DOI:10.13758/j.cnki.tr.1985.03.007

模糊(FUZZY)聚类分析在土壤资源评价制图中的应用*

周慧珍

（中国科学院南京土壤研究所）

一、问题的提出

土壤资源图的编制是以土壤图为基础着手进行的。编图前，首先要拟定出土壤资源的评价原则和选定评价项目[1—3]。应当把土壤图上每块图斑所包含的土壤类型的质量好坏，看作是若干个用评价项目来表示的因子对它综合影响的结果。评价项目及其所含的具体内容是进行土壤资源评价的信息。土壤资源评价的信息可以分为强信息和弱信息两种。所谓强信息，就是指具有精确统计数据的那些信息。例如有机质含量1.2%，1.5%……等，全磷含量0.05%，0.08%……等。所谓弱信息，通常指一些边界数据不明确，或用文字描述的信息。例如，产量低、中、平、高平、高。土壤质地描述：砂壤，砂壤—轻壤，轻壤，重壤—轻粘，轻粘—中粘，中壤，轻壤—中壤……等。在土壤资源评价工作的研究中，我们经常而且主要要利用这些定性描述的弱信息，但这些弱信息多般具有模糊的特性，经典数学中的统计学理论无法利用这些弱信息对土壤资源进行分类评价。那么，如何把这些定性的弱信息进行定量化，再用一种数学的模式进行土壤资源评价制图？这就提出了应用模糊数学的理论对土壤资源进行聚类分等。

二、应用模糊聚类进行土壤资源评价制图的原理

设给定论域U，U到［0，1］闭区间的任一映射μ_A

$$\mu_A: U \to [0, 1]$$

$$\mu \to \mu_A(\mu)$$

都确定U的一个模糊子集$\underset{\sim}{A}$，μ_A叫$\underset{\sim}{A}$的隶属函数，$\mu_A(\mu)$叫作μ对$\underset{\sim}{A}$的隶属度。

模糊子集$\underset{\sim}{A}$完全由其隶属函数所刻划。特别当μ_A的值域取［0，1］闭区间的两个端点，亦即{0，1}两个值时，$\underset{\sim}{A}$便退化为一个普通子集，隶属函数也就退化为特征函数[4]。

模糊聚类则是模糊等价关系在不同水平α上的图象。模糊关系$\underset{\sim}{R}$必须是模糊等价关系才能聚类。

设给定U上的一个模糊关系$\underset{\sim}{R} = (r_{ij})_{n\times n}$如果它满足以下三点：

（1）自反性$r_{ij} = 1$。

（2）对称性$r_{ij} = r_{ji}$　　（i，j=1，2……n）。

* 本文承赵其国　何同康、陈志诚、徐琪等同志的指正，特此感谢。

(3) 传递性 $\underset{\sim}{R}\circ\underset{\sim}{R}\subseteq\underset{\sim}{R}$。

则称 $\underset{\sim}{R}=(r_{ij})_{n\times n}$ 是一个模糊等价关系。

自反性是指矩阵的对角线上的元素全是1，即表示任一物体总与它自己属同一类；对称性是指 $\underset{\sim}{R}$ 为对称矩阵，即 $r_{ij}=r_{ji}$。也就是说如果 x 与 y 属一类，则 y 与 x 也属同一类。传递性不易直接看出，需要计算 $\underset{\sim}{R}\circ\underset{\sim}{R}$，它表示 $\underset{\sim}{R}$ 与它自身的合成，然后看其是否满足：

$$\underset{\sim}{R}\circ\underset{\sim}{R}\subseteq\underset{\sim}{R}$$

它是表示若 x 与 y 属一类，y 与 z 属一类，则必有 x 与 z 也属同一类。

根据 α 截矩阵的概念，我们可以得出：若模糊关系矩阵 $\underset{\sim}{R}$ 是模糊等价关系，则对于任意 $\alpha\in[0,1]$，所截得 α——截矩阵 R_λ 也是等价关系。由此可知，模糊等价关系 $\underset{\sim}{R}$ 确定之后，对给定的 $\alpha\in[0,1]$，便可相应得到一普通等价关系 R_λ，这也就是说可以决定一个 α 水平的分类。

