

# 土壤质量聚类分析——以封丘县为例

蔡瞳<sup>1,2</sup>, 徐惠, 吴群 (1. 南京农业大学土地管理学院, 江苏南京210095; 2. 江苏苏信房地产评估咨询有限公司, 江苏南京210018)

**摘要** 介绍了聚类分析的原理、方法和步骤; 并以封丘县土壤质量为例, 对其进行聚类分析, 将封丘土壤质量分为4类, 研究了其各因素状况, 在此基础上提出改进措施和方法。

**关键词** 土壤质量; 聚类分析; 封丘县; SPSS

中图分类号 S11+9 文献标识码 A 文章编号 0517-6611(2008)25-10998-02

## Clustering Analysis on the Soil Quality in Fengqiu County

CAI Tong et al (College of Land Management, Nanjing Agricultural University, Nanjing, Jiangsu 210095)

**Abstract** The principles, methods and steps of clustering analysis were introduced. Taking soil quality in Fengqiu County as an example, the clustering analysis was made on it. The soil quality in Fengqiu County was clustered into four types. The situations of each factor were studied and some improving measures and methods were put forward on this basis.

**Key words** Soil quality; Clustering analysis; Fengqiu County; SPSS

对土壤的研究是一项具有重要科学和生产意义的课题。一方面通过土壤研究可正确评价土壤质量, 另一方面可为土壤的整治、规划和合理利用提供科学依据。土壤科学的基础是土壤分类, 不同时期所拟定的土壤分析系统反映了该时期人们对土壤的认识水平和土壤科学本身的发展阶段。

以往的土壤质量评价多侧重于定性和单因素评价, 目前, 随着数值化评价方法的提出与进一步应用, 以及模糊数学方法和多元统计方法等的渗透, 土壤的评价研究已经由单纯的形态描述向指标化和定量化方向发展。聚类分析法作为现代多元统计中的一种模式识别方法, 由于为土壤评价提供了科学的方法, 近年来越来越广泛地用于土壤研究中。为了让更多的相关工作者了解聚类分析方法在土壤环境研究中的应用, 笔者针对聚类分析的原理、构造及应用前景等进行了综述, 并对封丘县土壤质量进行了聚类分析。

### 1 聚类分析原理

聚类分析的基本思想是认为所研究的数据集中的数据或者属性之间存在着程度不同的相似性。于是从数据集中取出一批数据, 具体找出一些能够度量数据值之间或者属性之间相似程度的量, 以这些量为中心作为划分类型的依据, 把一些相似程度较大的数据或属性聚合为一类, 把另外一些彼此之间相似程度较大的样品又聚合为另一类, 关系密切的聚合到一个小的分类单位, 关系疏远的聚合到一个大的分类单位, 直到所有数据或属性都聚合完毕, 把不同的类型一一划分出来<sup>[1]</sup>。

聚类分析是指对一组数据的群聚结构在无任何先验知识时, 根据样本间的距离与相似程度将样本分类, 包括以下3方面基本内容: 样本间相似性或不相似性的度量; 行成簇的算法; 聚类准则的确定<sup>[2]</sup>。有关聚类分析的方法比较多, 但用于土壤研究的聚类方法多为模糊聚类法和灰色聚类法。所谓模糊聚类分析法是把欲分类的对象, 如监测点、污染因子等作为样本, 这些样本之间存在多元模糊关系, 通过

模糊等价关系变换, 定量地确定各样本之间的亲疏关系, 从而对样本进行科学分类的方法<sup>[3]</sup>。由于事物本身在很多情况下都带有模糊性, 因此把模糊数学方法引入聚类分析, 就能使聚类分析更切合实际, 克服了靠单一因素进行分类的缺陷<sup>[4]</sup>。而灰色聚类法<sup>[5]</sup>是在聚类分析方法中, 引进灰色理论的白化函数。它将聚类对象(评价对象)对不同聚类指标所拥有的白化数, 用特定的数学模式按几个灰类进行归类, 最后得到聚类对象所属的分类级别的一种方法。但模糊聚类和灰色聚类法有时可能会出现分辨率低的问题, 所以还必须找到能弥补这个不足的方法, 而聚类分析却拥有他们不足的优点, 因而开始大规模应用于土壤质量的研究上。

