

DOI:10.16031/j.cnki.issn.1003-8035.2019.01.11

基于高斯过程分类的地震砂土液化预测及其应用

赵国彦,徐志炜,刘健

(中南大学资源与安全工程学院,湖南长沙 410083)

摘要:在砂土液化预测模型中,采用的数据一般来源单一且数据量少。本文选取了来自SPT场地、 V_s 场地、CPT场地的三个数据库,其数量分别为620、185、226例,每个数据库中影响因子的种类和个数都不相同。共有9种模型与高斯过程分类模型(GPC)进行综合对比分析,分别为FDA、DDA、BDA、MLR、ELM、LSSVM、SVM、CART、ANN。研究结果表明,GPC模型对每个数据库的液化判别都具有高准确率,分别为0.869 1、0.910 1、0.940 1,且其预测精度及稳健性均明显优于其他模型。由于数据来源的广泛性及代表性,可以得出GPC模型对于砂土液化预测具有极强的适应性。

关键词:砂土液化;高斯过程分类;对比分析;数据来源

中图分类号:P315

文献标识码:A

文章编号:1003-8035(2019)01-0093-07

Prediction method of seismic-induced sand liquefaction based on the Gauss Process classification

ZHAO Guoyan, XU Zhiwei, LIU Jian

(School of Resources and Safety Engineering, Central South University, Changsha, Hunan 410083, China)

Abstract: Many researches reveal that the data used in most present prediction models of sand liquefaction are simple and less. Three databases from SPT, VS, and CPT sites with numbers of 620, 185 and 226, respectively, were used in this paper. The numbers and types of influence factors in the three databases were different from each other. A total number of 9 models were compared with the Gauss Process Classification (GPC) model, as following: FDA, DDA, BDA, MLR, ELM, LSSVM, SVM, CART and ANN. The research results showed that the GPC model was accurate for the liquefaction of all databases, with accuracy rates of 0.869 1, 0.910 1 and 0.940 1, respectively. Moreover, the prediction accuracy and robustness of GPC model were significantly better than the other models. It concluded that the GPC model had a very strong adaptability to the prediction of sand liquefaction from the wide and representative data sources.

Keywords: sand liquefaction; Gaussian Process classification; comparative analysis; data sources

0 引言

砂土液化一般由地震诱发,能使土体完全丧失抗剪强度而处于“悬浮”状态,导致地基、房屋、道路、桥梁等地表结构物严重受损^[1]。比如,1964年的阿拉斯

加地震和新泻地震、1976年的唐山地震等均发生大面积因砂土液化现象,致使堤防和建筑物受到大规模的破坏,造成巨大经济损失,因而受到工程界广泛关注与深入研究^[2]。

目前,工程中常用的砂土液化预测方法主要是基

收稿日期:2018-04-03;修订日期:2018-04-24

基金项目:国家自然科学基金面上项目(51374244)

第一作者:赵国彦(1963-),男,湖南沅江人,博士,教授,主要从事采矿工程、矿山安全和岩石力学与工程等方面的研究工作。E-mail: 2469227190@qq.com

通讯作者:徐志炜(1992-),男,硕士,主要从事岩土工程方面的研究工作。E-mail:553966302@qq.com

于砂土液化真实案例的总结分析而建立起来的经验法,常见的有标准贯入法(SPT)、剪切波速法(V_s)及静力触探法(CPT),可以对相似场地进行液化判别^[3]。其中,以基于标贯值的 Seed 简化法应用最广泛,也是我国《建筑抗震设计规范》(GB 50011—2001)中建议的砂土液化判别方法的基础^[4]。影响砂土液化的因素很多,文献^[5]总结了 22 个,而传统的经验法选取的影响因素不够全面且某些参数的确定具有主观性,因此在不同地区砂土液化势判别中存在着较大的误差^[3]。近年来,随着计算技术、数学理论不断发展,各种高维数、非线性数学模型应用于砂土液化预测,比如人工神经网络^[3,4]、距离判别分析(DDA)^[6]、支持向量机(SVM)^[7]、Fisher 判别分析(FDA)^[8]、贝叶斯判别分析(BDA)^[9]等。但其数据来源均没有涵盖 SPT 场地、 V_s 场地和 CPT 场地,来源较单一。且文献^[3,6-8]的数据量非常少,这会对模型的适用性和可靠性产生一定质疑。本文选取的三个数据集来分别来自 SPT

