

模糊聚类分析方法在划分 农业区划评价图中的应用*

张新长 张克权

(武汉测绘科技大学, 430070)

【摘要】 客观、真实地评价一个地区农业经济区划差异性和农业经济水平, 从而预测该地区未来农业发展水平, 制作地区农业生产各部门区划评价图是非常必要的。本文结合某地区山区地域特点, 运用多元统计分析中模糊聚类分析的数学方法, 用统计信息进行定量分析, 使编制出的农业区划专题地图更具有实际应用价值和指导作用。

【关键词】 模糊聚类 专题制图 农业经济区划 数学模型

聚类分析是数理统计中研究“物以类聚”的一种多元统计分析方法, 即用数学方法定量地确定样品的亲疏关系, 从而客观地分型划类。由于很多事物本身都带有模糊性, 因此把模糊数学方法引入聚类分析, 就能使分类更切合实际。这在分析农业生产结构区划评价方面是一种有价值的探讨。

1 统计量选择

应用聚类分析方法, 关键在于要把各生产部门统计量选择合理, 即这些统计量具有明确的实际意义; 有较强的分辨力和代表性。以林业生产部门为例, 结合某山区地域特点, 主要选择了以下因子:

表 1

自然 条件 因子	地貌组合类型	社会 经济 条件 因子	人口密度
	斜坡坡度		森林覆盖率
	年日照时数		人均林业用地
	年平均气温		亩产活立木蓄积量
	年降水总量		用材林面积
	土层厚度		

2 模糊聚类方法

在统计量选定之后, 进行模糊聚类分析方法大致分以下三步:

- (1) 将各基准单元的统计量数据进行规格化预处理, 以便消除量纲, 使其可比。设有 n 个基准单元, 每个单元有 m 个数据, 变量记为 x_{ij} ($i = 1, 2, \dots, n; j = 1, 2,$

* 本文 1992 年 11 月 13 日收到, 截稿日期为 1993 年 7 月 20 日。

..., m)。

规格化后的变量为 x'_{ij} 即:

$$x'_{ij} = (x_{ij} - \bar{x}_j) / S_j \quad (1)$$

其中 \bar{x}_j 为第 j 个变量的平均数, S_j 为 j 个变量的标准差。

(2) 计算衡量被分类对象间相似程度值 γ_{ij}

设给定模糊关系矩阵 $\tilde{R} = [\gamma_{ij}]$, 如果它满足 $[\gamma_{ii}] = 1$; $[\gamma_{ij} = \gamma_{ji}]$; 则称 $\tilde{R} = [\gamma_{ij}]$ 为模糊相似矩阵, 它满足自反性和对称性。

计算模糊相似矩阵方法颇多, 本文选择了夹角余弦法, 即:

$$\gamma_{ij} = \frac{\sum_{k=1}^m x'_{ik} \cdot x'_{jk}}{\sqrt{\sum_{k=1}^m x'^2_{ik} \cdot \sum_{k=1}^m x'^2_{jk}}} \quad (2)$$

其中 $(i, j = 1, 2, \dots, n)$, $(k = 1, 2, \dots, m)$ 。

(3) 建立模糊等价矩阵、聚类

设给定论域 U 上一个模糊关系 $\tilde{R} = [\gamma_{ij}]_{n \times n}$, 如果满足:

$$\begin{aligned} \gamma_{ii} &= 1 && \text{自反性} \\ \gamma_{ij} &= \gamma_{ji} && \text{对称性} \\ \tilde{R} \circ \tilde{R} &\subseteq \tilde{R} && \text{传递性} \end{aligned}$$

则称 $\tilde{R} = [\gamma_{ij}]_{n \times n}$ 是一个模糊等价关系矩阵。

显然, 自反性、对称性是模糊相似矩阵满足的条件, 而传递性不易直接看出, 需计算 $\tilde{R} \circ \tilde{R}$ (记作 \tilde{R}^2), 即把数的运算加与乘改为并“ \vee ”与交“ \wedge ”:

$$\begin{aligned} \gamma_{ij} &= \bigvee_{k=1}^n [\gamma_{ik} \wedge \gamma_{kj}] = (\gamma_{i1} \wedge \gamma_{1j}) \vee (\gamma_{i2} \wedge \gamma_{2j}) \vee \dots \vee (\gamma_{in} \wedge \gamma_{nj}) \\ (i, j &= 1, 2, \dots, n) \end{aligned} \quad (3)$$

