

# 模糊聚类分析在耕地分类中的应用

陈争光<sup>1</sup>, 冯小娟<sup>2</sup>

(1. 黑龙江八一农垦大学, 大庆 163319; 2. 大庆油田技术培训中心, 大庆 163254)

**摘要:** 简单地介绍了模糊聚类分析方法, 并使用此方法, 以黑龙江农垦总局农场的耕地理化指标为依据, 对耕地进行分类, 并将分类结果和模糊物元分析方法所得的分析结果进行对照, 两者结果一致, 说明使用模糊聚类分析对耕地进行分类具有一定的实用价值。

**关键词:** 模糊聚类; 耕地; 分类

中图分类号: S155.4      文献标识码: A      文章编号: 1002-2767(2008)04-0042-03

## Application of Fuzzy Cluster Analysis on Classification of Cultivated Land

CHEN Zheng-guang<sup>1</sup>, FENG Xiao-juan<sup>2</sup>

(1. Heilongjiang August First Land Reclamation University, Daqing 163319; 2. Daqing Oil Field Technology Training Center, Daqing 163254)

**Abstract:** Fuzzy Cluster Analysis is introduced in this paper, and the Cultivated Land in one of the farms of Heilongjiang Land Reclamation Bureau was classified according to the land's physical and chemical index data. The results of fuzzy cluster analysis was consistent with that of fuzzy matter element analysis which showed some practicality of the application of fuzzy cluster analysis on classification of cultivated land.

**Key words:** Fuzzy Cluster Analysis; cultivated land; classification

耕地是农作物生产的基地, 是农业的基本生产资料, 是人类耕作劳动的对象, 与社会经济紧密联系。耕地的本质是肥力, 耕地的肥力是其各方面性质的综合反映, 体现了其在农业生产和科学研究中的重要地位。耕地肥力的高低直接影响着作物生长, 影响着农业生产的结构、布局和效益等方面。如何科学、合理、实用地评价耕地肥力, 为指导农业生产提供理论依据, 显得尤为重要。耕地肥力的评价指标很多, 大体上从反映耕地肥力的化学指标、物理性状、生物特征和土壤的环境条件出发进行肥力研究<sup>[1]</sup>。当然对于不同的地域, 评价耕地肥力的指标的选取可能有所侧重。本文以黑龙江农垦总局某农场耕地的部分化学指标和生物指标为依据, 使用模糊聚类的方法, 对该地区耕地进行了分类。

### 1 模糊聚类分析方法简介

聚类就是把具有相似性质的事物区分开来并加以分类, 由于耕地信息本身具有模糊性, 一块土地是

否属于某一类并不是泾渭分明, 这些模糊的概念便产生了经典分类学解决不了的困难, 因而模糊聚类分析就应运而生。应用模糊数学的理论和进行分析的聚类分析称为模糊聚类分析。应用模糊聚类分析得到的结果往往更符合实际<sup>[2]</sup>。模糊聚类分析的过程一般分为以下几个步骤<sup>[3-4]</sup>：

#### ① 建立模糊相似矩阵

假设被分类对象有  $n$  个, 每个对象有  $m$  个特性指标, 于是可以得到  $n \times m$  原始数据矩阵  $X$ ：

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

(1)

由于  $n$  个对象的  $m$  个特性指标的量纲和数量级不同, 不能直接使用原始数据(式 1)进行分类。必须首先对原始数据进行无量纲化处理, 使每一类指标值统一到共同的数据范围内, 这个过程就是数据的标准化。标准化工作在每一个指标数据内部进行, 也就是在原始数据矩阵  $X$  的每列上独立进行。首先按下式对原始数据矩阵  $X$  进行变换:

$$x'_{ik} = \frac{x_{ik} - \overline{x_k}}{\sigma_k} (i = 1, \cdots, n; k = 1, \cdots, m)$$

(2)

收稿日期: 2008-03-20  
第一作者简介: 陈争光(1973-), 男, 湖北黄陂人, 讲师, 主要从事计算在农业领域应用技术的研究和教学工作。 Tel: 13945606175; E-mail: ruzee@sin.com.

