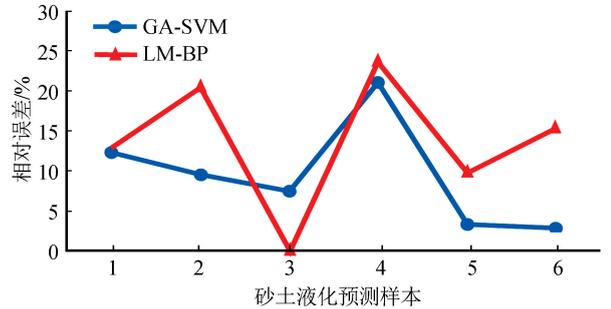
Fig.5 Relative error of LM-BP models' prediction results

表 9 与图 4 为属性约简前、属性约简后和属性约简后并进行主成分分析的三种 SVM 模型得到的预测结果。表 10 与图 5 为经过相同方法进行数据处理的 LM-BP 模型得到的预测结果。从图与表可知,在不约简情况下,SVM 模型与 LM-BP 模型评价结果较差,最大相对误差均超过了 20%。而约简后,两种模型的预测精度有了明显提升,最大相对误差分别降到 6.75%与 18.17%。同时对属性约简后的数据进行主成分分析,发现两种模型在运算效率与预测精度上又有所提高,其中经过 RS-GA-PCA-SVM 模型得到的预测结果准确率最高,最大相对误差仅为 5.34%。

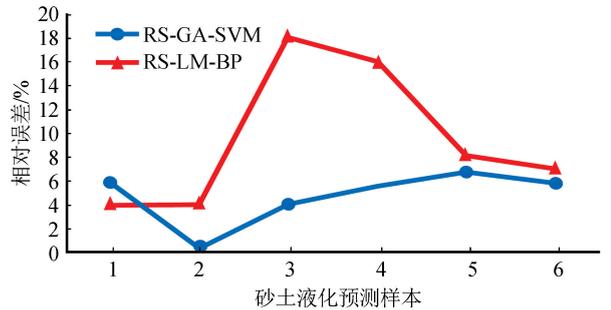
综合以上分析可知,经过 RS 属性约简后,SVM 与 LM-BP 预测模型得到了影响砂土液化的关键性指标(震级、震中距、标贯击数和地下水位),不仅提

高了砂土液化的预测精度,并在运算中降低了模型计算复杂度,提高了训练速度,减少了运算时间。同时以 RS 属性约简为基础,采用 PCA 对多维、互相关的砂土液化属性数据进行主成分提取,使预测模型的输入变量得到优化,消除了评价指标之间的相关影响,提高了两种模型的预测实时性与预测精度。

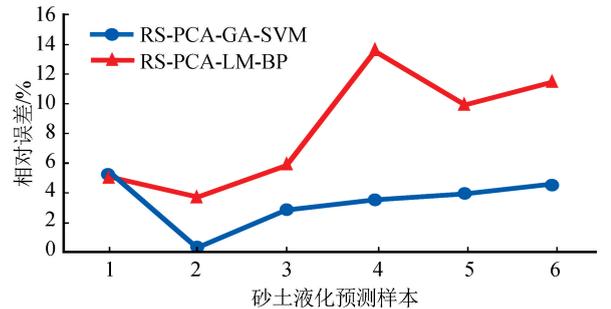
图 6 为不同数据处理阶段 SVM 与 LM-BP 模型



(a) GA-SVM与LM-BP模型



(b) RS-GA-SVM与RS-LM-BP模型



(c) RS-PCA-GA-SVM与RS-PCA-LM-BP模型

图 6 不同数据阶段模型的预测结果相对误差

Fig.6 Relative errors of different models' prediction results at different data phase

的预测结果对比。由图可知,SVM模型的预测效果更符合实际,而LM-BP模型的预测结果整体相对误差较大且预测效果不稳定,该模型受样本训练数目影响很大,对样本的依赖性较强,尤其对小样本统计学习泛化能力较差。

4 结论

(1) 震级、震中距、标贯击数和地下水位是影响砂土发生液化的关键性指标。

(2) 影响砂土液化的因素众多,本文基于RS-PCA-GA-SVM的支持向量机预测方法,根据有限的训练样本建立了砂土液化与影响因素之间非线性映射关系,不仅解决了各指标之间高维非线性的复杂关系问题,还提高了砂土液化预测的准确性与运算效率。