场地、 V_s 场地以及 CPT 场地,均是传统经验法主要应用场地,极具代表性。实例数量分别为 622、185、226 例,以此最大限度消除数据对模型判别的影响。

高斯过程(Gaussian Process, GP)是一种基于贝叶斯框架的具有统计学意义的机器学习方法,对小样本、高维数、非线性等复杂问题具有很好的自适性,特别适用于砂土液化预测问题^[10]。选取 Fisher 判别分析(FDA)、距离判别分析(DDA)、贝叶斯判别分析(BDA)、Logistic Regression(多元逻辑回归)、极限学习机(ELM)、最小二乘支持向量机(LSSVM)、支持向量机(SVM)、CART(分类回归树)、人工神经网络(ANN) 9 种极具代表性的判别模型作为参照,与 GPC 模型进行对比分析。

1 高斯过程分类模型

GPC 是一种基于高斯过程原理的分类方法,其分类过程如图 1 所示^[11]。具体过程如下:

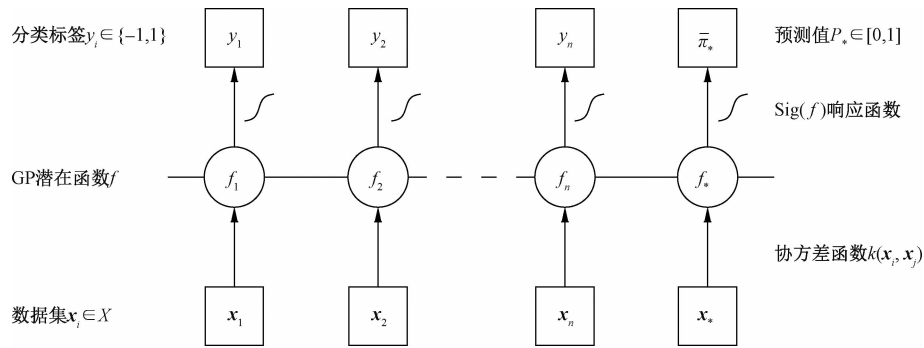


图 1 高斯过程二元分类模型在判别式框架中的图形描述

Fig. 1 The graphical representation of GPC in the discriminative framework

假定潜在函数 $f(x)$ 的先验分布为高斯分布,则

$$p(f | X) : N(f | m, K) \quad (1)$$

式中, K —— m 阶协方差矩阵,对称且正定。

$K_{ij} = k(x_i, x_j, \theta)$, $\theta = \{\theta_1, \theta_2\}$ 为正定协方差函数 $k(\cdot)$ 的超参数,可用极大似然法估计。

假设训练样本集为 $D = \{(x_i, y_i) | i = 1, \dots, n\}$, 输入数据集为 $X = [x_1, \dots, x_n]^T$, 输出值集合为 $y = [y_1, \dots, y_n]^T$, 潜在函数值 $f_i = f(x_i)$ 。类标签 y 是独立同分布的,在高斯过程二元分类中样本 x 属于类标签 y 的概率可以表示成:

$$p(y_i | f_i) = \text{Sig}(y f_i) \quad (2)$$

$\text{Sig}(\cdot)$ 函数的作用是将输出值转换为概率,以求得输入数据属于某类的概率。通常,称 $\text{Sig}(f)$ 为响应函数,其值在区间 $[0, 1]$ 上。由此,得到似然函数:

$$p(y | f) = \prod_{i=1}^n p(y_i | f_i) = \prod_{i=1}^n \text{Sig}(y f_i) \quad (3)$$