现在设一幅土壤图为一个给定的论域 U，土壤图中的图斑 a,b,x,y……等叫作元素。U 中某一部分元素的全体，即某些图斑构成评价中的某一等级，叫做 U 中的一个集合，常用 A，B，C……表示。

要想确定 U 中的一个集合 A，只要对 U 中任一元素 $x\in A$，或 $x\notin A$ 之间作一选择。

给定一块图斑 x，问它应该属于哪一等？如果由于评价因子的综合影响，确定这块图斑属于2等土壤，那么它就不可能再属于1等土壤了。这表明在土壤资源评价制图时，要求每块图斑都有它应有的等级范围，各图斑互不相交。就是说，这些图斑的分类关系应属于等价分类关系①②。

三、方法与步骤

为了说明模糊聚类分析在土壤资源评价制图中的应用，我们选定了江苏省1:100万土壤图的水稻土图斑作为试验样本，对江苏的水稻土作了粗略的肥力等级评价和制图工作，其方法和步骤如下：

1．拟订评价原则，选定评价项目，确定评价等级

如要评价江苏水稻土的肥力等级，不但应该考虑土壤本身的特性，还应该考虑其它自然因素，经济技术条件，耕作管理水平和改造利用土壤资源的能力等方面。所以我们选定了有机质含量及全氮(因为这两个因子是正相关的，所以实际计算时作为一个因子考虑)，全磷，质地，地貌部位和生产力状况(反映在产量上)及其参数作为评价项目和评价标准(表1)。

评价项目及其参数确定后，需要给单项评价项目的各参数给予对应的评价等级。评价等级确定的作用有二个。一个是对各个评价项目的文字参数予以数字化。另一个是对各参数起标准化的作用。因为各参数的量纲是不一样的，经过给予评定的等级以后，就把这些参数归到统一的量纲上来。本研究一共给予0，1，2，3，4，5六个等级(表1)。评价项目的各参数与它们一一对应，参数标准越高，等级也越高。

2．图斑的统计、整理和归并

1:100万江苏省土壤图上经统计，水稻土的图斑共有301块。我们可以把它看成301个单元，同时把它们编为1到301号。每一块图斑均代表了一种水稻土类型。而每一块图斑又都有5个因

① 耿春仁，赵以强编，模糊数学及其在气象上的应用，1982，讲义。

② 朱剑英，应用模糊数学，1982，讲义。

表1　　江苏省水稻土的评价项目、评价标准及其对应的等级*

评价等级	评价项目及标准					
	有机质(%)	全　氮(%)	全　磷(%)	质　　地	地　　貌	生产力状况(用产量来反映)
0	1.2—1.5	0.06—0.09	<0.05	沙　　壤	沿海平原	低
1	1.5—2.0	0.09—0.12	0.05—0.08	沙壤—轻壤，轻壤	低洼圩区，塘田	中
2	2.0—2.5	0.12—0.15	0.08—0.11	重壤—轻粘，轻粘—中粘	丘陵谷地	平
3	2.5—3.0	0.15—0.18	0.11—0.14	中壤，轻壤—中壤，中壤—重壤	圩　　区	高平
4	3.0—3.5	0.18—0.21	0.14—0.17		沿江平原，太湖平原	高
5	>3.5	>0.21	0.17—0.20			

*表中前三个评价项目及评价标准取自1966年出版的1:200万江苏土壤质地与土壤肥力要素图；第4、5个评价项目是作者与陈志诚同志根据江苏水稻土的质地和所处的地貌部位，人为分等的；最后一项取自陆彦椿同志对江苏水稻土研究的工作记录