### 2 聚类分析的方法和步骤

聚类的实质就是使属于同一类别的个体之间的距离尽可能地小, 而不同类别的个体间的距离尽可能地大。对于某一领域来说, 该领域一定包含许多东西, 把每一种东西当成一个样品, 并连同它的属性一起定义成一个独立的类。它的属性有的是用数来度量, 即定量的属性; 有的用文字来表示, 即定性的属性。需要把所有的属性抽象成一个数据阵。用数表示的属性直接提取具体的数据即可。对于用文字表示的属性按某种一对一的映射关系, 把它们映射成具体的数据, 这样就可以实现属性的数据化。用下面的映射关系:  $A \rightarrow B$ , 其中  $f$  表示一个映射函数;  $A$  是文字属性的集合, 对于不同属性的个数为  $n$ , 则  $B$  集合包含从1到  $n$  的  $n$  个自然数。按某个标准并根据属性所表示的含义的程度来与这  $n$  个数进行对应。对应的关系是一一对一或多对一的映射关系。程度最小的与1相对应, 程度最深的与  $n$  相对应。这样就可以把所有属性抽象成一个数据矩阵<sup>[6]</sup>。

此方法的基本步骤是:

第1步先将所有样本各自看成一类, 计算样本与样本之间的距离和类与类的距离。其中距离计算方法有3种: 绝对值距离法、欧氏距法和明斯基距离法<sup>[7]</sup>。以下采用绝对值距离法。

第2步, 合并距离最近的两类为一个新类, 新类的属性基于原类的属性构成。如果合并的新类中有一个公共类则这两个新类为一类。在实际生活中, 土壤的种类和名称很多, 首先把一个地块看成一个样品, 每一个地块定义为一个

基金项目 教育部哲学社会科学重大课题攻关项目: “工业化、城市化进程中的农村土地问题研究”(04JZD008)。

作者简介 蔡瞳(1985-), 男, 江苏泗洪人, 硕士研究生, 研究方向: 不动产评估与管理。

收稿日期 2008-06-16

独立的类。它有 $j$ 个属性,也就是每个样品有 $j$ 个变量, $x_j$ 表示属性。可以把许多地方的土壤的属性抽象化成一个数据阵,其中 $x_{ij}$ 表示第 $i$ 个样品的第 $j$ 个变量的值<sup>[8]</sup>。假设以 $n$ 个地块,每个地块的土壤有 $m$ 个属性为例。

第3步,把新类看成新的样品并提取出相应的属性值,同样采用绝对值距离公式来计算。不断计算新类与当前各类的距离并进行合并,直到类的个数为1。

第4步,作快速聚类表,确定类。

### 3 封丘县土壤质量的聚类分析

封丘县境南北长38.2 km,东西宽48.7 km,面积1 220.5 km<sup>2</sup>,耕地面积6.2 万hm<sup>2</sup>;辖3个镇22个乡,60个村民委员会,627个自然村,69.39万人;地处黄河故道,地貌复杂,沙

岗、平原、洼地兼有,黄河大堤以南滩地较高,其余地势低洼,地势由西南向东北倾斜;海拔65~72 m;黄河从县南和县东流过,境内流长56 km。

以封丘县的土壤质量为例,选取134个地块,每个地块都选取相同的能准确反映土壤质量的9个农艺属性:有机质、全N、全P、全K、碱解N、速效P、速效K、CEC、电导率。把这134个地块的属性抽象为数据矩阵,然后利用绝对值距离法<sup>[9]</sup>进行聚类分析。由于数据量大,借助SPSS软件的快速聚类法进行了计算,合并以后得到的最终聚类中心表见表1。