由贝叶斯公式可得到 f_* 后验概率:

$$p(f | X, y) = \frac{p(y | f) p(f | X)}{p(y | X)} \quad (4)$$

综上所述,对于指定数据 X_* ,则与其相对应的 f_* 的后验概率为:

$$p(f_* | X, y, X_*) = \int p(f_* | f, X, X_*) p(f | X, y) df \quad (5)$$

进一步推出 f_* 所对应的分类预测概率为:

$$p(y_* | X, x_*, y) = \int \text{Sig}(y_* f_*) p(f_* | X, y, x_*) df_* \quad (6)$$

在 GPC 模型中,常以 $p(y_* | X, x_*, y) = 0.5$ 为分类界限。当 $p > 0.5$ 时, $y_* = +1$,代表“液化”;反之,

$y_* = -1$, 代表“非液化”。GPC 模型的详细原理见文献[10]。

2 工程实例

2.1 砂土液化数据库

在本研究中,选取三个实例数据库用于构造和验证上述 10 种模型,分别命名为数据库一、二、三。

数据集一^[12]为 Hanna 等人在两大地震发生地(台湾,土耳其)记录的基于标准贯入值的 620 实例,其中

“液化”256 例,“非液化”364 例,涉及 12 个影响因素,其名称及统计描述见表 1。

数据集二^[13]为 Juang 和 Chen 总结的基于剪切波速的 185 个实例,其中“液化”107 例,非液化“78 例”,涉及 7 个影响因素,其名称及统计描述见表 2。

数据集三^[14]为 Goh 总结的基于静力触探值的 226 个实例,其中“液化”93 例,“非液化”133 例,涉及 6 个影响因素,其名称及统计描述见表 3。

表 1 砂土液化数据库一样本统计特性
Table 1 Discriminant results of the forecast sample NO. 1

影响因素	符号	描述说明	最小值	平均值	中位数	方差	最大值
a	D	土层埋深/m	0.80	7.66	6.70	4.89	19.80
b	$(N_1)_{60}$	标准贯入击数	1.00	14.48	11.00	11.38	75.00
c	$F \leq 75 \mu\text{m}$	黏粒含量/%	1.00	62.99	74.50	34.25	100.00
d	d_w	地下水位/m	0.35	1.45	1.10	1.20	10.00
e	σ_{t0}	竖向总应力/kPa	12.10	144.60	121.60	98.12	408.90
f	σ'_{t0}	竖向有效应力/kPa	7.50	82.48	68.15	52.80	233.70
g	a_t	加速度阈值 g	0.00	0.07	0.06	0.07	0.85
h	τ_{av}/σ'_{t0}	剪应力与竖向有效应力比	0.12	0.37	0.39	0.15	0.77
i	V_s	剪切波速/($\text{m}\cdot\text{s}^{-1}$)	37.00	166.98	155.00	67.03	500.00
j	ϕ'	内摩擦角/($^\circ$)	23.46	31.96	31.41	4.84	52.08
k	M_v	地震震级	7.40	7.49	7.40	0.10	7.60
l	$a_{\max}$	地表水平峰值加速度 g	0.18	0.38	0.40	0.15	0.67

表 2 砂土液化数据库二样本统计特性
Table 2 Discriminant results of the forecast sample NO. 2

影响因素	符号	描述说明	最小值	平均值	中位数	方差	最大值
a	D	土层埋深/m	2.00	5.30	4.50	2.34	14.80
b	σ_v	竖向总应力/kPa	38.40	94.77	85.40	40.37	251.60
c	σ'_v	竖向有效应力/kPa	27.80	64.60	55.48	29.78	140.80
d	SC	砂土类别	1.00	2.37	2.00	0.89	4.00
e	V_s	剪切波速/($\text{m}\cdot\text{s}^{-1}$)	79.00	141.61	133.00	30.90	274.00
f	$a_{\max}$	地表水平峰值加速度 g	0.02	0.21	0.18	0.12	0.51
g	M_w	地震震级	5.90	6.82	7.10	0.53	8.30