表2　　经整理归并后133个单元所对应的评价项目等级

No	有机质含量及全氮	全　磷	质　地	地　貌	生产力状况	指数和
1	0	1	3	1	0	5
2	0	1	3	2	3	9
14	3	3	3	4	4	17
42	2	4	3	4	4	17
75	2.4*	2.8*	2	4	4	15.2
76	1.8*	3	2	4	4	14.8
101	0	3	2	1	1	7
102	2.5*	3	2	1	1	9.5
132	1.5*	1	3	2	3	10.5
133	4	3	2	1	1	11

* 表示某图斑跨两个评价标准参数时，取的平均数。

子(即评价项目)在对它们起作用。在整理的过程中我们发现有很多图斑具有等级完全一一相等的共同因子。也就是说，许多图斑的有机质及全氮，全磷，质地，地貌，生产力状况的评定等级完全相等。这时，我们就可以把这些图斑都看成是一个共同的单元。经过归并，统计出整

个图幅上还有133个其评价因子等级不同的单元。同时将它们编为1到133号(表2)。这133个编号与上述的301个编号是有着内在的联系的。但进行模糊聚类计算是从133个单元着手的。

3. 建立X^2上的模糊相似关系R

因为本研究中的区域是江苏省土壤图上全体水稻土图斑的集合，它是一个有限集，所以可以用一个模糊矩阵来等价刻划模糊关系R，也就是说，建立一个模糊相似矩阵R来代表模糊相似关系。设$R=(r_{ij})_{n\times m}$是一个$n\times m$阶实数矩阵，如果R的每一个元素r_{ij}，$i=1,2\cdots\cdots n$，$j=1,2\cdots\cdots m$都满足：$0\leqslant r_{ij}\leqslant 1$，就称R是一个$n\times m$阶模糊矩阵，本研究中n与m相等，所以所建立的模糊相似矩阵是个方阵。

建立模糊相似矩阵的关键必须求出矩阵中每个元素所对应的隶属度r_{ij}，并要注意，在模糊矩阵中，r_{ij}的取值范围应该是$r_{ij}\in[0,1]$。

隶属度的求法有好几种，本研究是应用如下的公式与步骤取得r_{ij}的：

当$i=j$时，则$r_{ij}=1$　　$(i,j=1,2,\cdots\cdots 133)$

当$i\neq j$时，则$r_{ij}=\left(\sum_{k=1}^{5}a_{ik}\cdot a_{jk}\right)/M$　　$(i,j=1,2,\cdots\cdots 113)$

式中a_{ik}、a_{jk}分别表示第i行第k个分量和第j行中的第k个分量。M是一个适当选定的常数。

$$M=\sum_{k=1}^{5}a_{ik}\cdot a_{jk}+x$$

x是任意的正整数。加x的目的是使M值足够大，使隶属度$r_{ij}\leqslant 1$。

本研究中的M值是第14行与第47行的分量一一相乘，再求总和并加上11取得的。

现在我们就可以求得各元素所对应的隶属度：

$r_{1\cdot 1}=1$

$$r_{1\cdot 2}=\sum_{k=1}^{5}a_{1k}\cdot a_{2k}/M=(0\times 0+1\times 1+3\times 3+1\times 2+0\times 3)/70=0.17$$

$$r_{1\cdot 3}=\sum_{k=1}^{5}a_{1k}\cdot a_{3k}/M=(0\times 1+1\times 2+3\times 3+1\times 1+0\times 3)/70=0.17$$

$$\vdots$$

$$r_{1\cdot 133}=\sum_{k=1}^{5}a_{1k}\cdot a_{133k}/M=(0\times 4+1\times 3+3\times 2+1\times 1+0\times 1)/70=0.14$$

