

文章编号: 1000-7598 (2009) 08-2418-05

基于粗糙集理论的边坡稳定性评价中因素权重确定方法的研究

王广月, 崔海丽, 李倩

(山东大学 土建与水利学院, 济南 250061)

摘要: 为了准确计算影响边坡稳定性的各个因素的权重, 应用粗糙集理论对原始数据进行挖掘, 在实测数据离散化的基础上, 将权重确定问题转化为粗糙集中属性重要性评价问题, 建立了关于边坡稳定性评价的关系数据模型, 经过属性值特征化建立了知识系统, 在数据分析下通过分析评判方法对评价对象的支持度和重要性, 计算出边坡稳定性评价模型的权重。该方法克服了传统权重确定方法的主观性, 使得边坡稳定性评价方法更具客观性, 从而提高了边坡稳定性评价的精度, 通过实例说明该方法更加合理有效。

关键词: 边坡稳定性; 粗糙集理论; 属性重要性

中图分类号: TU 457

文献标识码: A

Investigation of method for determining factors weights in evaluating slope stability based on rough set theory

WANG Guang-yue, CUI Hai-li, LI Qian

(School of Civil and Hydraulic Engineering, Shandong University, Jinan 250061, China)

Abstract: To compute the weight factors that influence slope stability accurately, rough sets theory is used to analysis the raw data. Based on discretization of the measured data, the significance of attributes among rough sets is estimated instead of weight determination, and a relation data model about slope stability assessment is established. A knowledge system is built through making attribute value into eigenvalue. During data analysis, the weight of a slope stability assessment model is computed by analyzing the support and significance of forecasting method for the predicted object. The proposed approach overcomes the subjectivity of traditional weight determination method, improving the precision of risk assessment for slope stability. At last, an example is given to show the effectivity and rationality of this new method.

Key words: slope stability; rough set theory; importance of attribute

1 引言

岩土边坡的工作状况直接或间接地影响工程建筑物的稳定、耐久和安全, 边坡失稳破坏将带来巨大的人员和经济损失, 所以边坡稳定性的问题一直是一个十分重大的问题。边坡工程的稳定性受到众多因素的影响, 有很多因素既难以确定, 又难以定量估计, 因而其研究方法从确定性分析方法发展到不确定性分析方法, 并且为了克服边坡稳定性工程地质评价的随意性和不确定性, 人们尝试应用数学方法对整个评价过程进行定量或半定量描述, 并

获得了较大进展, 如: 模糊综合评价、灰色分析理论、信息量模型法、神经网络法等^[1-6], 而因素指标的权重是决定这些方法评价结果准确性的重要因素。粗糙集论是由波兰数学家Z. Pawlak在1982年提出的一种处理模糊和不确性问题的新型数学工具, 其主要思想就是在保持分类能力不变的前提下, 通过知识约简, 导出问题的决策或分类规则。本文将粗糙集理论引入到边坡稳定性评价的因素权重确定当中, 首先利用数据预处理方法对实测数据进行离散化处理, 然后利用知识约简方法计算出各属性的权重。

收稿日期: 2008-04-28

第一作者简介: 王广月, 男, 1963年生, 教授, 主要从事岩土力学和地基基础教学与研究工作。E-mail: wgyue@sdu.cn

2 粗糙集理论

2.1 粗糙集

设 R 是集合 A 上的二元关系, 如果它是自反、对称和传递的, 则它是 A 上的等价关系。设 $X \subseteq U$, R 是 U 上的等价关系, $A=(U, R)$ 是一个近似空间, 在 A 上, 如果 X 是一些 R 基本类的并集, 则称 X 是可定义的; 否则, 称 X 是不可定义的^[7]。 R 可定义集是全集 U 上那样一些子集, 这些子集在个体全集 U 上是恰好可被定义的, 而 R 不可定义集是子集 X 上不可能恰好被定义的。 R 可定义集被称作 R 一致集或 R 恰当集, 而 R 不可定义集也被说成是 R 不一致集或称 R -Rough 集, 简称不一致集或 Rough 集。

