

基于 Logistic 回归模型的膨胀土判别与分类

高卫东

(江苏师范大学 地理测绘与城乡规划学院,江苏 徐州 221116)

**摘 要:**采用《公路工程地质勘察规范》(JTG C20—2011)推荐的自由膨胀率、塑性指数、标准吸湿含水率作为膨胀土判别与分类指标,将膨胀潜势分为非、弱、中等、强膨胀土4个等级。以某高速公路沿线土样为例,利用SPSS软件建立了土膨胀潜势分级的有序Logistic回归模型,并利用所建模型对待判土样进行判别,结果与实际一致。研究表明:有序Logistic回归模型的判别性能良,能客观反映膨胀土分类的复杂状况,具有较好的工程应用前景。

**关键词:**膨胀土;膨胀潜势;有序Logistic回归;判别与分类;SPSS

中图分类号:TU443      文献标志码:A

文章编号:1001-5485(2020)06-0153-03

开放科学(资源服务)标识码(OSID):



Discrimination and Classification of Expansive Soil Based on Logistic Regression Model

GAO Wei-dong

(School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou 221116, China)

**Abstract:** According to *Code for highway engineering geological investigation* (JTG C20—2011), three indices, namely, free expansion rate, plasticity index, and water content of soil under standard moisture absorption, were selected as the factors for synthetic evaluation of expansive soil. The swelling potential of expansive soil was divided into four grades: non-expansive, weakly expansive soil, moderately expansive, and strongly expansive. With the soil samples along a highway as an example, the ordinal Logistic regression model of swelling potential classification was established by using SPSS software and was applied to testing other cases. The predicted results were in good agreement with the actual. The results indicate that the ordinal Logistic regression model performs excellently, and can objectively reflect the complicated situation of expansive soil classification, thus is of good prospect in practical engineering.

**Key words:** expansive soil; swelling potential; ordinal Logistic regression; discrimination and classification; SPSS

膨胀土主要由亲水矿物组成,具有随含水量增加而膨胀、随含水量减少而收缩的特性。这种胀缩易使土体稳定性受到破坏,强度降低,进而导致工程建筑变形破坏<sup>[1]</sup>。因此,在工程勘察过程中,对膨胀土进行正确判别与分类意义重大。

采用单一指标对膨胀土进行判别与分类往往存在一定的局限,多指标综合评价是提高膨胀土判别与分类准确性的有效途径<sup>[2]</sup>。

本文采用能够反映膨胀土性质的多个指标,以某高速公路沿线土样为例,尝试建立膨胀土判别与

分类的多分类有序Logistic回归模型,对建模样本的回判及待判样本的测试表明所建模型判别性能良好,在实际工程中可以运用。

1 膨胀土判别指标及分类标准

1.1 判别指标

目前,膨胀土的判别与分类指标主要有2种<sup>[3-6]</sup>:一种是直接指标,包括蒙脱石含量、比表面积、阳离子交换量等,虽然反映了膨胀土的本质特

性,但试验方法相对复杂,推广困难;另一种是间接指标,包括液限、塑限、塑性指数、自由膨胀率、黏粒含量等,由于其测试方法简单,在工程中普遍应用。

本文参照《公路工程地质勘察规范》(JTG C20—2011)(以下简称《规范》<sup>[7]</sup>),采用自由膨胀率、塑性指数及标准吸湿含水率作为膨胀土判别与分类指标。

(1)自由膨胀率  $F_s$ :自由膨胀率是人工制备的松散烘干土样在水中充分膨胀后增加的体积与原体积之百分比。该指标在一定程度上反映了土的黏土矿物成分、粒度成分、化学成分和交换阳离子成分等基本特性,同时又具有试验操作方便、简单易行等优点,是初步判别膨胀土和膨胀潜势强弱的指示性指标。

(2)塑性指数  $I_p$ :塑性指数是土的界限含水量液限与塑限的差值,综合反映了土的矿物成分和颗粒大小的影响,塑性指数愈大,表明土的颗粒愈细,比表面积愈大,土中的黏粒或亲水矿物含量愈高,土处在可塑状态的含水率变化范围愈大,则土的膨胀潜势就越强。

(3)标准吸湿含水率  $W_t$ :土的标准吸湿含水率是在标准条件下(温度为  $(25\pm 2)^\circ\text{C}$ 、相对湿度为  $(60\pm 3)\%$ )膨胀土试样恒重后的含水率。标准吸湿含水率能够反映土的矿物组成特性,与比表面积、阳离子交换量、蒙脱石含量之间线性相关程度高,较好地反映了膨胀土的本质特性<sup>[8]</sup>。