(3) 通过引入粗糙集理论对数据进行属性约简,剔除了信息的交叉和冗余,降低了SVM输入向量维数和计算复杂度,减少了训练时间。采用主成分分析处理数据,消除指标之间的相关影响。利用遗传算法对支持向量机的参数进行优化,避免了人工选择SVM参数的盲目性,解决了支持向量机方法无法实现选择内积核函数参数的问题。同时与LM-BP模型对比,该模型具有更强的外推能力和预测计算能力,为很多用传统的数学、力学方法难给出确定模型的领域中提供了新的思路。

参考文献(References)

[1] 高大钊,袁聚云.土质学与土力学[M].北京:人民交通出版社,2001:216-222.
GAO Dazhao, YUAN Juyun. Soil Science and Soil Mechanics [M]. Beijing: China Communications Press, 2001: 216-222.

[2] 李波,苏经宇,马东辉,等.地震砂土液化判别的灰色关联-逐步分析耦合模型[J].中南大学学报(自然科学版),2016,47(1):232-238.
LI Bo, SU Jingyu, MA Donghui, et al. Coupling Model Based on Grey Relational Analysis and Stepwise Discriminant Analysis for Seismic Liquefaction Discrimination of Sandy Soil[J]. Journal of Central South University (Science and Technology), 2016, 47(1): 232-238.

[3] 刘年平,王宏图,袁志刚,等.砂土液化预测的Fisher判别模型及应用[J].岩土力学,2012,33(2):554-546.
LIU Nianping, WANG Hongtu, YUAN Zhigang, et al. Fisher Discriminant Analysis Model and Its Application to Sand Liquefaction Prediction[J]. Hydrogeology & Engineering Geology, 2012, 33(2): 554-546.

[4] 祝百茹,刘海卿.基于模糊综合评判的砂土液化判别[J].广西大学学报(自然科学版),2013,38(1):144-150.

ZHU Bairu, LIU Haiqing. Fuzzy Comprehensive Evaluation on Liquefaction of Sand Soil[J]. Journal of Guangxi University (Nat Sci Ed), 2013, 38(1): 144-150.

[5] 高宗军,付青,郑秋霞,等.BP和Elman神经网络在砂土液化预测中的研究[J].中国安全生产科学技术,2013,9(6):58-62.
GAO Zongjun, FU Qing, ZHENG Qiuxia, et al. Study on Forecasting of Sand Liquefaction by Using BP Neural and Elman Neural Networks[J]. Journal of Safety Science and Technology, 2013, 9(6): 58-62.

[6] 李振林.砂土地震液化判别方法的综合应用[J].西安科技大学学报,2010,30(4):451-456.
LI Zhenlin. Discrimination Methods for Sandy Soil Seismic Liquefaction[J]. Journal of Xi'an University of Science and Technology, 2010, 30(4): 451-456.

[7] 夏建中,罗战友,龚晓南,等.基于支持向量机的砂土液化预测模型[J].岩石力学与工程学报,2005,24(22):4139-4144.
XIA Jianzhong, LUO Zhanyou, GONG Xiaonan, et al. Support Vector Machine Model for Predicting Sand Liquefaction[J]. Chinese Journal of Rock Mechanics and Engineering, 2005, 24(22): 4139-4144.

[8] 师旭超,郭志涛,韩阳.基于支持向量机的砂土液化预测分析[J].地震工程学报,2009,31(4):363-366.
SHI Xuchao, GUO Zhitao, HAN Yang. Analysis on Sand Seismic Liquefaction Prediction Based on the Support Vector Machine[J]. China Earthquake Engineering Journal, 2009, 31(4): 363-366.

[9] 张向东,冯胜洋,王长江.基于网格搜索的支持向量机砂土液化预测模型[J].应用力学学报,2011,28(1):24-28.
ZHANG Xiangdong, FENG Shengyang, WANG Changjiang. Support Vector Machine Model for Predicting Sand Liquefaction Based on Grid-Search Method[J]. Chinese Journal of Applied Mechanics, 2011, 28(1): 24-28.