这样计算 $\tilde{R}^2 = \tilde{R} \circ \tilde{R}$; $\tilde{R}^4 = \tilde{R}^2 \circ \tilde{R}^2$; 直到 $\tilde{R}^{2^k} = \tilde{R}^k$ 为止, 此时模糊矩阵 \tilde{R}^k 满足等价关系 $\tilde{R} \circ \tilde{R} \subseteq \tilde{R}^k$, 具有传递性。

在模糊等价关系 \tilde{R}^k 确定之后, 对给定的 λ 水平集 $\in [0, 1]$, 便可相应得到一普通等价关系 R^k_λ , 即可以决定一个 λ 水平的分类。将 $[\gamma^k_{ij}]$ 依大小顺序排列, 从 1 开始, 沿着 $[\gamma^k_{ij}]$ 自大到小依次取 λ 值, 定义:

$$\gamma^k_{ij} = \begin{cases} 1 & \gamma^k_{ij} \geq \lambda \\ 0 & \gamma^k_{ij} < \lambda \end{cases} \quad (4)$$

显然亦可得到若干个 0.1 R^k_λ 矩阵, 其中为 1 的表示这两个样品(区划)单元划为一类, 最后得到聚类结果, 画出动态聚类图。

3 统计数据的预处理

上面已经选择出了林业生产部门统计量。在实际应用中, 首先要对这些统计量按生产部门特点要求进行分类分级, 综合评判。比如对于本文所研究的这一地区具有中南山区特点, 林业生产一般分布在海拔 1200~2200 米左右。总之, 各种动、植物对自然条件有各自的最佳适应范围, 这是分级评分的主要依据。

对于自然条件因子, 由于各种自然条件因子是针对不同的区域单元, 因此首先必须归算

到统一的区域单元内。即把需要分析研究的乡、镇作为基准单元，对每个基准单元所包含其它区域单元的某种因子的不同级别评分值以面积比重作为权计算出各相应级别的评分值，最后计算出各基准单元内各类级别的评分值之和，即为该乡、镇单元各因子的总评分值。由于采用了评分值方法，实际上已经将基准单元各自然条件因子变成了一种模糊概念（评价等级概念）。

对于社会经济条件因子，由于该山区各乡、镇各种经济统计指标差异极大，若进行分级评分处理，就会造成两极本来差异很大的反而减小了它们之间的差异。所以对社会经济条件因子的统计数据仍以绝对数值参与聚类分析更符合本地区实际分布规律。

另外，在对自然条件因子统计数据的分级上一般不宜过多。主要考虑到自然条件具有较大的宽容，也就是说农业生产具有较大范围的适应性，无需对它们再进行较细的分级。较细的分级反而会带来很大的随机波动，而较粗的分级能较客观地反映农业经济区域的差异性。

4 应用分析研究

本地区有 142 个乡、镇作为样本 ($n = 1, 2, \dots, 142$) 针对林业生产特点选择了 m 个统计量 ($m = 1, 2, \dots, 11$)。它们分别是：

- x_1 : 地貌组合类型(米); x_2 : 斜坡坡度($^\circ$); x_3 : 年日照时数(小时);
- x_4 : 年平均温度($^\circ\text{C}$); x_5 : 年降水总量(毫米); x_6 : 土层厚度(厘米);
- x_7 : 人口密度(人/平方公里); x_8 : 森林覆盖率(%); x_9 : 有林地面积(亩);
- x_{10} : 人均林业用地(亩/人); x_{11} : 亩产活立木蓄积量(米³/亩)。

其中 $x_1 \sim x_6$ 项采用分级评分值； $x_7 \sim x_{11}$ 项采用绝对值。组成原始数据矩阵：

$$x = [x_{ij}] = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

其中 $i = 1, 2, \dots, 142; j = 1, 2, \dots, 11$

(因数据量较大，仅列出原始数据的首尾部分)，(见表 2)。并在此基础之上，对数据矩阵进行规格化预处理。(表中双横线下部分为经规格化预处理后的数据)。

消除量纲干扰后，建立模糊相似矩阵 R 。由于矩阵样本太大 (142×142)，所以本文无法全部列出，仅以模糊相似矩阵表达式说明：

$$\tilde{R} = \begin{pmatrix} \mathcal{V}_{11} & \mathcal{V}_{12} & \dots & \mathcal{V}_{1 \ 142} \\ \mathcal{V}_{21} & \mathcal{V}_{22} & & \mathcal{V}_{2 \ 141} \\ \vdots & \vdots & & \\ \mathcal{V}_{142 \ 1} & \mathcal{V}_{142 \ 2} & \dots & \mathcal{V}_{142 \ 142} \end{pmatrix}_{142 \times 142}$$

其中 $\mathcal{V}_{ij} = \mathcal{V}_{ji}; \mathcal{V}_{ii} = 1 (i, j = 1, 2, \dots, 142)$