42  
黑龙江农业科学

其中:  $\overline{x_k}$  是  $n$  个对象的第  $k$  个指标的平均值,  $\sigma_k$  是  $n$  个对象的第  $k$  个指标的均方差, 其计算公式分别如下:

$$\overline{x_k} = \frac{1}{n} \sum_{i=1}^n x_{ik} \tag{3}$$

$$\sigma_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ik} - \overline{x_k})^2} \tag{4}$$

经过上述变换之后, 每个指标的均值为 0, 均方差为 1, 且消除了量纲的影响。但是,  $x'_{ik}$  不一定在区间  $[0, 1]$  上。此时, 可以通过下面的极差变换使其满足要求:

$$x''_{ik} = \frac{x'_{ik} - \min_{1 \leq i \leq n} \{x'_{ik}\}}{\max_{1 \leq i \leq n} \{x'_{ik}\} - \min_{1 \leq i \leq n} \{x'_{ik}\}} \quad (k = 1, \dots, m) \tag{5}$$

上述过程称为数据的标准化, 完成这一工作之后, 下一步就是依据标准化的数据矩阵建立模糊相似矩阵  $R = (r_{ij})_{n \times m}$ 。矩阵  $R$  中数据元素  $r_{ij}$  表示了第  $i$  个对象和第  $j$  个对象之间的相似程度。确定  $r_{ij}$  的方法很多, 比如相似系数法、距离法等, 每一类方法又分为很多具体方法。根据耕地分类的实际要求, 本文选用相似系数法中的数量积法。由标准矩阵  $X = (x_{ij})_{n \times m}$  建立相似矩阵的数量积法的公式如下:

$$r_{ij} = \begin{cases} 1 & (i = j) \\ \frac{1}{M} \sum_{k=1}^m x_{ik} \times x_{jk} & (i \neq j) \end{cases} \tag{6}$$

其中:  $M = \max_{i \neq j} \{ \sum_{k=1}^m x_{ik} \times x_{jk} \}$ 。

至此, 模糊相似矩阵建立起来了。

②聚类

一般说来, 依据上述方法建立起来的模糊矩阵只是一个相似矩阵  $R$ , 不一定是一个模糊等价矩阵, 即矩阵  $R$  具有自反性和对称性, 但是不具备传递性。为了进行分类, 须将  $R$  改造为模糊等价矩阵  $R^*$  (实际上, 也可直接在相似矩阵  $R$  上直接进行分类, 但是基于模糊等价矩阵的分类方法理解起来方便, 实现也相对容易, 因此本文使用此方法)。使用传递闭包的方法将相似矩阵改造为等价矩阵。其依据是下面的定理:

定理: 若  $R$  是  $n$  阶模糊相似矩阵, 则存在一个最小的自然数  $k (k \leq n)$ , 使得传递闭包  $t(R) = R^k$ , 对于一切大于  $k$  的自然数  $l$ , 恒有  $R^l = R^k$ , 此时,  $t(R)$  为模糊等价矩阵。

注意, 定理中的  $R^k = R^{k-1} \circ R$ , 即  $r_{ij}^k = \bigvee_{k=1}^m (r_{ik}^{(k-1)} \wedge r_{jk})$ 。这也为我们从模糊相似矩阵出发求模糊等价矩阵提供了计算公式。

在实际应用中, 使用的是平方法求等价矩阵, 其

方法是将模糊矩阵自乘  $R \circ R = R^2$ , 得到  $R^2$ , 然后再自乘  $R^2 \circ R^2 = R^4, \dots$ , 继续下去, 直到某次  $R^{2k} = R^k$  时为止, 此时  $t(R) = R^k$  即为所求。

有了  $t(R)$  之后, 下一步就是动态聚类。动态聚类的过程就是求模糊等价矩阵  $R^*$  的  $\lambda \in [0, 1]$  水平截集的过程<sup>[5]</sup>,  $\lambda$  水平截集  $R_\lambda$  是一个布尔矩阵, 也是一个等价矩阵。其元素为:

$$r_{ij}(\lambda) = \begin{cases} 1 & r_{ij} \geq \lambda \\ 0 & r_{ij} < \lambda \end{cases} \tag{7}$$

当  $\lambda$  从 1 逐渐变化到 0 时,  $R_\lambda$  不断变化, 这个过程形成了一个动态聚类关系。

2 应用实例(农场耕地聚类分析)

