



最小生成树法在环境背景值识别中的应用

辛世崇 王广珍

(吉林化工学院)

摘要 本文用最小生成树法解决环境中的聚类分析问题。该法用 x^2 距离系数矩阵描述样本变量间的相似关系, 同时用 Kruskal 法构造最小生成树, 最后以分类图的形式给出分类结果。

关键词: 最小生成树法; 聚类分析; 土壤重金属背景值。

模糊聚类分析广泛用于环境质量评价, 已提出的典型分类程序包括最小支撑树法和最大树的 Prim 算法等。但上述聚类分析过程的模糊非相关矩阵的建立, 运算过于复杂, 必须借助计算机完成^[1,2]。图论中最小生成树问题以往曾用于交通业中的线路及给排水管路最短距离的研究, J. Devillers 等近年首次提出用最小生成树法研究微生物对某些金属的耐受性问题^[3]。有关该法在环境质量评价及其类似问题中的应用尚未见报道, 本文通过最小生成树法在土壤重金属背景值识别中的应用实例, 说明它在环境科学领域应用的可行性和其快速、简便、结果直观等特点。

一、方法简介

图论中有关构造最小生成树的理论已有专著^[4], 本文仅简述如下:

最小生成树问题的数学模型是: 给定无向图 $G_n = [N, E]$ 及其中每一条边的长度 $l(x_i, x_j)$, 即无向图为已知, 要求 G_n 的部分树 $T_n = [N, E']$, 使连结所有节点各边总长为最小, 即:

$$\sum_{(x_i, x_j) \in E'} l(x_i, x_j) \rightarrow \min \quad (1)$$

对于具体问题, 其最小树的总长是固定的, 但其形式不一定是唯一的。

最小树的构造需先将原始数据(经适当处理)转换为距离矩阵。如果把 N 个样本(X 中 N 个列)看成 p 维空间中 N 个点, 则两样本间的相似程度可用 p 维空间中两点的距离来度量, 此距离称之为马氏距离。如果变量 x_1, x_2, \dots, x_p 互不相关, 马氏距离即为

欧氏距离, 也称之为 x^2 距离^[5], 即:

$$d_{jk} = \sqrt{\sum_{i=1}^p (x_{oj} - x_{ok})^2} \quad (2)$$

x^2 距离满足下列关系式:

$$d(a, b) = d(b, a) \quad (3)$$

$$d(a, b) = 0, \text{ 有 } a = b \quad (4)$$

$$d(a, c) \leq d(a, b) + d(b, c) \quad (5)$$

计算任两个样本 x_j 与 x_k 之间的距离, d 值越小表示两个样本相似程度越大, d 值越大, 表示两个样本相似程度越小。算出任何两两样本的距离后, 可排出距离系数矩阵:

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \dots & d_{mN} \end{bmatrix} \quad (6)$$

由式(3)可知, 该矩阵为实对称矩阵, 所以只需计算其上三角或下三角部分即可。根据 D 可按 Kruskal 法构造出最小生成树^[6], 按给定值对最小生成树进行切割后, 可得到分类树。据此可对 N 个点进行聚类分析, 距离近的点归为一类, 距离远的点属不同类。下面以某地土壤背景值研究数据为例, 用最小生成树法对各类土壤按其背景值高低进行识别分类, 并说明分类步骤。

二、应用实例

某地区调查了 8 类典型土样, 10 种重金属的背景值(36 个剖面, 108 个土样), 所调查的土壤类型分别为 A 暗棕色森林土, 花岗岩母质、B 暗棕色森林土, 玄武岩母质、C 白浆土, 玄武岩母质、D 白浆土, 第四纪沉积物母质、E 草甸土, 江河近代冲积物母

表 1 某地区土壤中 10 种重金属背景值 (ppm)

土壤 \ 背景值	Cd	Pb	Cr	Ni	Zn	Mn	Hg	As	Cu	Ti
A	0.26	12.86	47.26	14.00	57.39	309.7	0.04	3.16	37.44	7930
B	0.29	11.26	51.05	18.27	54.54	416.9	0.09	4.39	45.25	8780
C	0.19	12.30	54.35	18.71	53.87	301.7	0.05	1.87	14.37	7850
D	0.23	12.21	56.03	21.73	50.76	374.3	0.033	3.32	19.00	8060
E	0.20	12.63	66.09	22.47	65.46	452.7	0.126	3.34	20.66	7778
F	0.20	13.72	52.26	18.31	56.73	400.8	0.03	1.90	9.09	7825
G	0.31	16.23	62.40	23.19	67.82	568.3	0.09	2.72	17.93	9005
H	0.11	18.72	65.78	25.75	115.3	620.8	0.20	9.38	21.82	10580
全地区总背景值	0.25	12.01	56.98	19.31	56.87	364.1	0.07	3.46	29.61	8700