表 3 砂土液化数据库三样本统计特性
Table 3 Discriminant results of the forecast sample NO. 3

影响因素	符号	描述说明	最小值	平均值	中位数	方差	最大值
a	q_c	锥尖阻力/MPa	0.90	5.82	4.90	4.08	25.00
b	R_f	CPT 摩擦比/%	0.10	1.22	0.90	1.05	5.20
c	σ'_v	竖向有效应力/kPa	22.50	74.65	62.80	34.32	215.20
d	σ_v	竖向总应力/kPa	26.60	106.89	90.30	55.24	274.00
e	$a_{\max}$	地表水平峰值加速度 g	0.08	0.29	0.25	0.14	0.80
f	M_w	地震震级	6.00	6.95	7.10	0.44	7.60

2.2 影响因素分析

互信息 (Mutual Information)^[15] 法能计算各个影响因素与实测结果之间的相关性,其值越大相关性越高(图 2)。由图 2 可知,无论哪个数据库,对结果影响最大的都是竖向有效应力(σ'_v)和竖向总应力(σ_v)。这是因为容易液化的土体黏结力很低,其稳定性主要取决于土体颗粒间的摩擦力,而摩擦力与其所受到的

应力成正比。数据库一中,另一个比较突出的影响因素是内摩擦角(ϕ'),它反映的是摩擦特性。地震震级(M_w)的互信息值最低,不到 0.01,说明在 SPT 场地可不考虑这个因素。其余值大体相当,对判定结果都有一定影响。数据库二、三中,互信息值差距均小于 0.12,说明包含剪切波速(V_s)、CPT 摩擦比(R_f)在内的每个影响因子都很重要,没有无关变量。

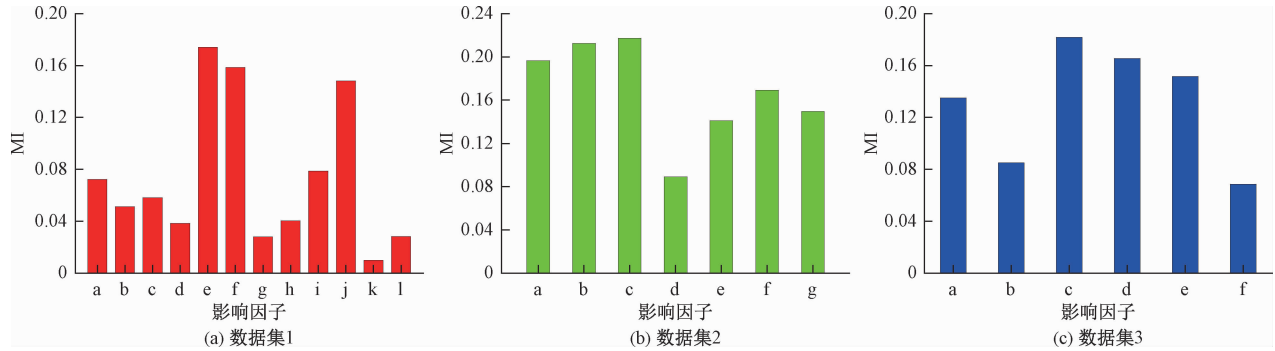


图 2 砂土液化影响因子互信息值

Fig.2 The mutual information of the influence factors of soil liquefaction

2.3 模型构建及对比分析

为综合比较分析 GPC 及另外 9 种判别方法在砂土液化预测方面的性能,首先对数据库里的数据进行归一化处理,使影响因子的值属于 $[0,1]$;然后在每个数据库中随机抽取 80% 的数据作为训练样本,其余为测试样本。ELM、ANN 初始参数为隐含层神经元个数, LSSVM、SVM 初始参数为常规化参数 C 和核函数参数 σ^2 。针对 ELM、LSSVM、SVM,采用网格算法和 10 次 10 折交叉验证确定最佳初始参数,由于 ANN 模型

(单隐层 BP 网络)训练时间较长,为节省时间,采用 1 次 10 折交叉验证(ANN 模型运行 200 次需耗时 47 小时);FDA、DDA、BDA、MLR、CART 无需初始参数,而 GPC 可进行超参数自适应获取。模型训练完毕后,用测试样本进行检验,获得测试样本预测精度。上述过程重复 200 次,求取平均预测精度,以保证结果的可靠性和代表性,计算流程如图 3 所示。上述所有运算均在 MATLAB 环境中运行,测试结果见表 4,预测精度见图 4。