### 4 聚类结果分析

根据以上的最终聚类中心表可以得出4类土壤质量分类,见表2。

表1 聚类中心表

Table 1 The clustering center

聚类	有机质 g/kg	全N g/kg	全P g/kg	全K g/kg	碱解N mg/kg	速效P mg/kg	速效K mg/kg	CEC	电导率 ms/cm
Cluster	Organic matter	Total N	Total P	Total K	Alkaline hydrolytic N	Available P	Available K	cmol/kg	Electric conductivity
1	14.10	0.81	0.75	21.52	52.97	100.05	10.43	11.58	778
2	14.34	0.82	0.76	21.96	53.59	91.45	8.23	11.38	635
3	12.68	0.74	0.77	21.90	50.12	85.97	9.36	11.08	155
4	11.81	0.73	0.73	21.19	50.54	68.82	6.31	9.00	509

表2 封丘土壤质量分类

Table 2 The classification of the soil quality in Fengqiu County

土壤质量分类	地块点
Classification of soil quality	Block sites
第1类 The first class	15, 19, 23, 41, 52, 55, 65, 81, 87, 88, 89, 97, 98, 102, 108, 116, 117, 129
第2类 The second class	2, 3, 4, 8, 13, 14, 16, 20, 21, 22, 24, 26, 32, 33, 40, 44, 45, 46, 47, 49, 50, 56, 59, 64, 69, 70, 71, 72, 73, 74, 76, 83, 84, 95, 100, 103, 105, 106, 107, 109, 110, 114, 115, 119, 121, 124, 126, 128, 130, 132, 134
第3类 The third class	10, 11, 17, 18, 35, 36, 37, 38, 39, 58, 60, 86, 99, 133
第4类 The fourth class	1, 5, 6, 7, 9, 12, 25, 27, 28, 29, 30, 31, 34, 42, 43, 48, 51, 53, 54, 57, 61, 62, 63, 66, 67, 68, 75, 77, 78, 79, 80, 82, 85, 90, 91, 92, 93, 94, 96, 101, 104, 111, 112, 113, 118, 120, 122, 123, 125, 127, 131

由于目前对土壤质量的评价还没有统一的标准,为此国家“973计划”中的“土壤质量演变规律与持续利用”课题组在充分调研的基础上提出了我国土壤质量指标体系的初步建议方案,该研究采用的标准是其中的绿色食品地土壤肥力分级。

第1类土壤主要由18个地块组成,其中有机质在10.00~20.00 g/kg,全N在0.75~1.00 g/kg,全P在0.70~1.00 g/kg,全K在20.00~30.00 g/kg,碱解N在30.00~60.00 mg/kg,速效P在50.00 mg/kg以上,速效K在5.00~10.00 mg/kg,CEC在10.00~20.00 cmol/kg。该类土壤质量尚可,要广辟有机肥源,增施有机肥。

第2类土壤主要由51个地块组成,其中有机质在5.00~15.00 g/kg,全N在0.75~1.00 g/kg,全P在0.50~1.00 g/kg,全K在20.00~30.00 g/kg,碱解N在30.00~60.00 mg/kg,速效P在70.00 mg/kg以上,速效K基本都在5.00 mg/kg以下,CEC在5.00~20.00 cmol/kg。该类土壤质量较差,要重视有机肥和K肥施用,协调NPK比例。

第3类土壤主要由14个地块组成,其中有机质在15.00

~25.00 g/kg,全N在0.50~1.00 g/kg,全P在0.80~1.00 g/kg,全K在20.00~25.00 g/kg,碱解N在40.00~60.00 mg/kg,速效P在70.00 mg/kg以上,速效K在6.00~12.00 mg/kg,CEC在10.00~20.00 cmol/kg。该类土壤质量优良,要大力推广平衡施肥。

