

土地利用类型划分的计算机系统聚类分析

陈 寿 勤

(广西计算中心)

摘 要

本文介绍如何应用多元统计分析方法中的聚类分析方法,利用电子计算机科学地、合理地解决农业区域规划中的土地利用类型划分和农作物种植结构的最优化方案问题。文中以武鸣县的实例作了引证。而所介绍的计算机程序设计方法和实际应用程序具有解决聚类分析问题的通用性和可移植性。

一个地区的农业规划工作中,要根据国家计划和本地区的农业生产条件,经过农作物种植结构的最优化过程,合理地确定一个地区的农作物各品种种植计划。这样做,可以在各项主要生产条件保持基本不变的情况下,达到增产增收的目的。随之而来的问题便是如何安排这个地区各个区域的种植计划。过去,许多地方的农业计划部门,由于计算分析手段落后,往往凭经验甚至照人口或耕地面积、历年产量等简单比例摊派农作物种植面积,下达产量指标,忽视了本地区内土地利用类型上的差异性和相似性,即使考虑到了,其定性的分析也是相当粗糙的,无法真正做到因地制宜,合理安排。应用聚类分析方法,采用现代化计算工具——电子计算机进行土地利用类型的划分研究,可以使传统的定性分析与定量分析方法结合起来,提高农业计划制订工作的科学性,提高农作物种植计划摊派的合理性。

本文叙述应用电子计算机进行土地利用类型划分的系统聚类分析方法,探讨武鸣县在落实农作物种植计划上宜于采取的一些措施。

一、系统聚类分析方法

系统聚类分析方法是对事物进行分类的一种有效方法,是数理统计方法中多元分析的一种。这种方法的基本思想是:先将若干个样本各自为一类,计算它们的距离,选择距离最小的二个样本归为一个新类,计算新类与其它类的距离,选择距离最小的两个类归为一个新类,每次合并缩小一个类,直到所有样本都划为一个类为止,再根据不同级别划分为若干类型。

根据上述方法,在进行聚类分析时,首先是计算样本之间的距离。

(3) 极差正规化:

$$x'_{ij} = \frac{x_{ij} - \min_{1 \leq i \leq n} \{x_{ij}\}}{\max_{1 \leq i \leq n} \{x_{ij}\} - \min_{1 \leq i \leq n} \{x_{ij}\}}, \quad j = 1, 2, \dots, m$$

通过数据的标准化处理并计算样本两两之间的距离之后,就可得到距离矩阵,它是一个对称阵,取其下三角阵,记为 $D(0)$:

$$D(0) = \begin{matrix} & G_1 & G_2 & G_3 & \dots & G_{n-1} & G_n \\ \begin{matrix} G_1 \\ G_2 \\ G_3 \\ \vdots \\ G_{n-1} \\ G_n \end{matrix} & \begin{pmatrix} 0 & & & & & & \\ G_{21} & 0 & & & & & \\ G_{31} & G_{32} & 0 & & & & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\ G_{n-1,1} & G_{n-1,2} & G_{n-1,3} & \dots & 0 & & \\ G_{n,1} & G_{n,2} & G_{n,3} & \dots & G_{n,n-1} & 0 \end{pmatrix} \end{matrix}$$

括号外的 G_1, G_2, \dots, G_n 表示类,最初,每个样本为一类。 $d_{pq} = d_{qp}$,表示两个样本间的距离。

第二步,选择 $D(0)$ 的最小元素,设为 D_{pq} ,则将 G_p 和 G_q 合并成一新类,记为 $G_r, G_r = \{G_p, G_q\}$ 。

第三步,计算新类 G_r 与其它类的距离。

新类与其它类的距离计算可有最短距离法、最长距离法、中间距离法、重心法、类平均法、可变类平均法、可变法、离差平方和法等多种方法。这些方法可以用一个统一公式计算。我们实际应用中,采用最短距离法和最长距离法。

采用最短距离法计算时,新类 G_r 与其它类的距离 D_{rk} 为:

$$D_{rk} = \min_{i \in G_r, j \in G_k} d_{ij} = \min \left\{ \begin{matrix} \min_{i \in G_p, j \in G_k} d_{ij}, \\ \min_{i \in G_q, j \in G_k} d_{ij} \end{matrix} \right\} = \min \{ D_{pk}, D_{qk} \}$$