表 4 测试结果
Table 4 Test results

数据库	统计特征	FDA	DDA	BDA	MLR	ELM	LSSVM	SVM	GPC	CART	ANN	均值
一	平均值	0.741 8	0.750 2	0.782 4	0.743 8	0.723 1	0.745 2	0.830 0	0.869 1	0.805 6	0.784 3	0.777 6
	标准差	0.033 0	0.038 0	0.034 4	0.035 2	0.039 4	0.038 2	0.031 5	0.028 9	0.035 7	0.055 1	
二	平均值	0.771 4	0.767 6	0.809 7	0.783 5	0.771 2	0.769 1	0.804 5	0.910 1	0.832 8	0.779 9	0.800 0
	标准差	0.062 1	0.061 3	0.060 2	0.060 1	0.059 7	0.073 6	0.060 7	0.048 9	0.063 6	0.082 0	
三	平均值	0.880 3	0.825 3	0.881 3	0.889 8	0.884 2	0.811 1	0.925 0	0.940 1	0.845 6	0.914 1	0.879 7
	标准差	0.045 6	0.059 2	0.046 4	0.044 1	0.049 5	0.052 1	0.035 9	0.034 9	0.048 6	0.043 7	
均值		0.797 8	0.781 0	0.824 5	0.805 7	0.792 8	0.775 1	0.853 2	0.906 4	0.828 0	0.826 1	

注:红色粗体表示最大值,蓝色粗体表示次大值,黑色粗体表示第三大值。

由表 4 可知,模型表现最优的是 GPC,其次是 SVM,再次是 CART,平均正确率分别为 0.906 4、0.853 2、0.828 0, LSSVM 表现最差,平均正确率只有 0.775 1。图 4 中,红色数字表示最大值,蓝色数字表示第二大值,黑色数字表示第三大值。GPC 模

型在数据库一、二、三中均取得最大值,说明 GPC 模型适用于各种场地砂土液化的预测,且效果较于其他 9 种都好。从数据库平均值来看,数据库一的平均值最低,为 0.777 6,原因是存在无关变量(地震震级)。图 4 显示, GPC 模型的上下线间距、箱长