第4类土壤主要由51个地块组成,其中有机质在5.00~15.00 g/kg,全N在0.50~1.00 g/kg,全P在0.60~1.00 g/kg,全K在20.00~25.00 g/kg,碱解N在30.00~60.00 mg/kg,速效P在40.00 mg/kg以上,速效K在10.00~15.00 mg/kg,CEC在5.00~10.00 cmol/kg。该类土壤质量较差,要加大各类养分的施放力度,并同时要发展预报、控制和利用肥料的方法,增加肥料的有效利用率,每3年对土壤质量进行一次评估,对土壤中各种营养元素的含量进行测评,以正确施放肥料,提高肥料的利用效率。

总体来说,封丘县的土壤质量一般,在因地制宜的同时,对第2、4类地块还需加大治理力度,增加资金投入,改善土壤环境,提高农作物产量,增加当地政府的财政收入。

### 5 问题与展望

近年来,由于聚类分析对样本进行科学分类具有独到之处,因而在土壤环境质量评价研究中的应用逐广,许多土壤研究者与环境工作者运用聚类分析在土壤质量评价中开展了卓有成效的工作。但由于土壤研究的复杂性,在运用聚类分析处理土壤问题这一新领域时,还存在主观随意性,分辨率和信息利用率不高<sup>[10]</sup>,多因子的综合评价精度较低等问题,对这些问题的研究工作还有待加强,并和其他科学方法相结合。聚类分析作为一种模式识别方法,在识别某些性质相近的问题时,具有较大的优势。对于复杂土壤数据处理,用传统的方法,往往得到的结果不理想,或方法繁琐。聚类分析近年在土壤研究中取得的结果充分说明其应用前景将

(下转第11001页)

显著性差异水平,而各个施钾处理之间大葱产量并没有显著性差异。

表3 不同钾肥处理的大葱产量

Table 3 The yield of scallion under different K fertilizer treatments

处理 Treatment	小区产量 Hot yield kg				折合产量 kg/hm <sup>2</sup> Converted yield	较对照 ± % Compared with CK
	重复1 Repeat 1	重复2 Repeat 2	重复3 Repeat 3	平均 Average		
1-NP	132.11	120.41	126.02	126.18	63 090.0	0
2-NPK1	137.07	143.09	140.44	140.20	70 100.0	11.1
3-NPK2	142.68	148.05	141.76	144.16	72 081.7	14.3
4-NPK3	144.57	146.60	142.11	144.43	72 213.3	14.5

表4 Duncan 法多重比较结果

Table 4 The multiple comparison results by Duncan method

处理 Treatment	均值 Mean	5% 显著水平 5% significant level	1% 极显著水平 1% extremely significant level
4	144.43	a	A
3	144.16	a	A
2	140.20	a	A
1	126.18	b	B