2.2 信息系统

粗糙集理论中的知识表达方式一般采用信息系统的形式, 一个信息系统 S 是一个系统 (U, A) , 其中 $U=\{u_1, u_2, \dots, u_{|U|}\}$ 是有限非空集, 称为论域或对象空间, U 中的元素称为对象; $A=\{a_1, a_2, \dots, a_{|A|}\}$ 也是一个非空有限集, A 中的元素称为属性; 对于每个 $a \in A$, 有一个映射 $a:U \rightarrow a(U)$, 且 $a(U)=\{a(u)|u \in U\}$ 称为属性 a 的值域^[8]。一个信息系统可以用一个信息表来表示, 当没有重复元组时, 信息表是一个关系数据库。如果 $A=CU D$, $C \cap D = \emptyset$, 则信息系统 (U, A) 为一个决策表, 其中 C 中的属性成为条件属性, D 中的属性称为决策属性。

2.3 上近似、下近似

设给定集合 X 和 Y , 又设 P 和 Q 是分别定义在 X 和 Y 上的不分明关系 (近似于等价关系)。设 $R \subseteq XY$ 是 XY 上任意二元关系, 又设 $I=PR$ 是不分明关系的积, 则 R 的 I -上近似和 I -下近似分别被定义如下:

$$\begin{aligned} I^*(R) &= \{(x, y) \in XY : I((X, Y)) \cap R \neq \emptyset\} \\ I_*(R) &= \{(x, y) \in XY : I((X, Y)) \subseteq R\} \end{aligned} \quad (1)$$

上近似和下近似的差为

$$BN_I(R) = I^*(R) - I_*(R) \quad (2)$$

式 (2) 被称为 R 的边界线区域。

下近似又称为 R 的 I 正区域, 被记成 $POS_I(R) = I_*(R)$, 它是如此一些个体元素的集合, 这些元素完全属于 X 的成员。

2.4 不可分辨关系

设论域为 U , P 为条件属性 C 的一个子集, 则由 P 决定的不可分辨关系 $IND(P)$ 为

$$IND(P) = \{(x, y) \in U \times U / \forall a \in P, f(x, a) = f(y, a)\} \quad (3)$$

式中: $f(x, a)$ 表示论域元素 $x \in U$ 关于属性 a 的取值; 不可分辨关系 $IND(P)$ 构成了对论域 U 的一个分类, 记作 $U/IND(P)$ 。

2.5 知识约简

知识约简是粗糙集理论的核心内容。众所周知, 知识库中知识 (属性) 并不是同等重要的, 甚至其中某些知识是冗余的。所谓知识约简, 就是在保持知识库分类能力不变的条件下, 删除其中不相关或不重要的知识。

设 R 是一个等价关系族, $r \in R$ 。如果 $IND(R) = IND(R - \{r\})$, 则称 r 在 R 中是可被约去的知识; 如果 $P = R - \{r\}$ 是独立的, 则 P 是 R 的一个约简。

2.6 属性重要性

令 P 和 Q 分别为条件属性 C 和决策属性 D 的一个子集, 则

$$K = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|} \quad (4)$$

式 (4) 表示知识 Q 是 K 度依赖于知识 P 的, 记为 $P \Rightarrow_K Q$ 。其中: $|U|$ 表示 U 的基, 即 U 中元素的个数; 同理, $|POS_P(Q)|$ 表示 $POS_P(D)$ 中元素的个数。属性子集 $C_i \subseteq C$ 关于 D 的重要性定义为

$$\sigma_D(C_i) = \gamma_C(D) - \gamma_{C-C_i}(D) \quad (5)$$

$\sigma_D(C_i)$ 的值越大, 表明相应属性的重要性越大; 反之, 重要性越小。

2.7 计算步骤

(1) 在不改变数据的分类能力的前提下, 将各属性值进行离散化处理, 建立边坡稳定性的评价指标体系。粗糙集方法是一类符号化分析方法, 需要将连续的属性离散化。连续属性的离散就是将连续属性值域划分为若干个区间, 每个区间用不同代码表示属性值。常用的方法有神经网络法、模糊聚类法、等频离散法、等宽离散法及利用相关领域知识的方法等。