1.2 分类标准

依据《规范》<sup>[7]</sup>,膨胀土的判别标准见表 1。

表 1 膨胀土判别标准

Table 1 Criterion of identifying expansive soil

自由膨胀率 $F_s/\%$	塑性指数 $I_p$	标准吸湿含水率 $W_t/\%$
$\geq 40$	$\geq 15$	$\geq 2.5$

当土样符合表 1 中的 2 项指标时,应定为膨胀土。《规范》进一步将膨胀潜势划分为非、弱、中等和强膨胀土 4 个等级,标准见表 2,至于需要符合几项指标,《规范》没有给出明确规定。

表 2 膨胀潜势分级

Table 2 Classification of swelling potential

等级	自由膨胀率 $F_s/\%$	塑性指数 $I_p$	标准吸湿含水率 $W_t/\%$
非膨胀土	$< 40$	$< 15$	$< 2.5$
弱膨胀土	$[40, 60)$	$[15, 28)$	$[2.5, 4.8)$
中等膨胀土	$[60, 90)$	$[28, 40)$	$[4.8, 6.8)$
强膨胀土	$\geq 90$	$\geq 40$	$\geq 6.8$

2 多分类有序 Logistic 回归

多分类有序 Logistic 回归模型<sup>[9]</sup>是分析 3 个及

其以上取值水平的具有等级关系的被解释变量与其他自变量之间关系的多元回归模型。设被解释变量有  $k$  个分类,如果各分类的概率表示为  $\pi_1, \pi_2, \dots, \pi_k$ ,对  $p$  个自变量可拟合以下  $k - 1$  个 Logistic 回归方程:

$$\text{Logit}_1 = \text{Logit}\left(\frac{\pi_1}{1 - \pi_1}\right) = -\alpha_1 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad ; \quad (1)$$

$$\text{Logit}_2 = \text{Logit}\left(\frac{\pi_1 + \pi_2}{1 - \pi_1 - \pi_2}\right) = -\alpha_2 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad ; \quad (2)$$

...

$$\text{Logit}_{k-1} = \text{Logit}\left(\frac{\pi_1 + \pi_2 + \dots + \pi_{k-1}}{1 - \pi_1 - \pi_2 - \dots - \pi_{k-1}}\right) = -\alpha_{k-1} + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad \circ \quad (3)$$

第  $k$  个类作为参照类。式中:  $\alpha_1, \alpha_2, \dots, \alpha_{k-1}$  为回归方程的常数项;  $x_1, x_2, \dots, x_p$  为解释变量;  $\beta_1, \beta_2, \dots, \beta_p$  为解释变量的系数。上述各式中的参数估计可采用最大似然估计法求得,对于有序 Logistic 回归,解释变量的系数相同。

模型估计完成后,对模型进行检验。对回归模型整体检验常用指标有皮尔逊  $\chi^2$  (Pearson  $\chi^2$ ),若  $\chi^2$  统计量很小,则观测值与预测值之间差别不显著,表示模型对数据的拟合程度高,反之,则拟合不佳;偏差  $D$  统计量是采用似然比的自然对数乘以  $-2$  来构建的,当样本的规模足够大时,  $D$  服从  $\chi^2$  分布,其自由度为所设模型中协变类型数目与系数个数之差,  $D$  值较大,表示模型拟合不佳,反之,  $D$  值很小,表示模型拟合程度高。对回归系数的显著性检验通常采用 Wald 检验。

按照式(1)一式(3),分别求出  $\pi_1, \pi_1 + \pi_2, \dots, \pi_1 + \pi_2 + \dots + \pi_{k-1}$ ,然后根据  $\pi_1 + \pi_2 + \dots + \pi_{k-1} + \pi_k = 1$ ,可求出  $\pi_k$ :

$$\pi_1 = \frac{\exp(-\alpha_1 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(-\alpha_1 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} \quad ; \quad (4)$$

$$\pi_2 = \frac{\exp(-\alpha_2 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(-\alpha_2 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} - \pi_1 \quad ; \quad (5)$$

...

$$\pi_{k-1} = \frac{\exp(-\alpha_{k-1} + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(-\alpha_{k-1} + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} - \pi_1 - \pi_2 - \dots - \pi_{k-2} \quad ; \quad (6)$$

$$\pi_k = 1 - \pi_1 - \pi_2 - \dots - \pi_{k-1} \quad \circ \quad (7)$$

### 3 膨胀土判别与分类的 Logistic 回归模型

#### 3.1 数据来源

以文献[10]提供的合肥—六安—叶集高速公路沿线所取的 21 个土样为例,土样 1—17 为建模样本,18—21 为测试样本,膨胀潜势记为 1、2、3、4 四个等级,分别表示为非、弱、中等和强膨胀土,21 个土样数据见表 3。