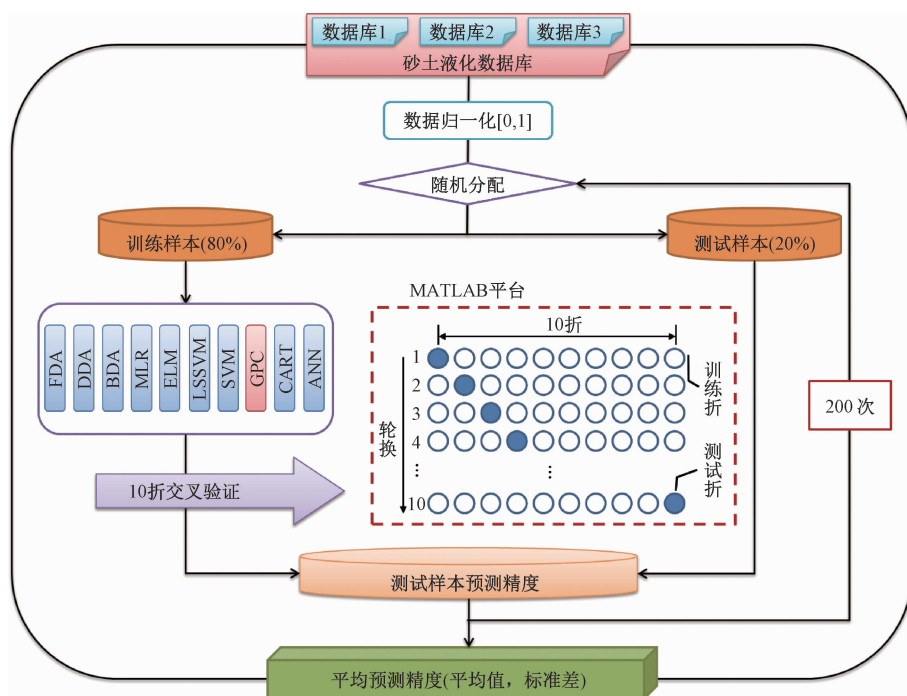


图 3 计算流程

Fig. 3 Illustration of the calculation procedure

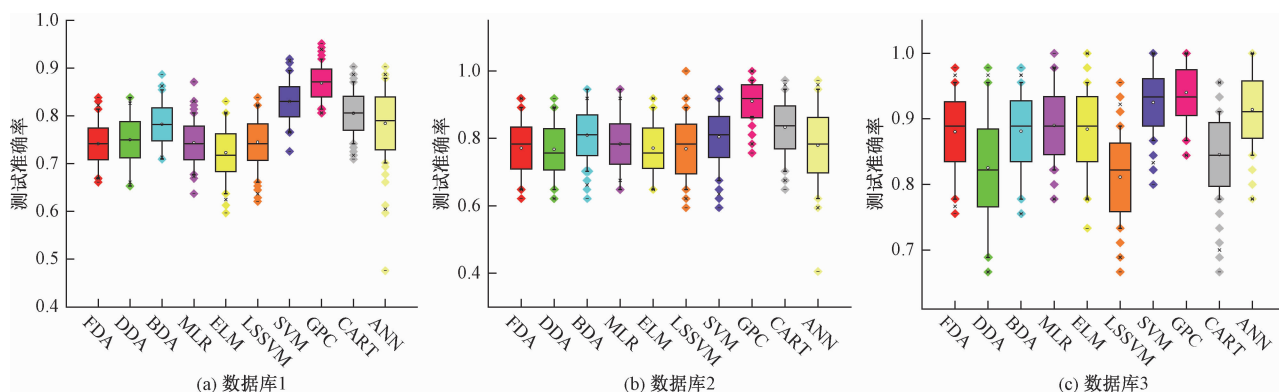


图 4 测试样本预测精度箱线图

Fig. 4 The boxplots of prediction accuracy of the test samples

都小于其他 9 种模型,进一步显示了 GPC 模型的稳健性。

图 5 给出了 10 种判别模型在三个数据库中的测试准确率累计分布曲线(CDF)。模型的 CDF 越低,说明该模型越好。从图 5 中可以看出,GPC 模型明显优于其他方法。

2.4 影响因子敏感性分析

用傅立叶振幅灵敏度分析方法(Fourier Amplitude Sensitivity Test, FAST)^[16]对三个数据库中的影响因子进行敏感度分析,得到各影响因子的一阶灵敏度指数 FOSI,表示仅由一个参数变异对模型

输出值变异所做出的贡献,其值如图 6 所示。由图 6 可知,数据库一敏感性最高的是标准贯入击数($(N_1)_{60}$);数据库二敏感性最高的是土层埋深(D),其次为剪切波速(V_s),两者相差不大,且均远大于其他值;数据库三敏感性最高的是锥尖阻力(q_c),其次为地表水平峰值加速度(a_{max}),再次为 CPT 摩擦比(R_f),其中锥尖阻力和 CPT 摩擦比为静力触探法(CPT)所得出的重要指标。由此可以看出三个数据库所包含的数据都极具代表性,GPC 模型对三大最主要场地的砂土液化都能做出很好的预测。