(2) 形成决策表。采用离散化后的属性值形成一张二维表格, 每一行描述一个对象, 每一列描述对象的一种条件属性, 最后一列为决策属性。

(3) 知识约简删除属性值相同的对象, 也就是重复的数据行, 使决策表达到最简单, 以减少计算

工作量。

(4) 分别按条件属性和决策属性对论域进行等价类划分。

(5) 为了找出某些属性的重要性, 需要从表中去掉一些属性, 再来考察没有该属性后分类会怎样变化。若去掉该属性相应分类变化较大, 则说明该属性强度大, 即重要性高; 反之, 说明该属性的强度小, 即重要性低。

(6) 利用属性重要性的计算公式计算各条件属性对决策属性的重要性。

(7) 将重要性进行归一化处理, 得到更直观的比较结果。

3 实例分析

表 1 是文献[9]搜集到的 32 个土质边坡工程实例的数据, 在这张信息表中, 论域 $U=\{1, 2, 3, 4, 5, \dots, 32\}$, 条件属性集 $C=\{C_1, C_2, C_3, C_4, C_5, C_6\}$, 决策属性集 $D=\{D_1, D_2\}$, 其中: C_1 为重度, C_2 为黏聚力, C_3 为摩擦角, C_4 为边坡角, C_5 为边坡高度, C_6 为孔隙压力比, D_1 为安全系数, D_2 为边坡状态。按照上面提到的计算步骤计算属性重要性。

将上面的数据进行离散化处理, 得到信息系统的约简见表 2。将上面的表格进行数据约简, 得到简化的信息系统数据表, 见表 3。

表 1 边坡工程实例的数据
Table 1 Data of slope projects

序号	重度 (kN/m ³)	黏聚力 (kPa)	摩擦角 (°)	边坡角 (°)	边坡高度 (m)	孔隙压力比	安全系数	边坡状态
1	22.4	10.0	35.0	45.0	10.0	0.40	0.90	0
2	20.0	20.0	36.0	45.0	50.0	0.50	0.83	0
3	20.0	0.1	36.0	45.0	50.0	0.25	0.79	0
4	20.0	0.1	36.0	45.0	50.0	0.50	0.67	0
5	22.0	0.0	40.0	33.0	8.0	0.35	1.450	1
6	24.0	0.0	40.0	33.0	8.0	0.30	1.58	1
7	20.0	0.0	24.5	20.0	8.0	0.35	1.37	1
8	18.0	0.0	30.0	20.0	8.0	0.30	2.05	1
9	27.0	40.0	35.0	43.0	420.0	0.25	1.15	0
10	27.0	50.0	40.0	42.0	407.0	0.25	1.44	1
11	27.0	35.0	35.0	42.0	359.0	0.25	1.27	1
12	27.0	37.5	35.0	37.8	320.0	0.25	1.24	1
13	27.0	32.0	33.0	42.6	301.0	0.25	1.16	0
14	27.0	32.0	33.0	42.2	289.0	0.25	1.30	1
15	27.3	14.0	31.0	41.0	110.0	0.25	1.25	1
16	27.3	31.5	29.7	41.0	135.0	0.25	1.25	1
17	27.3	16.8	28.0	50.0	90.5	0.25	1.25	1
18	27.3	26.0	1.0	50.0	92.0	0.25	1.25	1
19	27.3	10.0	39.0	41.0	511.0	0.25	1.47	1
20	27.3	10.0	39.0	40.0	470.0	0.25	1.43	1
21	25.0	46.0	35.0	47.0	443.0	0.25	1.28	1
22	25.0	46.0	35.0	44.0	435.0	0.25	1.37	1
23	25.0	46.0	35.0	46.0	432.0	0.25	1.23	1
24	26.0	150.0	45.0	30.0	200.0	0.25	1.20	0
25	18.0	25.0	0.0	30.0	6.0	0.25	1.09	0
26	18.0	12.0	0.0	30.0	6.0	0.25	0.78	0
27	22.4	10.0	35.0	30.0	10.0	0.25	2.00	1
28	21.4	10.0	30.3	30.0	20.0	0.25	1.70	1
29	22.0	10.0	36.0	45.0	50.0	0.25	1.02	0
30	22.0	20.0	36.0	45.0	50.0	0.25	0.89	0
31	12.0	0.0	30.0	35.0	4.0	0.25	1.46	1
32	12.0	0.0	30.0	45.0	8.0	0.25	0.80	0