2.3 不同施钾处理对大葱生产收益的影响 按大葱0.6元/kg,氯化钾肥料2.1元/kg,各处理经济效益情况见图1。

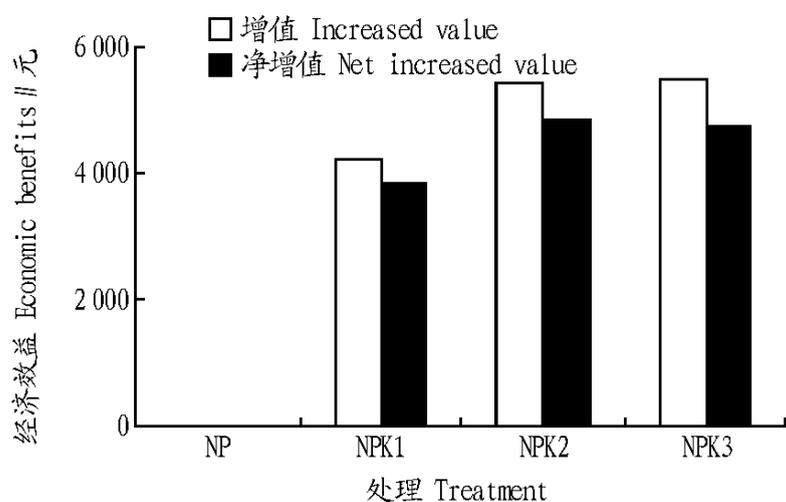


图1 施钾对大葱生产经济效益的影响

Fig.1 The effects of applying K on the economic benefits of scallion production

大葱施用氯化钾肥料后增产值4 206 ~5 474 元/hm<sup>2</sup>,净增产值828 ~4 844 元/hm<sup>2</sup>。增产值处理4 最高,处理3 次之,但净增产值处理3 最高。从产出投入比看,处理2、3 的产出投入比分别为11.1、9.8,施用钾肥的回报率很高,虽然处理3 产出投入比小于处理2,但钾肥的多投入,增加了1 000 元/hm<sup>2</sup>以上的收益。说明在当地土壤养分条件下和一定氮磷基础上,施钾262.5 kg/hm<sup>2</sup>,增产效果最好,并且经济效益最佳。

### 3 结论

(1) 种植大葱在氮磷基础上增施钾肥,一定用量范围内对大葱有增产作用,随着钾肥用量的增加,大葱产量也不断增加,但超过一定量大葱增产幅度明显减小。

(2) 在试验条件下,各施钾处理比不施钾处理增产达到极显著性差异水平。而各个施钾处理之间大葱产量没有显著性差异。

(3) 在当地土壤养分条件下及一定氮磷基础上,施钾262.5 kg/hm<sup>2</sup>,增产效果最好,并且经济效益最佳。

### 参考文献

- [1] S-L 蒂斯代夫,LU-L·纳尔逊,J·D·毕腾.土壤肥力与肥料[M].金继运,刘荣乐,译.北京:中国农业科技出版社,1998:213.
- [2] 金继运,自由路,杨丽苹,等.高效土壤养分测试技术与设备[M].北京:中国农业出版社,2006:74-87,148-153.
- [3] 何其伟,苏德恕,赵德婉,等.山东名产蔬菜[M].济南:山东科学技术出版社,1990:178-180.
- [4] 从同林,王在超,刘卫东,等.钾肥对大葱的增产效果及适宜用量[J].山东农业科学,1996(4):26.
- [5] 王在超,刘卫东.大葱钾肥施用指标初探[J].中国蔬菜,1997(3):25.
- [6] 邓聚龙.灰色系统:社会、经济[M].北京:国防工业出版社,1985.
- [7] HICHEL J, SAWERRHI, CZARNOVSKA K. Spatial and temporal distribution of metals in soils in Warsaw Poland[J]. Environmental Pollution, 1997, 98(2): 169-174.
- [8] 万存绪,张效勇.模糊数学在土壤质量评价中的应用[J].应用科学学报,1991(4):359-365.
- [9] 邓志鸿,唐世渭,张铭,等.Ordogy 研究综述[J].北京大学学报,2002(5):730-738.
- [10] 何晓群.现代统计分析方法与应用[M].北京:中国人民大学出版社,1998.
- [11] 唐守正.多元统计分析[M].北京:中国林业出版社,1989.

(上接第10999页)

会十分广阔。

### 参考文献

- [1] 郑丽萍,梁永全.基于聚类分析法的本体构造方法[J].青岛大学学报,2005(3):55-58.
- [2] 刘凡丰.聚类分析在土壤研究中的应用[J].四川理工学院学报,2005,6(2):66-72.
- [3] 奚旦力,孙裕生,刘秀英.环境监测[M].修订版.北京:高等教育出版社,1996.
- [4] 吴万铎.模糊数学与电子计算机应用[M].北京:电子工业出版社,