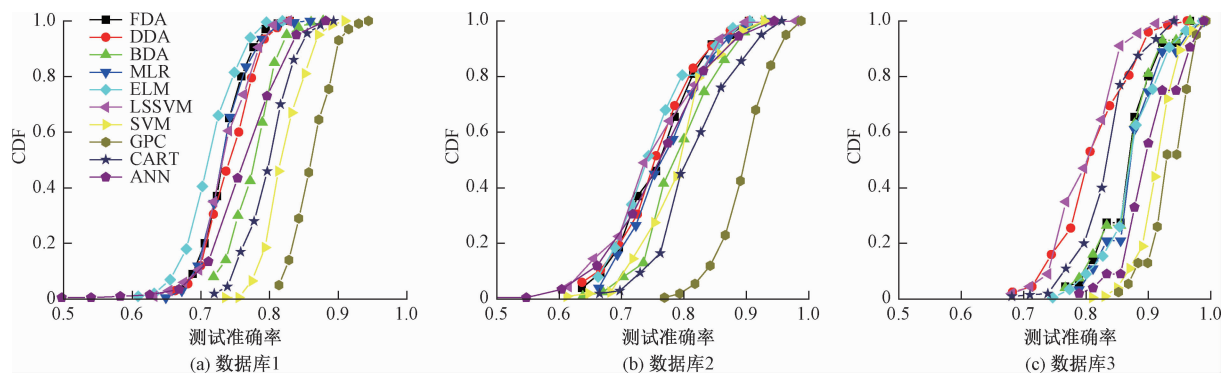


图5 测试准确率累计分布曲线

Fig.5 Cumulative distribution curves of the test accuracy

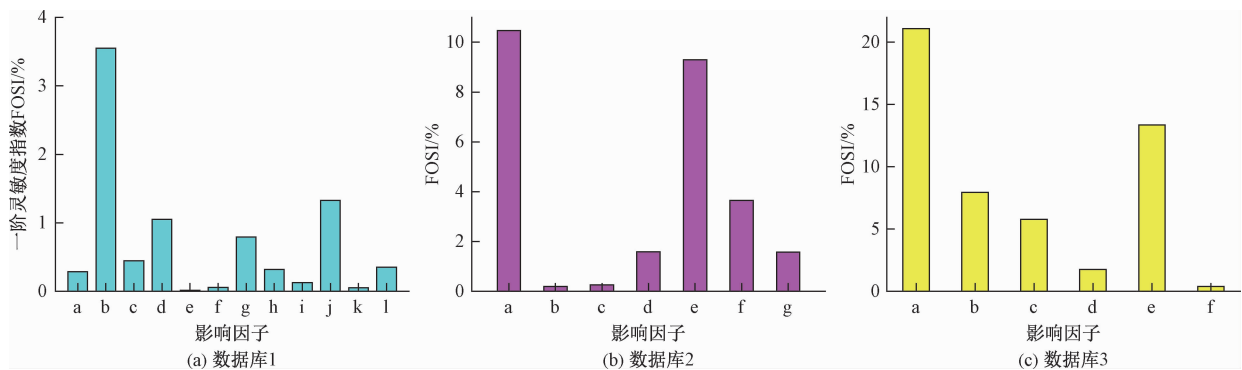


图6 判别因子 FOSI

Fig.6 The FOSI of initial influence factors

3 结论

(1) 针对不同地区砂土液化主要影响因素不同的特点,本文选取三个数据库,分别取自 SPT 场地、 V_s 场地和 CPT 场地,以消除样本选择对模型分析的影响。灵敏度分析显示,三个数据库中的数据都极具代表性。

(2) 本文选取 9 中判别分析方法与 GPC 模型进行对比分析,结果显示 GPC 模型在预测准确率、预测精度、模型稳健性方面均明显优于其他模型。

(3) 由三个数据库的代表性可知,GPC 模型具有很强的适应性,能很好的预测世界各地砂土液化问题。

参考文献:

- [1] 刘红军,杨东海. 饱和砂土地震液化危害及液化机理分析[J]. 森林工程, 2005, 21(4): 29-31.
LIU Hongjun, YANG Donghai. Seismic liquefaction hazards of saturated sand and analysis of liquefaction mechanism [J]. Forest Engineering, 2005, 21(4): 29-31.

- [2] 刘勇健. 基于聚类—二叉树支持向量机的砂土液化预测模型[J]. 岩土力学, 2008, 29(10): 2764-2768.

LIU Yongjian. Support vector machine model for predicting sand liquefaction based on clustering binary tree algorithm [J]. Rock and Soil Mechanics, 2008, 29(10): 2764-2768.

- [3] 李方明,陈国兴. 基于 BP 神经网络的饱和砂土液化判别方法[J]. 自然灾害学报, 2005, 14(2): 108-114.

LI Fangming, CHEN Guoxing. Saturated sand liquefaction potential estimation method based on BP neural network [J]. Journal of Natural Disasters, 2005, 14(2): 108-114.