表 2 某地区土壤中 10 种重金属的 M'_{oi} 值

土 样 \ M'_{oi}	Cd	Pb	Cr	Ni	Zn	Mn	Hg	As	Cu	Ti
A	1.04	1.071	0.8294	0.7250	1.009	0.8507	0.5174	0.9133	1.264	0.9115
B	1.16	0.9376	0.8959	0.9461	0.9590	1.145	1.286	1.269	1.528	1.009
C	0.76	1.024	0.9538	0.9689	0.9472	0.8287	0.7143	0.5405	0.4853	0.9023
D	0.92	1.017	0.9833	1.125	0.8926	1.028	0.4714	0.9595	0.6417	0.9264
E	0.80	1.052	1.160	1.164	1.151	1.243	1.800	0.9653	0.6977	0.8940
F	0.80	1.142	0.9172	0.9482	0.9975	1.101	0.4286	0.5491	0.3070	0.8994
G	1.24	1.351	1.095	1.201	1.193	1.561	1.286	0.7861	0.6055	1.035
H	0.44	1.559	1.154	1.324	2.027	1.705	2.857	2.711	0.7369	1.216

表 3 各类土样的 x^2 距离系数矩阵

	A	B	C	D	E	F	G	H
A	0							
B	0.9349	0						
C	0.9622	1.493	0					
D	0.8035	1.294	0.5943	0				
E	1.539	1.169	1.304	1.390	0			
F	1.119	1.712	0.4555	0.6024	1.538	0		
G	1.389	1.263	1.191	1.146	0.8463	1.220	0	
H	3.414	2.790	3.457	3.333	2.380	3.575	2.824	0

质、F 砂土, G 壤土和 H 粘土. 原始数据如表 1 所示.

为便于分类和结果比较, 对原始数据进行处理. 定义一土壤中某重金属背景值指数 M'_{oi} 如下:

$$M'_{oi} = \frac{\text{土壤中某重金属 } Me_i \text{ 的背景值}}{\text{全地区某重金属 } Me_i \text{ 的总背景值}} \quad (7)$$

原始数据按式(7)处理得表 2.

设各类型土样的集合为:

$$N = \{x_1, x_2, \dots, x_n\} \quad n = 8$$

每类土样有 10 种重金属背景值指数 M'_{oi} 指标, 即:

$$x_i = (x_{i1}, x_{i2}, \dots, x_{im}), (m = 10, i = 1, 2, \dots, 8)$$

其中 x_{ik} 表示第 i 类土样第 k 种重金属的背景值指数 M'_{oi} . 具体聚类分析步骤如下.

1. 根据式(2)、(6)由表 2 数据建立 x^2 距离系数矩阵, 得表 3.

2. 设 $G = [N, E]$ 为连通的模图, 用 Kruskal 算法求出 G 的最小生成树 $T_n = [N, E']$. 其作图要点是:

(1) 将图 G 的 m 条边按从短至长的顺序依次编号为 e_1, e_2, \dots, e_m (假定 G 有 m 条边, n 个节点, 且 $m > n - 1$), 若有几条边长度都相等, 则这些边之间可任意排序;

(2) 作出图 G 中的 n 个孤立节点, 按 e_1, e_2, \dots, e_m 的顺序试着加边. 每次加入边 e_k 均要保持“无圈性”, 否则考虑加入下一条边 e_{k+1} ;

(3) 当加入 $n - 1$ 条边后, 即得到 G 的最小生成树. 按上述作图要点, 由表 3 得到其最小生成树如图 1 所示.

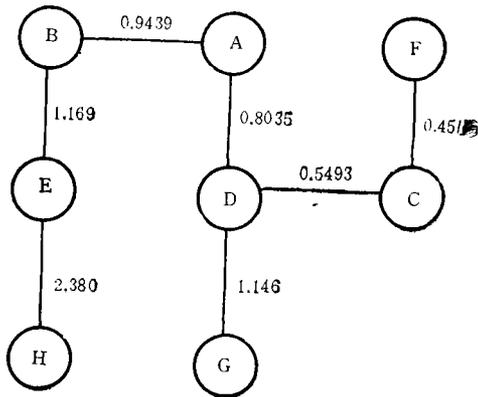


图 1 最小生成树

3. 根据图 1 所示的最小生成树, 取 $d = 1.000$, 将权重大于 1.000 的边切割, 得到如图 2 所示的分类树, 据此可对土样进行聚类分析.

三、结果与讨论

由图 2 所得的各类土壤样本的分类为: $\{B, A, D, C, F\}$, $\{E\}$, $\{G\}$, $\{H\}$. 参照表 1、表 2 数据可见 $\{B, A, D, C, F\}$ 代表了重金属背景值处于平均水平的土壤类型, $\{H\}$ 则属于典型的高重金属背景值土样. $\{E\}$ 、 $\{H\}$ 虽各单独归为一类, 但由表 3 数据可见 d_{BG} 的距离系数为 0.8463, 远低于 d_{EA} 、 d_{EB} 、 d_{EC} 、 d_{ED} 、 d_{EF} 、 d_{EH} , 故土样 E 和 G 的重金属

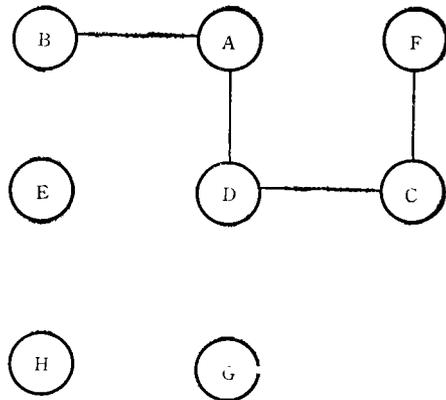


图 2 取 $d = 1.000$ 切割后所得的分类树

背景指数更为接近(相似程度更大), 故 $\{E\}$ 、 $\{G\}$ 可划归为一类 $\{E, G\}$. 参照表 1、表 2 可见 $\{E, G\}$ 代表了重金属背景值略偏高的土样.