现有黑龙江农垦总局某农场 12 块耕地土壤理化指标的测量值(见表 1)。聚类分析的数据来自于耕地速效氮、速效磷、速效钾、有机质、pH 的等五项测量指标。

表 1 耕地理化指标的测量值

地块编号	耕地土壤的理化指标				
	速效氮	速效磷	速效钾	有机质	pH
1	177.59	36.35	116.88	39.41	5.79
2	170.04	31.95	77.66	34.42	5.78
3	210.46	31.61	153.91	52.08	6.17
4	182.11	30.03	114.75	40.69	5.75
5	181.92	30.40	99.02	38.24	6.01
6	239.25	31.92	145.27	56.02	4.71
7	213.08	35.15	132.94	45.42	5.94
8	179.75	32.87	117.03	40.66	5.97
9	166.72	29.86	91.45	33.69	5.67
10	197.18	33.36	130.69	41.69	5.84
11	261.47	30.13	125.56	75.37	5.84
12	216.40	29.91	144.72	50.85	5.85

表 1 中五项指标的前三个的单位是  $\text{mg} \cdot \text{kg}^{-1}$ , 有机质的单位是  $\text{g} \cdot \text{kg}^{-1}$ , pH 是无量纲值。如果统一到同一个量纲上来, 那么有机质含量和其他数据之间的数量级相差太大, 会造成大特性指标排斥小特性指标对分类的作用。因此, 需要对数据进行标准化处理。标准化方法很多, 在此采用上面介绍的相似系数法(式 2)。标准化之后的数据使用数量积法(式 6)得到模糊相似矩阵(见表 2)。

在聚类之前需要将相似矩阵转换为等价矩阵, 转换方法使用平方法求矩阵的传递闭包。表 2 所示的模糊相似矩阵计算结果为:  $R^4 = R^8$ 。表 3 即为  $t(R) = R^4$ 。由表 3 可见, 此时的传递闭包为等价关系。

表 2 对数据进行标准化之后再利用数量积法得到的相似矩阵

1	0.375	0.706	0.365	0.398	0.403	0.823	0.607	0.250	0.662	0.515	0.499
0.375	1	0.363	0.232	0.296	0.060	0.389	0.340	0.207	0.326	0.274	0.259
0.706	0.363	1	0.583	0.568	0.673	0.922	0.708	0.361	0.798	1.000	0.900
0.365	0.232	0.583	1	0.353	0.282	0.475	0.400	0.240	0.426	0.513	0.491
0.398	0.296	0.568	0.353	1	0.197	0.489	0.429	0.275	0.434	0.493	0.464
0.403	0.060	0.673	0.282	0.197	1	0.615	0.345	0.069	0.490	0.806	0.605
0.823	0.389	0.922	0.475	0.489	0.615	1	0.688	0.296	0.781	0.827	0.724
0.607	0.340	0.708	0.400	0.429	0.345	0.688	1	0.285	0.585	0.570	0.550
0.250	0.207	0.361	0.240	0.275	0.069	0.296	0.285	1	0.274	0.270	0.291
0.662	0.326	0.798	0.426	0.434	0.490	0.781	0.585	0.274	1	0.680	0.635
0.515	0.274	1.000	0.513	0.493	0.806	0.827	0.570	0.270	0.680	1	0.905
0.499	0.259	0.900	0.491	0.464	0.605	0.724	0.550	0.291	0.635	0.905	1

表 3 模糊等价矩阵

1	0.389	0.823	0.583	0.568	0.806	0.823	0.708	0.361	0.798	0.823	0.823
0.389	1	0.389	0.389	0.389	0.389	0.389	0.389	0.361	0.389	0.389	0.389
0.823	0.389	1	0.583	0.568	0.806	0.922	0.708	0.361	0.798	1	0.905
0.583	0.389	0.583	1	0.568	0.583	0.583	0.583	0.361	0.583	0.583	0.583
0.568	0.389	0.568	0.568	1	0.568	0.568	0.568	0.361	0.568	0.568	0.568
0.806	0.389	0.806	0.583	0.568	1	0.806	0.708	0.361	0.798	0.806	0.806
0.823	0.389	0.922	0.583	0.568	0.806	1	0.708	0.361	0.798	0.922	0.905
0.708	0.389	0.708	0.583	0.568	0.708	0.708	1	0.361	0.708	0.708	0.708
0.361	0.361	0.361	0.361	0.361	0.361	0.361	0.361	1	0.361	0.361	0.361
0.798	0.389	0.798	0.583	0.568	0.798	0.798	0.708	0.361	1	0.798	0.798
0.823	0.389	1	0.583	0.568	0.806	0.922	0.708	0.361	0.798	1	0.905
0.823	0.389	0.905	0.583	0.568	0.806	0.905	0.708	0.361	0.798	0.905	1