将 $D(0)$ 中 p 行、 q 行、 p 列、 q 列用上式合并成一个新行新列。新行新列对应 G_r ,所得到的矩阵记作 $D(1)$ 。

最长距离法定义为新类 G_r 与其它类 G_k 的距离用两类之间最远的距离来表示:

$$D_{rk} = \max \{ D_{pk}, D_{qk} \}$$

其它方法同最短距离法。

第四步,对 $D(1)$ 重复上述对 $D(0)$ 的两步,得 $D(2)$,如此进行下去,直到所有的元素成为一类为止。

如果某一步 $D(k)$ 中最小元素不止一个,则对应这些元素的类可以同时合并。

二、程序设计的考虑

我们分别用BASIC语言和FORTRAN语言设计了一个聚类分析的通用程序。在程序设

计时，作了如下一些处理（参考程序框图）：

1. 为了决定是否进行数据标准化处理，设置字符变量Y\$,若输入的Y\$值等于“Y”，则进行数据标准化处理，否则转入下一步。

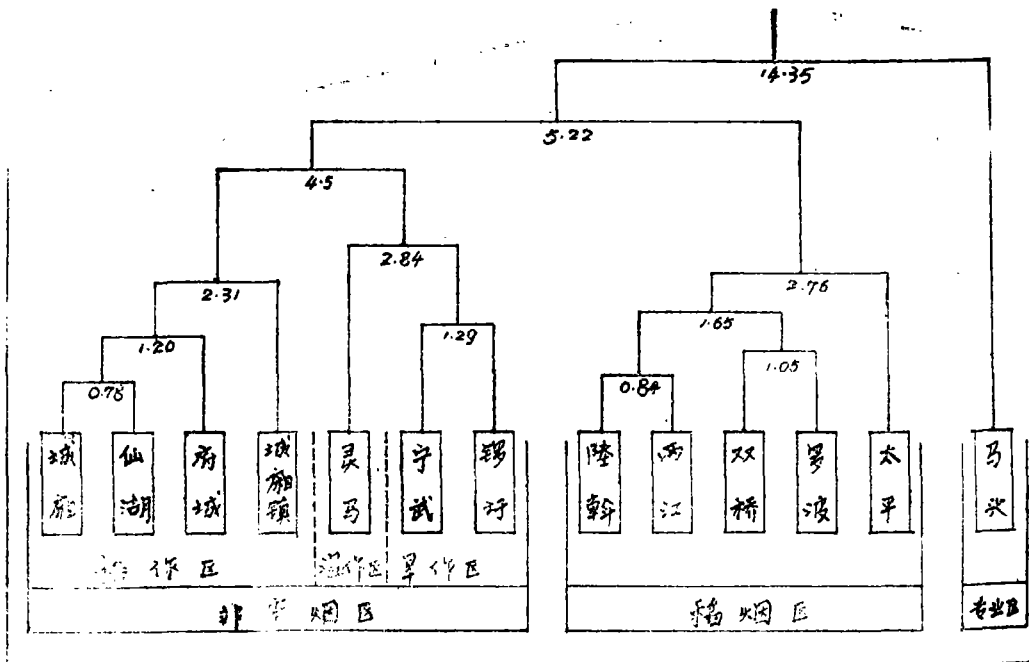
2. 在数据标准化程序段，根据输入的E值决定选用何种标准化方法，E取值1, 2, 3, 分别对应于选用标准差标准化、极差标准化或极差正规化三种方法。各标准化处理分别编一子程序解决。

3. 合并出一个新类后，计算新类G_r与其它类的距离D_{r,k}时，依所输入的L\$值（“S”或“L”）不同，判断应采用何种距离算法（最短距离法或最长距离法）。