表 3 膨胀土样本  
Table 3 Samples of expansive soils

土样号	自由膨胀率 $F_s/\%$	塑性指数 $I_p$	标准吸湿 含水率 $W_L/\%$	膨胀潜势 实际等级
1	25	16.5	2.8	1
2	24	18.9	4.3	1
3	31	17.3	3.2	1
4	31	26.3	4.3	2
5	38	22.6	4.0	2
6	27	26.4	4.3	2
7	40	23.4	4.0	2
8	48	28.6	4.9	2
9	44	29.6	4.5	2
10	48	32.7	5.3	2
11	46	35.5	5.4	3
12	54	31.1	6.2	3
13	55	42.6	6.0	3
14	64	39.0	8.5	3
15	71	48.9	9.5	4
16	71	38.7	7.9	3
17	60	53.3	7.0	4
18*	54	31.1	6.2	3
19*	47	29.7	5.8	2
20*	38	19.4	3.8	2
21*	69	48.3	7.6	4

注:带\*为待判样本。

#### 3.2 模型计算结果

自由膨胀率、塑性指数、标准吸湿含水率 3 个指标分别记为  $x_1, x_2, x_3$ , 作为膨胀土判别与分类的解释变量;膨胀潜势为被解释变量,取值为 1、2、3、4。利用 SPSS(Statistical Product and Service Solutions)中 Ordinal Regression 模块<sup>[11]</sup>,采用 Logit 函数,拟合出 3 个 Logit 回归方程:

$$\text{Logit}\left(\frac{\pi_1}{1 - \pi_1}\right) = 35.949 + 0.285x_1 + 0.899x_2 + 2.608x_3 \quad ; (8)$$

$$\text{Logit}\left(\frac{\pi_1 + \pi_2}{1 - \pi_1 - \pi_2}\right) = 54.548 + 0.285x_1 + 0.899x_2 + 2.608x_3 \quad ; (9)$$

$$\text{Logit}\left(\frac{\pi_1 + \pi_2 + \pi_3}{1 - \pi_1 - \pi_2 - \pi_3}\right) = 75.184 + 0.285x_1 + 0.899x_2 + 2.608x_3 \quad . (10)$$

自变量回归系数的 Wald 统计量值分别为 0.567,2.713,0.289,Wald 统计量的伴随概率  $p$  值分别为 0.451,0.100,0.591,3 个回归系数在显著性水平为 0.05 下都不显著,说明膨胀土判别与分类不取决于单个因素,而是由多个评判因素共同决定<sup>[12]</sup>,塑性指数( $x_2$ )系数的 Wald 统计量最大,表明塑性指数对膨胀土潜势判别的解释程度最高。SPSS 软件对模型中是否所有自变量偏回归系数全为 0 进行似然比检验,由  $p < 0.001$  知至少有一个自变量的偏回归系数不为 0,说明解释变量全体与连接函数 Logit 之间的线性关系显著,模型选择正确;计算 Pearson $\chi^2$  和偏差  $D$  这 2 个统计量数值分别为 1.148,1.942,2 个统计量的伴随概率  $p$  值均为 1.00,>0.05,表明自变量和因变量是相关的,模型的整体拟合优度好。

由拟合出的式(8)一式(10)3 个 Logit 回归方程,可以计算出各土样属于 4 个膨胀潜势等级的概率,分别记为  $\pi_1, \pi_2, \pi_3, \pi_4$ ,见表 4。

表 4 膨胀潜势概率及判别等级

Table 4 Swelling potential probability and discrimination grade						
土样号	$\pi_1$	$\pi_2$	$\pi_3$	$\pi_4$	实际等级	判别等级
1	1.00	0.00	0.00	0.00	1	1
2	0.96	0.04	0.00	0.00	1	1
3	0.99	0.01	0.00	0.00	1	1
4	0.00	1.00	0.00	0.00	2	2
5	0.03	0.97	0.00	0.00	2	2
6	0.01	0.99	0.00	0.00	2	2
7	0.01	0.99	0.00	0.00	2	2
8	0.00	0.99	0.01	0.00	2	2
9	0.00	1.00	0.00	0.00	2	2
10	0.00	0.63	0.37	0.00	2	2
11	0.00	0.17	0.83	0.00	3	3
12	0.00	0.17	0.83	0.00	3	3
13	0.00	0.00	1.00	0.00	3	3
14	0.00	0.00	0.99	0.01	3	3
15	0.00	0.00	0.00	1.00	4	4
16	0.00	0.00	0.98	0.02	3	3
17	0.00	0.00	0.01	0.99	4	4
18*	0.00	0.17	0.83	0.00	3	3
19*	0.00	0.92	0.08	0.00	2	2
20*	0.46	0.54	0.00	0.00	2	2
21*	0.00	0.00	0.03	0.97	4	4

根据模型计算土样归属某等级的最大概率确定其膨胀潜势。由表 4 可知,土样 1—17 建模样本的回判准确率为 100%,将 18—21 样本数据代入模型(8)一(10)计算,膨胀潜势等级分别为 3,2,2,4,均为膨胀土,与实际情况一致。文献[10]采用 5 个指标