$r_{2\cdot 2}=1$

$$r_{2\cdot 3}=\sum_{k=1}^{5}a_{2k}\cdot a_{3k}/M=(0\times 1+1\times 2+3\times 3+2\times 1+3\times 0)/70=0.19$$

$$\vdots$$

$$r_{2\cdot 133}=\sum_{k=1}^{5}a_{2k}\cdot a_{133k}/M=(0\times 4+1\times 3+3\times 2+2\times 1+3\times 1)/70=0.20$$

$$\vdots$$

$r_{133\cdot 133}=1$

求得各元素所对应的隶属度后，就能建立一个模糊相似矩阵(表3)。

表3　　模糊相似矩陈 R

	1	2	3	———	131	132	133
1	1	0.17	0.17	———	0.14	0.17	0.14
2		1	0.19	———	0.16	0.33	0.20
3			1	———	0.24	0.21	0.24
131					1	0.23	0.40
132						1	0.29
133							1

4．求模糊等价关系矩阵

从上面的模糊矩阵R中，可以看出，该矩阵中的主对角线元素均为1，而且矩阵中二个元素相应的隶属度对于主对角线来说都是对称的。因此称它为模糊相似矩阵R。但我们的目的是要完成土壤资源的等价分类制图，同时只有模糊等价关系才能聚类，故还须求出R的可传闭包t(R)。如果

$$t(R) = R \circ R \circ \cdots\cdots \circ R = R^k$$

则说明 R^k 就是一个模糊等价关系。由此方能对U论域中的元素在任意水平α上进行分类。这样，就可以对土壤图上全部水稻土图斑集合中的图斑在任意给定的水平上进行分类。本研究中 $0 < \alpha < 1$。

求R的可传闭包，实质上是求模糊相似矩阵的合成运算。

设　r_{ij} 是R矩阵的元素。

则　$R \circ R = R^2 = r_{ij}^{(2)}$

$$r_{ij}^{(2)} = \bigvee_{k=1}^{133} (r_{ik} \wedge r_{kj})$$

符号∧为逻辑乘，即取小。∨为逻辑加，即取大。

$\bigvee_{k=1}^{133} (r_{ik} \wedge r_{kj})$ 是表示先将第一个矩阵的第 i 行元素与第二个矩阵的第 j 列相应元素作比较，选取最小值，然后在133个最小值中再选取最大值，作为新矩阵中第 i 行第 j 列的元素值。

按上述公式进行计算，最后求得　　$t(R) = R^4 = R^8 = \cdots\cdots$(表4)

表4　模糊等价关系矩阵 $R^4 = R^8 = \cdots\cdots$

	1	2	3	·····	131	132	133
1	1	0.26	0.26	·····	0.26	0.26	0.26
2		1	0.36	·····	0.46	0.49	0.49
3			1	·····	0.36	0.36	0.36
131					1	0.46	0.46
132						1	0.51
133							1

说明本研究中的分类关系R已达到可传性的要求。所以 R^4 是一个模糊等价关系矩阵。利用 R^4 矩阵进行模糊聚类分析就能得到土壤资源评价的等级。

5．画出聚类图，进行模糊聚类分析

以归并后的133个图斑编号为横坐标，α值为纵坐标。按α值从大到小依次截取对应的图斑编号 $r_{ij} \geq \alpha$ 所对应的图斑编号应聚在一起，这样就形成了一棵系统树(图1)。然后根据土壤学家或权威意见，或根据自己在工作中的要求，在系统树上根据α值截取所需要的等级类型。

6．在土壤图上勾绘土壤资源评价等级图

根据系统树得出的土壤资源评价等级作为上图单元，再按133个编号追逆到301个图斑中去。在301个图斑编号中寻找与133个编号中一一相同的单元，并在这些图斑中填上与该单元相应的资源评价等级符号，直至301个图斑完全填完为止。