再将相似矩阵转化为模糊等价矩阵 ($R^k = R^{2k}$) 满足传递性。 R^k 仍然是一个 142×142 方阵。

聚类是在对模糊等价矩阵 $R^k = [\mathcal{V}_{ij}^k]_{142 \times 142}$ 的数据分布特征分析基础上，选择不同的 λ 水平值进行聚类的。在划分林业区划评价等级时，为了得到五种评价区划（最好、较好、

表 2

编 号	项 目	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1		1.00	1.38	5.00	5.00	1.00	0.00	318.00	8.20	5861.00	2.07	0.80
2		1.00	0.00	5.00	5.00	1.00	1.34	20.90	10.00	6421.00	3.06	0.94
3		1.00	0.00	5.00	5.00	1.00	1.34	20.90	10.00	6421.00	3.06	0.94
4		1.00	1.28	5.00	5.00	1.00	0.45	213.00	3.60	59594.00	4.38	0.37
5		1.37	1.52	3.00	4.00	2.00	1.22	143.00	35.30	143864.00	8.37	1.45
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
138		2.48	2.31	3.00	4.00	5.00	2.94	97.00	54.72	105878.00	10.38	0.73
139		3.43	2.31	1.00	3.00	4.00	2.89	27.00	67.94	319674.00	42.75	3.34
140		3.36	2.76	1.00	3.00	4.00	2.98	39.00	47.93	84568.00	26.02	0.01
141		2.78	2.35	1.00	3.00	4.00	2.31	58.00	66.65	79495.00	21.28	0.89
142		2.67	1.93	1.00	3.00	5.00	2.73	67.00	56.09	172863.00	14.36	0.32
1		-1.57	-0.27	2.06	1.68	-1.36	-4.25	0.44	-1.53	-1.30	-0.09	-0.42
2		-1.57	-0.58	2.06	1.68	-1.36	-2.07	0.19	-1.44	-1.29	-0.09	-0.30
3		-1.57	-0.58	2.06	1.68	-1.36	-2.07	0.19	-1.44	-1.29	-0.09	-0.30
4		-1.57	-0.29	2.06	1.68	-1.36	-3.52	0.08	-1.76	-0.20	-0.08	-0.80
5		-1.06	-0.24	0.60	0.57	-0.66	-2.26	-0.16	-0.21	1.51	-0.08	0.15
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
138		0.49	-0.06	0.60	0.57	1.45	0.54	-0.31	0.74	0.74	-0.08	-0.80
139		1.81	-0.06	-0.86	-0.53	0.75	0.46	-0.55	1.38	5.08	-0.08	1.82
140		1.71	0.04	-0.86	-0.53	0.75	0.61	-0.51	0.41	0.30	-0.08	-1.12
141		0.90	-0.05	-0.86	-0.53	0.75	-0.48	-0.45	1.32	0.20	-0.08	-0.34
142		0.75	-0.15	-0.86	-0.53	1.45	0.20	-0.42	0.80	2.10	-0.08	-0.84

一般、较差、最差),按 λ 值从大至小进行逐次归并。分别取:0.97、0.94、0.91、0.89、0.86、0.83、0.80、0.64,在 λ 值 ≤ 0.89 时,得到五类林业评价区划的单元编号。根据分类特征,绘制出林业评价区划动态聚类图(见图1)和由此而制作的林业区划评价图(见图2)。

依据聚类结果可以初步看出:林业生产水平具有按自然条件差异分布的特点,尤其受地

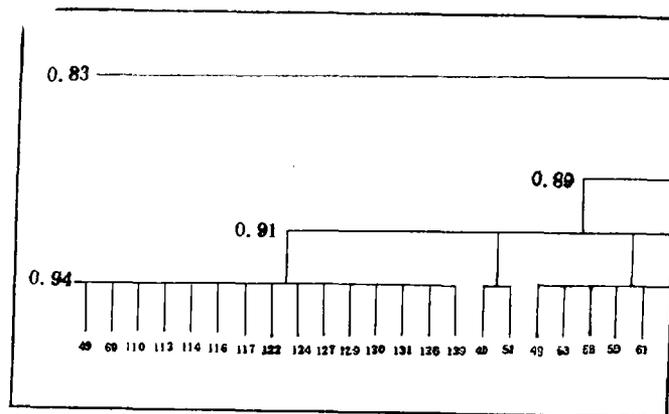


图 1 林业区划评价动态聚类图(部分)