由以上应用实例可见, 最小生成树法适用于环境科学领域的聚类分析问题, 具有较高的推广应用价值. 该法由于采用 x^2 距离系数描述样本间的相似关系, 避免了大量的矩阵复合运算, 使计算过程更为简便、快速. 以分类树图的形式给出聚类分析结果, 使计算结果直观可靠. 用该法对文献[1]、[2]给出的实例进行了验证, 其结果均与最小支撑树法和最大树的 Prim 算法相同. 当所分类的样本数较大时, 此聚类分析过程可按 Kruskal 算法编制的计算机程序进行.

参 考 文 献

- [1] Yoshioka, Y. et al., Ecotoxicol. Environ. Safety, (12), 15--21 (1986).
- [2] 于连生, 中国环境科学, 2(6), 33(1982).
- [3] 白希孟等, 中国环境科学, 5(6), 38(1985).
- [4] 何泳玲, 网络技术, 第 10—20 页, 上海交通大学出版社, 上海, 1989 年.
- [5] 王学仁, 地质数据多变量统计分析, 第 207—215 页, 科学出版社, 北京, 1982 年.
- [6] 王朝瑞, 图论, 第 79—88 页, 人民教育出版社, 北京, 1981 年.

(收稿日期: 1990 年 4 月 7 日)

Investigation on Environmental Radioactivity in Shenyang Area.

Li Shuqing et al. (Liaoning Provincial Institute of Labour Sanitation, Shenyang): *Chin. J. Environ. Sci.*, 11(6), 1990, pp. 67—70

This paper deals with the radioactive levels of gross β , ^{90}Sr , ^{137}Cs , ^{238}U , ^{232}Th , ^{226}Ra , ^{40}K in some environmental media in Shenyang area during 1981—1987. After Chernobyl accident on April 26, 1986, large amounts of gross β and ^{131}I were found in radioactive fallout on May 13 of that year, their values were 131 Bq/m².d and 281.2 Bq/m².d respectively. The highest value of ^{131}I in milk was 5.1 Bq/l. and that in spinach 48.1 Bq/kg. So it is shown that the environment was affected by Chernobyl accident. However, the residents' thyroid doses in the area were estimated, and the effective dose equivalents for adults were less than 2 μSv , and for infants less than 11 μSv not beyond the dose limits, according to "Basic Sanitary Standard for Radioactive Protection".

Key Words: environmental radioactivity, Chernobyl accident, Shenyang.

A Conception on Integrated Investigation of the Exterior Environment in the Environmental Impact Assessment (EIA) Projects.

Hou Huan, Pei Chenghu, Ping Yingdeng (Beijing Municipal Research Institute of Environmental Protection):

Chin. J. Environ. Sci., 11(6), 1990, pp. 70—72

Based on the practical experience of environmental impact assessment (EIA) projects, this paper presents a conceptional framework for analyzing an integrated investigation of the exterior environment in an EIA work. The framework involves two investigative dimensions: natural environment and social environment, in which attention should be paid respectively to some items so as to avoid certain mistakes.

Key Words: environmental impact assessment, practical experience, exterior environment.

A Simple Algorithm for Identification of Environmental Background Values.

Xin Shichong, Wang Guangguan (Jilin College of Chemical Engineering): *Chin. J. Environ. Sci.*, 11(6), 1990, pp. and for infants less than 11 μSv 73—75

In order to solve the cluster set problems in identification of environmental background values, a rapid and simple algorithm, the method of minimum spanning tree, is presented in this paper. The similar relation of sample variables is described with χ^2 distance coefficient matrix, and the minimum spanning tree is drawn in accordance with Kruskal method. The classified results by the algorithm are shown in a graph form. In this paper an applied case has also been given.

Key Words: cluster analysis, identification of environmental background value.

"Design Thinking of A Greening Project in and around A Petrochemical Complex"

Design Thinking of A Greening Project in and around A Petrochemical Complex. Chen Ziqin (Anqin Petrochemical Complex, Anhui Province): *Chin. J. Environ. Sci.*, 11(6), 1990, pp. 76—78

This article introduces the experience of a green project that has been performed in Anqin Petrochemical Complex for more than ten years. In order to purify ambient air around the Complex, the project was designed in line with different conditions of the branch factories where pollution sources varied from the processes of production. The design implicated a scientific choice of plant species and rational arrangement of planting trees and grass so as to attain a good environmental benefit. By means of greening work, a clean fine working and living environment has been successfully created.

Key Words: petrochemical plant, design of green project, choice of tree species.