7．地图整饰

根据系统树，在土壤图上勾绘出土壤资源评价的制图单元，然后着墨清绘制图单元的界线，绘出图例，进行着色，完成整幅土壤资源评价图的编制工作(图 2)。

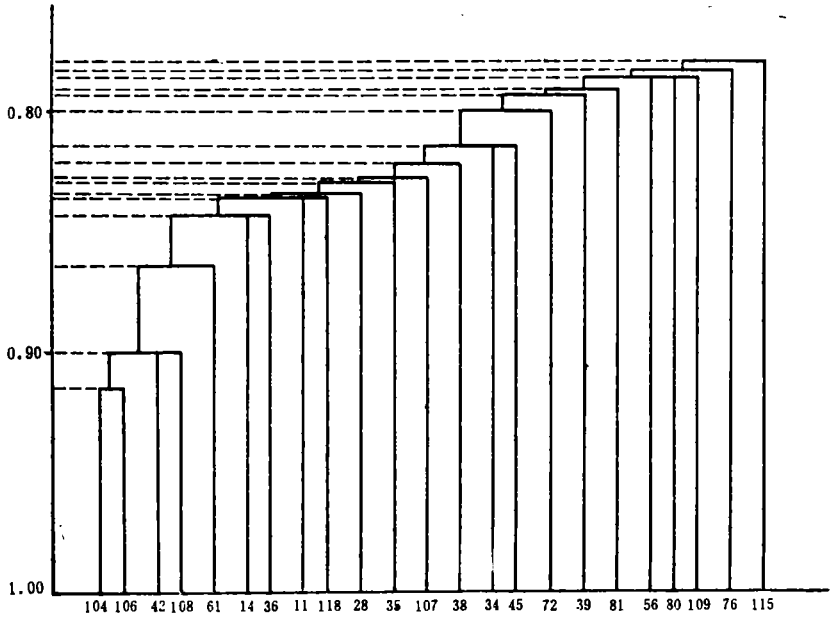

图1 模糊聚类系统树示意图

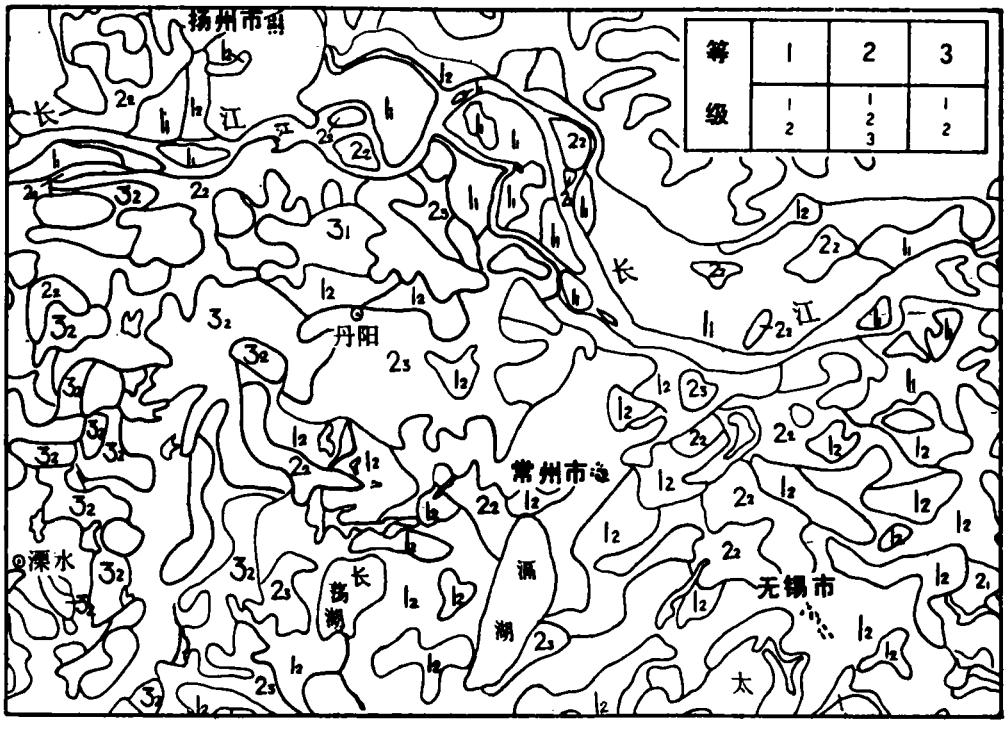

图2 江苏省水稻土肥力等级评定示意图

四、结果和讨论

利用模糊聚类分析的方法，我们把江苏省的水稻土资源分为一、二、三，三个等级。其

中 $\alpha \geqslant 0.68$ 为一等田。$0.68 > \alpha \geqslant 0.39$ 为二等田。$\alpha < 0.39$ 为三等田。一等田又分为二级。$\alpha > 0.80$ 为一等一级田。$0.80 \geqslant \alpha \geqslant 0.68$ 为一等二级田。二等田又分为三级。$0.68 > \alpha \geqslant 0.61$ 为二等一级田。$0.61 > \alpha \geqslant 0.46$ 为二等二级田。$0.46 > \alpha \geqslant 0.39$ 为二等三级田。三等田又分为二级。$0.39 > \alpha \geqslant 0.34$ 为三等一级田。$0.34 > \alpha > 0$ 为三等二级田。经征求有关同志的意见，他们认为最终编制的江苏省1:100万水稻土肥力等级图是能够反映出客观情况的。

土壤工作者以往都习惯于用指数相加的方法进行土壤资源评价制图，根据实践，这种方法是肯定可以应用的一种资源评价方法。指数相加的方法，容易使评价等级划分绝对化。在有的图斑中会出现与客观事实不符的制图单元。而通过模糊聚类分析获得的资源等级，理论依据比较充分，并能更科学地反应出客观存在的自然现象。