对于表 3 中不同的 $\lambda \in [0, 1]$ , 当 $\lambda$ 从 1 变化到 0 时的水平截集 $R_\lambda$ 形成动态聚类过程如下:

取 $\lambda = 1.000$ , 分为 11 类,  $\{11, 3\}, \{1\}, \{2\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9\}, \{10\}, \{12\}$ ;

取 $\lambda = 0.922$ , 分为 10 类,  $\{11, 3, 7\}, \{1\}, \{2\}, \{4\}, \{5\}, \{6\}, \{8\}, \{9\}, \{10\}, \{12\}$ ;

取 $\lambda = 0.905$ , 分为 9 类,  $\{11, 3, 7, 12\}, \{1\}, \{2\}, \{4\}, \{5\}, \{6\}, \{8\}, \{9\}, \{10\}$ ;

取 $\lambda = 0.823$ , 分为 8 类,  $\{11, 3, 7, 12, 1\}, \{2\}, \{4\}, \{5\}, \{6\}, \{8\}, \{9\}, \{10\}$ ;

取 $\lambda = 0.806$ , 分为 7 类,  $\{11, 3, 7, 12, 1, 6\}, \{2\}, \{4\}, \{5\}, \{8\}, \{9\}, \{10\}$ ;

取 $\lambda = 0.798$ , 分为 6 类,  $\{11, 3, 7, 12, 1, 6, 10\}, \{2\}, \{4\}, \{5\}, \{8\}, \{9\}$ ;

取 $\lambda = 0.708$ , 分为 5 类,  $\{11, 3, 7, 12, 1, 6, 10, 8\}, \{2\}, \{4\}, \{5\}, \{9\}$ ;

取 $\lambda = 0.583$ , 分为 4 类,  $\{11, 3, 7, 12, 1, 6, 10, 8, 4\}, \{2\}, \{5\}, \{9\}$ ;

取 $\lambda = 0.568$ , 分为 3 类,  $\{11, 3, 7, 12, 1, 6, 10, 8, 4, 5\}, \{2\}, \{9\}$ ;

取 $\lambda = 0.389$ , 分为 2 类,  $\{11, 3, 7, 12, 1, 6, 10, 8, 4, 5, 2\}, \{9\}$ ;

取 $\lambda = 0.361$ , 分为 1 类,  $\{11, 3, 7, 12, 1, 6, 10, 8, 4, 5, 2, 9\}$ ;

聚类过程随 $\lambda$ 的取值从大到小变化, 聚类类别越来越少。当 $\lambda = 0.361$ 时, 12 个地块合为一类, 该类地块编号的顺序正好是 $\lambda$ 从大到小变化时地块编号的加入的顺序。我们同时使用模糊物元分析方法对这 12 块耕地的地力等级进行了分析, 这 12 块地的地力等级从大到小的顺序依次为: 11、3、6、12、7、10、1、8、4、5、2、9。由此可见, 由模糊物元分析得到的顺序和使用模糊聚类分析得到的聚类顺序基本一致, 从一个方面说明聚类分析在耕地划分中的有效性和实用性。

3 结束语

模糊聚类分析在进行耕地分类、确定耕地等级方面具有客观性和有效性。为建立科学的耕地等级体系具有非常重要的意义。

参考文献:

[ 1 ] 骆东奇, 白洁, 谢德体. 论土壤肥力评价指标和方法[ J ]. 土壤与环境, 2002, 11(2): 202-205.  
[ 2 ] 王新洲, 史文中, 王树良. 模糊空间信息处理[ M ]. 武汉: 武汉大学出版社, 2002(10): 59.  
[ 3 ] 谢季坚, 刘承平. 模糊数学方法及其应用[ M ]. 2 版. 武汉: 华中科技大学, 2004.  
[ 4 ] 梁榕辉, 黄文锋. 模糊聚类分析法及计算机实现[ J ]. 成组技术与生产现代化, 1997(4): 41-44.  
[ 5 ] 李吉鸿. 模糊数学基础及使用算法[ M ]. 北京: 科学出版社, 2005.