- [4] 陈国兴,李方明. 基于径向基函数神经网络模型的砂土液化概率判别方法[J]. 岩土工程学报, 2006, 28(3): 301-305.

CHEN Guoxing, LI Fangming. Probabilistic estimation of sand liquefaction based on neural network model of radial basis function [J]. Chinese Journal of Geotechnical Engineering, 2006, 28(3): 301-305.

- [5] TANG Xiaowei, HU Jilei, QIU Jiangnan. Identifying significant influence factors of seismic soil liquefaction and analyzing their structural relationship [J]. Ksce Journal of Civil Engineering, 2016; 1 - 9.
- [6] 金志仁. 基于距离判别分析方法的砂土液化预测模型及应用 [J]. 岩土工程学报, 2008, 30 (5) : 776 - 780.
- JIN Zhiren. Prediction of sand liquefaction based on distance discriminant analysis and its application [J]. Chinese Journal of Geotechnical Engineering, 2008, 30 (5) : 776 - 780.
- [7] 夏建中, 罗战友, 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, et al. Support vector machine model for predicting sand liquefaction [J]. Chinese Journal of Rock Mechanics and Engineering, 2005, 24 (22) : 4139 - 4144.
- [8] 薛新华, 钟声. 基于 Fisher 判别法的砂土液化势判别 [J]. 沈阳建筑大学学报 (自然科学版), 2016, 32 (6) : 1070 - 1074.
- XUE Xinhua, ZHONG Sheng. Sand liquefaction prediction using fisher discriminant analysis method [J]. Journal of Shenyang Jianzhu University (Natural Science), 2016, 32 (6) : 1070 - 1074.
- [9] 胡记磊, 唐小微, 裘江南. 基于贝叶斯网络的地震液化概率预测分析 [J]. 岩土力学, 2016, 37 (6) : 1745 - 1752.
- HU Jilei, TANG Xiaowei, QIU Jiangnan. Prediction of probability of seismic induced liquefaction based on Bayesian network [J]. Rock and Soil Mechanics, 2016, 37 (6) : 1745 - 1752.
- [10] RASMUSSEN C E, WILLIAMS C K I. Gaussian processes for machine learning [M]. Cambridge: MIT Press, 2006.
- [11] NICKISCH H, RASMUSSEN C E. Approximations for binary Gaussian Process classification [J]. Journal of Machine Learning Research, 2008 (9) : 2035 - 2078.
- [12] ADEL M. HANNA, DERIN URAL, et al. Neural network model for liquefaction potential in soil deposits using Turkey and Taiwan earthquake data [J]. Soil Dynamics and Earthquake Engineering, 2007, 27: 521 - 540.
- [13] JUANG C H, CHEN C J. A rational method for development of limit state for liquefaction evaluation based on shear wave velocity measurements [J]. International Journal for Numerical and Analytical Methods in Geomechanics, 2000, 24: 1 - 27.
- [14] ANTHONY T C GOH, S H GOH. Support vector machines: Their use in geotechnical engineering as illustrated using seismic liquefaction data [J]. Computers and Geotechnics, 2007, 34: 410 - 421.
- [15] 童楚东, 史旭华. 基于互信息的 PCA 方法及其在过程监测中的应用 [J]. 化工学报, 2015, 66 (10) : 4101 - 4106.
- TONG Chudong, SHI Xuhua. Mutual information based PCA algorithm with application in process monitoring [J]. Ciesc Journal, 2015, 66 (10) : 4101 - 4106.
- [16] CHONGGANG XU, GEORGE ZDZISLAW GERTNER. Reliability of global sensitivity indices [J]. Journal of Statistical Computation and Simulation, 2011, 81: 11 - 12.

敬告作者和读者

为适应我国科技信息化建设的需要,扩大作者、读者学术交流渠道,本刊已加入《中国学术期刊(光盘版)》、《中国期刊网》及《中文科技期刊数据库》。今后,本刊将作者著作权使用费与本刊稿酬一次性付给。如果作者不同意将文章编入上述数据库,请在投稿时声明,本刊将作适当处理。谢谢!

《中国地质灾害与防治学报》编辑部