例如，在用模糊聚类分析法编制的江苏水稻土肥力等级图上，我们发现119，123，126号图斑(都是夹沙土）的五个评价因子的等级均为 0 、 4 、 3 、 4 、 3，指数和为14，它们被归入一等一级田中。而同样为夹沙土的106，107，113号图斑，它们的指数和也为14，只是评价因子的等级分别为 2 、2 、3 、4、3，却被归入一等二级中去了。经分析，我们发现119，123，126号图斑的地理位置都靠近江边，第二个评价项目的全磷含量特别高(等级分数为 4)，因而它们被归入一等一级田中去了。由此可见，某个评价因子的特高或特低，在模糊聚类的分类中，分类的等级会出现升一级或降一级的情况，使之更符合客观实际情况。如果用指数相加的方法，这几个图斑就会全部归入一个等级类型中去。

利用模糊聚类分析法进行土壤资源评价制图的方法有好几种，但前几步的基本步骤大致相同，仅最后一步，有的是用 α 截取值进行聚类分析，也有用编网法和画最大树枝法进行聚类分析。本研究采用比较简便的画系统树法。画系统树的时候要注意从大到小截取 α 值，否则整个图形会出现不收敛的现象，无法进行聚类分析。

模糊聚类分析借助于电子计算机的帮助，不仅可以用于土壤资源评价制图，应用同样的原理，它还可以用于土壤自动分类制图中。成图速度快，逻辑性强。但是数学模型的建立必须要求有丰富经验的土壤工作者去选定合理的评价因子，或从数据库中抽调合理的成土因子，最后去检验聚类分析结果的正确性。科学的方法加上土壤工作者丰富的土壤知识和经验才能推动该学科的前进。

参 考 文 献

[1] 何同康，土壤资源研究方法简介。土壤专报，37号，51～79页，1980。

[2] 中国科学院南京土壤研究所黑龙江队著，黑龙江省与内蒙古自治区东北部土壤资源，209～214 ，科学出版社，1982。

[3] 杨艳生、冉家铭，县一级土壤资源评价研究。土壤专报，38号，57—66，1983。

[4] 贺仲雄编，模糊数学及其应用，98、152～156，天津科学技术出版社，1983。