[10] WANG F, SU J, WANG Z. Forecasting of Building Settlements Due to Earthquake Liquefaction Based on LS-SVM with Mixed Kernel[J]. Electronic Journal of Geotechnical Engineering, 2015, 20(1): 11-19.

[11] FAN Z J, LENG Y Q, XU Y L, et al. A Discrimination Method of Saturated Sand Liquefaction Possibility Based on Support Vector Machine[J]. Applied Mechanics & Materials, 2014, 509: 38-43.

[12] PAWLAK Z. Rough Set[J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356.

[13] 高爽,冬雷,高阳,等.基于粗糙集理论的中长期风速预测[J].中国电机工程学报,2012,32(1):32-37.
GAO Shuang, DONG Lei, GAO Yang, et al. Mid-long Term Wind Speed Prediction Based on Rough Set Theory[J]. Proceedings of the CSEE, 2012, 32(1): 32-37.

[14] 张文修,吴伟志.粗糙集理论介绍和研究综述[J].模糊系统与数学,2000,14(4):1-12.
ZHANG Wenxiu, WU Weizhi. An Introduction and a Survey for the Studies of Rough Set Theory[J]. Fuzzy Systems and Mathematics, 2000, 14(4): 1-12.

- [15] 张国英,王娜娜,张润生,等.基于主成分分析的 BP 神经网络在岩性识别中的应用[J].北京石油化工学院学报,2008,16(3):43-46.
ZHANG Guoying, WANG Nana, ZHANG Runsheng, et al. Application of Principal Component Analysis and BP Neural Network in Identifying Lithology[J].Journal of Beijing Institute of Petro-chemical Technology,2008,16(3):43-46.
- [16] VAPNIK V N. The Nature of Statistical Learning Theory [M].New York:SpringVerlag,1995.
- [17] 杨健,陈庆寿.砂土液化影响因素及其判别方法[J].西部探矿工程,2004,93(2):1-2.
YANG Jian, CHEN Qingshou. Affecting Factors and Distinguishing Methods of Sandy Soil Liquefaction[J]. West-China Exploration Engineering,2004,93(2):1-2.
- [18] 蔡煜东,宫家文.砂土液化预测的人工神经网络模型[J].岩土工程学报,1993,15(6):53-58.
CAI Yudong, GONG Jiawen. Artificial Neural Network Model for Prediction of Sand Liquefaction[J]. Chinese Journal of Geotechnical Engineering,1993,15(6):53-58.
- [19] 王举范,陈卓.基于信息熵的粗糙集连续属性多变量离散化算法[J].青岛科技大学学报(自然科学版),2013,34(4):423-426.
WANG Jufan, CHEN Zhuo. Multiple Variable Discretization Algorithm of Continuous Attributes in Rough Set Theory Based on Information Entropy[J].Journal of Qingdao University of Science and Technology (Natural Science Edition), 2013,34(4):423-426.
- [20] WONG S K M, ZIARKO W. On Optiongal Decision Rule in Decision Table [J]. Bulletin of Polish Academy of Science, 1985,33:693-696.
- [21] 郭超,宋卫华,魏威.基于网格搜索-支持向量机的采场顶板稳定性预测[J].中国安全科学学报,2014,24(8):31-36.
GUO Chao, SONG Weihua, WEI Wei. Stope Roof Stability Prediction Based on Both SVM and Grid-search Method[J]. China Safety Science Journal,2014,24(8):31-36.
- [22] 谢玮,王彦春,刘建军,等.基于粒子群优化最小二乘支持向量机的非线性 AVO 反演[J].石油地球物理勘探,2016,51(6):1187-1194.
XIE Wei, WANG Yanchun, LIU Jianjun, et al. Non-linear AVO Inversion Based on PSO-LSSVM[J]. Oil Geophysical Prospecting,2016,51(6):1187-1194.
- [23] 季斌,周涛发,袁峰.遗传算法优化支持向量机矿产预测方法[J].测绘科学,2015,40(10):106-109.
JI Bin, ZHOU Taofa, YUAN Feng. The Detection Method of Maglev Gyroscope Abnormal Data Based on the Characteristics of Two Positioning[J]. Science of Surveying and Mapping,2015,40(10):106-109.