程序框图见38页

三、武鸣县土地利用类型划分及评价

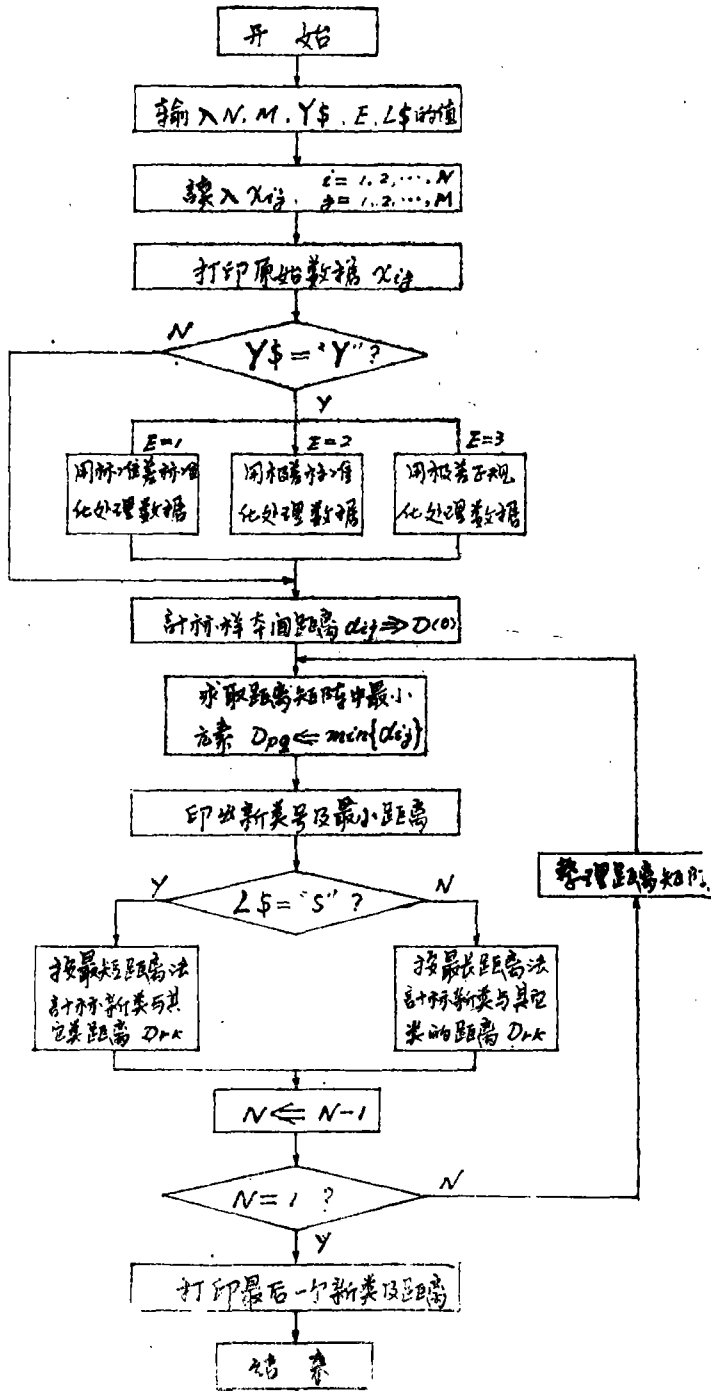
我们应用上述聚类分析方法和通用程序，对武鸣县土地利用类型划分进行了聚类分析，并结合实际情况给出各个类型土地利用的评价。



聚 类 图

在实际进行聚类分析时，我们选择影响武鸣县13个公社农作物结构的五个因子：水田贡献率、水稻贡献率、玉米贡献率、花生贡献率、烤烟贡献率作为样本指标（见表2），各贡献率是按如下方法计算的：

$$\text{贡献率} = \frac{\text{某公社某作物占全县该作物产量百分比}}{\text{该公社耕地占全县耕地百分比}}$$



程序粗框图

由于不存在量纲和量级的影响，故不需进行数据标准化处理，而直接采用绝对距离公式计算各公社各因子之间的差异，从而得到绝对距离矩阵(见表3)，然后用最长距离法进行聚类分析，并作出聚类图(见37页)。由聚类图，在最短距离5.50处划一平行于横坐标的直线，可将13个区域划分为2个土地利用类型：专业产烟区和非专业产烟区；在最短距离5.0处划一平行于横坐标的直线，又可划分为非产烟区、稻烟区和专业产烟区三个土地利用类型；在最短距离2.80处划一平行于横坐标的直线，可以将13个区域划分为稻作区、混作区、旱作区、稻烟区和专业产烟区五个土地利用类型。表2列出了五个土地利用类型的范围、农业生产特点及对各类型的土地利用评价。

表1 武鸣县各乡影响农作物结构主要因素指标

X _i 乡	项 目	水田贡献率	水稻贡献率	玉米贡献率	花生贡献率	烤烟贡献率
城 厢		0.95	1.12	0.54	1.14	0.29
太 平		1.07	0.97	0.74	0.64	2.44
双 桥		1.16	1.30	0.43	0.91	1.43
宁 武		0.57	0.30	2	0.99	0
锣 圩		0.67	0.67	2.16	0.33	0
灵 马		1.13	1.08	1.97	2.06	0.05
仙 湖		0.95	1.04	0.51	0.62	0.14
府 城		0.95	0.92	0.93	0.79	0.03
陆 斡		1.16	1.19	0.54	1.15	1.04
两 江		1.38	1.13	0.69	1.52	1.08
罗 波		1.21	1.30	1.06	1.23	1.38
马 头		1.47	1.28	0.5	2.47	9.14
城 厢 镇		1.5	1.67	0.33	1.17	0