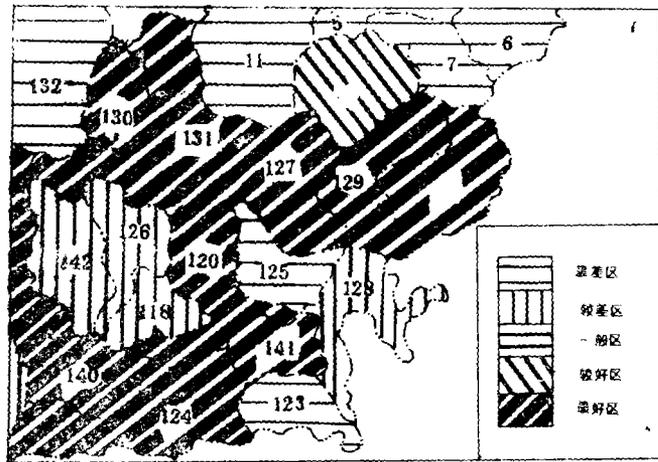


图 2 某地区林业区划评价图(部分)

貌类型结构的控制,反映了林业生产在山区的发展趋势。西北部低中山区,林业用地面积很大,且主要分布在该地区西北,优势区占该区总面积50%以上;东北部江河沿岸区,因沿江河及库区,所以不宜发展林业生产;中部中低山区,林业用地面积更大,优势区占全区80%以上,以有林地为主,是一个发展型的林业基地;中南部低山丘陵区,林业生产也有一定规模,主要以用材林、经济林为主;南部高山区,林地面积极为广阔,全区均为林业分布优势区,为全地区之冠,是发展林业生产的重要基地。

经过以上分析我们可得出如下结论:

(1) 将定量分析与定性分析方法相结合,相互比较印证,是进一步完善各种区划分析研究的科学方法。

(2) 用多元统计分析划分各类农业区划,合理采集和构造自然条件、社会经济条件因子是关键一步。

(3) 应该承认:模糊数学在对农业各部门区划划分及评价上有独特作用,但不能完全照搬,而是要根据具体情况灵活运用。

(4) 由于采集的是某一年的原始变量数据,某些统计数据波动性较大,因而反映在划分某些农业区划上不够合理,因此建议在可能的条件下,应尽量采用多年平均指标进行聚类。

参 考 文 献

- (1) 丁士晟编著. 多元分析方法及其应用. 吉林人民出版社, 1979
- (2) 贺仲雄. 模糊数学及其应用. 天津科学出版社, 1983
- (3) 张克权, 郭仁忠. 专题制图数学模型. 测绘出版社, 1991
- (4) Lajos S, Ferenc C. Statistical Determination of Class Intervals for Maps, 1978

APPLICATION OF THE ANALYTICAL METHOD OF FUZZY CLUSTER IN MAKING AGRICULTURAL REGION EVALUTING MAPS

Zhang Xinchang, Zhang Kequan

(Wuhan Technical University of Surveying and Mapping)

Abstract

In order to evaluate more adjectively and more authentically agricultural economical divisions of a region and agricultural economical level, and so to forecast agricultural economy developing direction of this region, it is necessary to make this region evaluating maps of agricultural economy in all branches. This paper combines an area with mountain area characteristic and utilizes multielement statistical analysis, cluster analysis or fuzzy cluster as a mathematical method by constructing all sorts of statistical factors and surveying them. To carry on quantitative analysis based on statistic information and to be practically valuable and conducted affect for study of agricultural results and for compilation of various of thematic maps of agriculture by these results.

Key words Fuzzy cluster, Thematic map, Region of agricultural economy, Mathematical model

(责任编辑: 张燕燕)

论文发表、论文降重、论文润色请扫码



免费论文查重，传递门 >> <http://free.paperyy.com>

阅读此文的还阅读了：

- [1. DBSCAN聚类算法的研究与改进](#)
- [2. 改进的模糊BP神经网络及在犯罪预测中的应用](#)
- [3. 浅谈聚类分析及最大树聚类法之应用](#)
- [4. 模糊聚类分析方法在划分农业区划评价图中的应用](#)
- [5. 机群系统上基于映射和抽样划分的并行聚类算法](#)
- [6. 模糊聚类分析在种子分类上的应用](#)
- [7. 聚类分析的快速预处理算法](#)
- [8. 模糊聚类灰关联度分析的决策方法](#)
- [9. 网格资源超图划分聚类任务调度](#)
- [10. 聚类融合控制算法研究](#)
- [11. 基于隶属度改进的模糊K-均值算法](#)
- [12. 模糊联想记忆系统规则提取的新方法](#)
- [13. 基于粒子群优化的模糊K-Means目标分类算法](#)
- [14. 海量中文短信文本最佳聚类数研究](#)
- [15. 基于信息理论的合作聚类算法研究](#)