表 2 决策表
Table 2 Table of decision-making

序号	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	D ₁	D ₂
离散值为 0	<23	<21	<36	<45	<60	<0.26	<1	不稳定
离散值为 1	≥23	≥21	≥36	≥45	≥60	≥0.26	≥1	稳定
1	0	0	0	1	0	1	0	0
2	0	1	1	1	0	1	0	0
3	0	0	1	1	0	0	0	0
4	0	0	1	1	0	1	0	0
5	0	0	1	0	0	1	1	1
6	1	0	1	0	0	1	1	1
7	0	0	0	0	0	1	1	1
8	0	0	0	0	0	1	1	1
9	1	1	0	1	1	0	1	0
10	1	1	1	0	1	0	1	1
11	1	1	0	0	1	0	1	1
12	1	1	0	0	1	0	1	1
13	1	1	0	0	1	0	1	1
14	1	1	0	0	1	0	1	1
15	1	0	0	0	1	0	1	1
16	1	1	0	0	1	0	1	1
17	1	0	0	1	1	0	1	0
18	1	1	0	1	1	0	0	0
19	1	0	1	0	1	0	1	1
20	1	0	1	0	1	0	1	1
21	1	1	0	1	1	0	1	0
22	1	1	0	1	1	0	1	0
23	1	1	0	1	1	0	1	0
24	1	1	1	0	1	0	1	0
25	0	1	0	0	0	0	1	0
26	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	1	1
28	0	0	0	0	0	0	1	1
29	0	0	1	1	0	0	1	0
30	0	0	1	1	0	0	0	0
31	0	0	0	0	0	0	1	1
32	0	0	0	1	0	0	0	0

表 3 知识约简后的决策表
Table 3 Table of decision-making after the reduced knowledge

序号	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	D ₁	D ₂
1	0	0	0	1	0	1	0	0
2	0	0	1	1	0	1	0	0
3	0	0	1	1	0	0	0	0
4	0	0	1	0	0	1	1	1
5	1	0	1	0	0	1	1	1
6	0	0	0	0	0	1	1	1
7	1	1	0	1	1	0	1	0
8	1	1	1	0	1	0	1	1
9	1	1	0	0	1	0	1	1
10	1	0	0	0	1	0	1	1
11	1	0	0	1	1	0	1	0
12	1	1	0	1	1	0	0	0
13	1	0	1	0	1	0	1	1
14	1	1	0	1	1	0	1	0
15	1	1	1	0	1	0	1	0
16	0	1	0	0	0	0	1	0
17	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	1	1
19	0	0	1	1	0	0	1	0
20	0	0	0	1	0	0	0	0