表2 武鸣县土地利用类型划分及评价

区域类型			范围	农业生产特点	评价
非 产 烟 专 业 区	非 产 烟 区	稻 作 区	包括城厢、仙湖、府城、城厢镇等四个乡	耕地占全县总数28.14%，以水稻生产为主，稻田占耕地51.4%，主要作物有水稻、玉米、甘蔗、花生、木薯及少量烤烟，其贡献率分别是1.03, 0.67, 1.37, 0.88, 1.11及0.2。是该县水稻、甘蔗、花生、木薯主要生产基地之一。	今后可适当减少粮地，扩种甘蔗，轮作的春玉米改种花生，注意挖掘潜力，进一步提高单产。
		旱 作 区	包括宁武、锣圩两个乡	耕地占全县总数21.79%，以旱作为主，畲地占耕地63.2%，玉米、水稻、甘蔗、花生、木薯的贡献率分别是2.09, 0.55, 0.92, 0.63, 1.16，旱作亩产低，是黄豆、黑豆最主要产区，其贡献率高达3.16。	今后以发展甘蔗、花生为主，经济作物与粮食作物并举，以提高农民收入。
		混 作 区	灵马乡	耕地占全县总数3.88%，旱地虽只占33.5%，但旱地作物的花生、玉米贡献率分别高达2.05, 1.99，木薯也达1.01，水稻及甘蔗贡献率分别是1.07, 0.41，水稻、花生、甘蔗亩产都较高。	今后可适当扩种甘蔗。
	稻 烟 区	包括陆斡、两江、太平、罗波、双桥等五个乡。	耕地占全县总数42.62%，是该县烤烟产区，烤烟贡献率为1.47，其余作物以水稻种植为主，稻田占耕地62.8%，水稻、玉米、甘蔗、花生、木薯的贡献率分别为1.18, 0.61, 0.90, 1.05, 0.2，是该县水稻、花生、甘蔗主要生产基地之一。	今后，可适当减少粮地，扩种甘蔗，轮作的春玉米改种花生，注意挖掘潜力，进一步提高单产。	
产 烟 专 业 区	产 烟 专 业 区	产 烟 专 业 区	马头乡	耕地占全县总数3.57%，是该县最有名的烟叶产区，其烤烟不但产量高，而且质量好，是高级香烟的原料烟。烤烟贡献率高达9.2，水稻、玉米、甘蔗、花生、木薯的贡献率分别为1.28, 0.5, 0.55, 2.50, 0.04。	今后重点发展水稻、烤烟、花生等作物，特别注意提高烤烟的种植技术，甘蔗种植也可适当扩大。

表3 武鸣县各公社影响农作物结构主要因子绝对距离

公社 d_{ij}	城厢	太平	双桥	宁武	锣圩	灵马	仙湖	府城	陆斡	两江	罗波	马头	城厢镇
城厢	0	3.12	1.87	3.1	3.45	2.81	0.78	1.20	1.04	1.76	2.14	10.9	1.63
太平		0	2.01	5.22	4.87	5.21	2.74	2.92	2.42	2.76	2.44	9.48	4.5
双桥			0	4.67	4.86	4.32	2.13	2.61	0.85	1.61	1.05	9.67	2.5
宁武				0	1.29	2.49	3.12	2.3	4.14	4.56	4.2	14	4.15
锣圩					0	2.84	2.73	2.25	4.49	4.91	4.55	14.35	4.5
灵马						0	3.21	2.67	3.47	3.15	3.37	11.51	3.54
仙湖							0	0.82	1.82	2.54	2.92	11.62	2.05
府城								0	2.24	2.66	2.56	12.1	2.31
陆斡									0	0.84	1.10	9.86	2.09
两江										0	1.30	9.44	2.45
罗波											0	9.84	2.83
马头												0	11.03
城厢镇													0

参 考 文 献

- [1] 张尧庭, 方开泰: 多元统计分析引论, 科学出版社, 1982年版。