对上面的数据论域分别按条件属性和决策属性进行分类得^[5]:

$$U/IND(C)=\{\{1\},\{2\},\{3,19\},\{4\},\{5\},\{6\},\{7,12,14\},\{8,15\},\{9\},\{10\},\{11\},\{13\},\{16\},\{17,18\},\{20\}\}$$

$$U/IND(D)=\{\{1,2,3,12,17,20\},\{4,5,6,8,9,10,13,18\},\{7,11,14,15,16,19\}\}$$

分别去掉一个条件属性后的论域分类得

$$U/IND(C-C_1)=\{\{1\},\{2\},\{3,19\},\{4,5\},\{6\},\{7,12,14\},\{8,15\},\{9\},\{10\},\{11\},\{13\},\{16\},\{17,18\},\{20\}\}$$

$$U/IND(C-C_2)=\{\{1\},\{2\},\{3,19\},\{4\},\{5\},\{6\},\{7,11,12,14\},\{8,13,15\},\{9,10\},\{16,17,18\},\{20\}\}$$

$$U/IND(C-C_3)=\{\{1,2\},\{3,19,20\},\{4,6\},\{5\},\{7,12,14\},\{8,9,15\},\{10,13\},\{11\},\{16\},\{17,18\}\}$$

$$U/IND(C-C_4)=\{\{1,6\},\{2,4\},\{3,19\},\{5\},\{7,9,12,14\},\{8,15\},\{10,11\},\{13\},\{16\},\{17,18,20\}\}$$

$$U/IND(C-C_5)=\{\{1\},\{2\},\{3,19\},\{4\},\{5\},\{6\},\{7,12,14\},\{8,15\},\{9\},\{10\},\{11\},\{13\},\{16\},\{17,18\},\{20\}\}$$

$$U/IND(C-C_6)=\{\{1,20\},\{2,3,19\},\{4\},\{5\},\{6,17,18\},\{7,12,14\},\{8,15\},\{9\},\{10\},\{11\},\{13\},\{16\}\}$$

分析每条属性关于决策属性的重要性:

$$POS_C(D)=\{1,2,4,5,6,9,10,11,13,16,20\}$$

$$POS_{C-\{C_1\}}(D)=\{1,2,4,5,6,9,10,11,13,16,20\}$$

$$POS_{C-\{C_2\}}(D)=\{1,2,4,5,6,9,10,20\}$$

$$POS_{C-\{C_3\}}(D)=\{1,2,4,5,6,10,11,13,16\}$$

$$POS_{C-\{C_4\}}(D)=\{5,13,16\}$$

$$POS_{C-\{C_5\}}(D)=\{1,2,4,5,6,9,10,11,13,16,20\}$$

$$POS_{C-\{C_6\}}(D)=\{1,4,5,9,10,11,13,16,20\}$$

$$\gamma_C(D)=\frac{|POS_C(D)|}{|U|}=\frac{11}{20}$$

$$\sigma_D(C_1)=\gamma_C(D)-\gamma_{C-\{C_1\}}(D)=\frac{11}{20}-\frac{11}{20}=0$$

$$\sigma_D(C_2)=\gamma_C(D)-\gamma_{C-\{C_2\}}(D)=\frac{11}{20}-\frac{8}{20}=\frac{3}{20}$$

$$\sigma_D(C_3) = \gamma_C(D) - \gamma_{C-\{C_3\}}(D) = \frac{11}{20} - \frac{9}{20} = \frac{2}{20}$$

$$\sigma_D(C_4) = \gamma_C(D) - \gamma_{C-\{C_4\}}(D) = \frac{11}{20} - \frac{3}{20} = \frac{8}{20}$$

$$\sigma_D(C_5) = \gamma_C(D) - \gamma_{C-\{C_5\}}(D) = \frac{11}{20} - \frac{11}{20} = 0$$

$$\sigma_D(C_6) = \gamma_C(D) - \gamma_{C-\{C_6\}}(D) = \frac{11}{20} - \frac{9}{20} = \frac{2}{20}$$

经归一化处理, 得各条件属性的重要性为 0.000、0.200、0.133、0.533、0.000、0.133。可见在影响土质边坡稳定性的因素中, 边坡角的影响最大, 内聚力的影响比较大, 摩擦角和孔隙压力比的影响比较小, 重度和边坡高度的影响最小。这个结果可以用于土质边坡的加固实例中, 但是应该结合具体的工程情况及实践经验等等。

4 结 语

粗糙集理论完全从实际数据出发, 挖掘数据中隐含的知识揭示客观的内在规律, 不受任何主观因素的影响, 因此得出的结论更可观, 也更具有实际意义。随着粗糙集理论的研究越来越深入, 它的应用也越来越广泛。本文给出了粗糙集理论在边坡稳定性分析当中的应用, 通过对实测数据的计算得到了影响边坡稳定性的主要因素和次要因素, 从而为边坡加固工作提供了一些依据。但是其中还存在不足之处, 就是数据离散化的划分界限没有明确的标准, 有待于进一步研究。

参 考 文 献

- [1] 薛新华, 张我华, 刘红军. 基于遗传算法和模糊神经网络的边坡稳定性评价[J]. 岩土力学, 2007, 28(12): 2643—2648.
- XUE Xin-hua, ZHANG Wo-hua, LIU Hong-jun. Evaluation of slope stability based on genetic algorithm and fuzzy neural network[J]. **Rock and Soil Mechanics**, 2007, 28(12): 2643—2648.
- [2] 刘端伶, 谭国焕, 李启光, 等. 岩石边坡稳定性和 Fuzzy 综合评判法[J]. 岩石力学与工程学报, 1999, 18(2): 170—175.
- LIU Duan-ling, THAM L G, LEE P K K. The stability of rock slope and fuzzy comprehensive evaluation method[J]. **Chinese Journal of Rock Mechanics and Engineering**, 1999, 18(2): 170—175.
- [3] 薛新华, 张我华, 刘红军. 基于SOFM神经网络的边坡稳定性评价[J]. 岩土力学, 2008, 29(8): 2236—2240.
- XUE Xin-hua, ZHANG Wo-hua, LIU Hong-jun. Evaluation of slope stability based on SOFM neural network[J]. **Rock and Soil Mechanics**, 2008, 29(8): 2236—2240.
- [4] 陈新民, 罗国煜. 基于经验的边坡稳定性灰色系统分析与评价[J]. 岩土工程学报, 1999, 21(5): 638—641.
- CHEN Xin-min, LUO Guo-yu. Grey system analysis and evaluation of slope stability based on experience[J]. **Chinese Journal of Geotechnical Engineering**, 1999, 21(5): 638—641.
- [5] 方国柱, 张正禄, 章传银. 工程高边坡稳定性评价的信息量法[J]. 武汉测绘科技大学学报, 1996, 21(4): 344—349.
- FANG Guo-zhu, ZHANG Zheng-lu, ZHANG Chuan-yin. Information value method for stability study of high rock slopes of engineering[J]. **Journal of Wuhan Technical University of Surveying and Mapping**, 1996, 21(4): 344—349.
- [6] 黄志全, 崔江利, 刘汉东. 边坡稳定性预测的混沌神经网络方法[J]. 岩石力学与工程学报, 2005, 23(22): 3808—3812.
- HUANG Zhi-quan, CUI Jiang-li, LIU Han-dong. Chaotic network method for slope stability prediction[J]. **Chinese Journal of Rock Mechanics and Engineering**, 2005, 23(22): 3808—3812.
- [7] 刘清. Rough 集及 Rough 推理[M]. 北京: 科学出版社, 2001.
- [8] 张文修, 吴伟业. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
- [9] 贺可强, 雷建和. 边坡稳定性的神经网络预测研究[J]. 地质与勘探, 2001, 37(6): 72—75.
- HE Ke-qiang, LEI Jian-he. A predication of the slope stability based on neural network[J]. **Geology and Prospecting**, 2001, 37(6): 72—75.
- [10] 王力, 袁长春, 吕大海, 等. 粗糙集理论在高层建筑结构知识发现中的应用[J]. 哈尔滨工业大学学报, 2006, 38(12): 2073—2076.
- WANG Li, YUAN Chang-chun, LÜ Da-gang, et al. Application of rough sets in high-rise building structure's knowledge discovery[J]. **Journal of Harbin Institute of Technology**, 2006, 38(12): 2073—2076.