

# 空间分析技术和图论树算法在农业区划中的应用

史舟 吴宏海 王人潮

(浙江大学浙江省农业遥感与信息技术重点研究实验室, 杭州 310029)

**摘要** 地域分异规律是区划研究的理论依据, 而进行区划的技术关键是如何根据空间相似性实现其合理分割. 该研究以地理信息系统的空间数据采集、分析工具为支撑, 利用 GIS 的空间邻接分析与图论的树算法相结合的方法对浙江省龙游县农业分区进行了研究. 利用该方法建立的自动分区模型, 运行效率高, 且分区结果与当地实际较为符合.

**关键词** 最小树分类 农业区划 地理信息系统

中图分类号: P208, S126 文献标识码: A 文章编号: 1006-8961(2000)03-0244-05

## Application of Spatial Analysis and Optimal-Tree Cluster in Agricultural Division

SHI Zhou, WU Hong-hai, WANG Ren-chao

(Institute of Agricultural Remote Sensing and Information System, Zhejiang University, Hangzhou 310029)

**Abstract** Because of the existence of regional variability in natural and economical condition in agricultural production, proper agricultural division is essential that different developing strategy can be adopted for every zone. However, establishment of agricultural division requires large, unwieldy geographical-referenced data sets and spatial adjacency relationship, it is difficult to utilize many conventional mathematical methods. In the study, Geographical Information System was used to manipulate and analysis spatial data, Optimal-tree Cluster was adopted for division. This integrated methodology was programmed as applied model and embedded in Soil Information System, case on Longyou county, Zhejiang province. Advantage of this integrated methodology is easily and quickly with low cost, compared to agricultural division by local bureau. Furthermore, the result is also suitable for the practical situation.

**Keywords** Optimal-tree cluster, Agricultural division, Geographic information system

## 0 引言

所谓区划就是根据地域分异规律把性质特征不同的地域分开, 将相似的归为一类, 以便在工作中实现分区分类指导. 区划一般应符合两个要求, 一是区内自然、经济条件具有相对一致性; 二是保持一定的行政区界的完整性. 目前, 区划一般采用定性方法和统计、随机、动态等定量方法, 而在多因素条件下的分类与分区分析法有主成分得分、因子得分、回归方

程数值划分、聚类分析、模糊综合评判和判别分析等方法<sup>1~3</sup>. 由于在区划定量化研究中, 数学模型的应用存在两个问题, 一是模型所需大量的原始数据较难准确获取, 从而影响了模型的使用和区划结果的正确性; 二是利用该模型所得的区划结果很难保持区块的空间连接和行政区界的相对完整, 因而不符合区划的要求. 因此, 如何将空间信息(包括空间分异和空间邻接)、数学工具、人的经验有机地结合起来是当前区划研究的新方向.

本研究以地理信息系统的空间数据采集、分析

工具为支撑, 利用 GIS 的空间邻接分析与图论的树算法相结合来进行浙江省龙游县农业种植分区. 同时将分区算法程序化后嵌入到红壤资源信息系统中, 并与龙游县空间数据库结合, 建立了县级农业种植分区模块.

### 1 空间分析法与图论树算法结合下的区划方法

由于区划既要求各分区内部的自然、经济条件具有相对一致性, 又要保持空间上的连通性和行政区划界的完整性. 如采用普通的聚类方法, 就很难使分类结果满足分区的原则, 因此, 需引入图论的空间表示法及相关算法.

分区原则的数学表达为: 设  $U$  代表将要划分的区域 (有  $n$  个分区单元  $V_i \subseteq U, i = 1, 2, \dots, n$ ), 对  $U$  进行区划, 就是求一割集  $\{U_i, i = 1, 2, \dots, m\}$ , 其中  $m$  为分区数. 该割集需满足: ①  $V_i \subseteq U_j$ ,  $\forall V_i, V_j = 1$  (连通性); ②  $d(V_i, V_j)$  定义为一个  $V_i$  和

$V_j$  之间的关联度量,  $\max \{d(V_i, V_j) \mid V_i, V_j \in U_i\} < A, A$  为一常数 (内部一致性)<sup>[4]</sup>.

根据图的几何概念, 图是空间一些点与连接这些点的线的集合, 即  $G = (V, E)$ , 其中  $V$  表示点的集合,  $E$  表示边的集合. 若分区用无向图来表示, 则  $V$  是各分区单元集合,  $E$  是各单元邻接关系集合. 由于分区在考虑空间邻接性的同时, 更重要的是各分区单元之间的内在相似性. 因此, 分区需用加权连通图 (网络) 来表示, 即  $G = (V, E, D)$ , 其中  $D$  在图论的树算法中表示各点之间的权值. 分区中采用各单元间的相似系数来构造权值矩阵  $|D|$ , 采用各分区单元间的邻接关系来构造邻接矩阵  $|E|$ . 现在区划则转变为求加权连通图  $G = (V, E, D)$  的合理割集, 即首先采用图论最小树法求得一个最小生成树, 然后根据区内分异小, 区间分异大的原则, 选择适当的  $A$  值, 将最小生成树分割成  $m$  个子树, 即将研究区域分成  $m$  个种植利用区. 利用 GIS 的空间分析手段与图论的树算法相结合的方法进行区划研究的具体技术思路见图 1 所示.

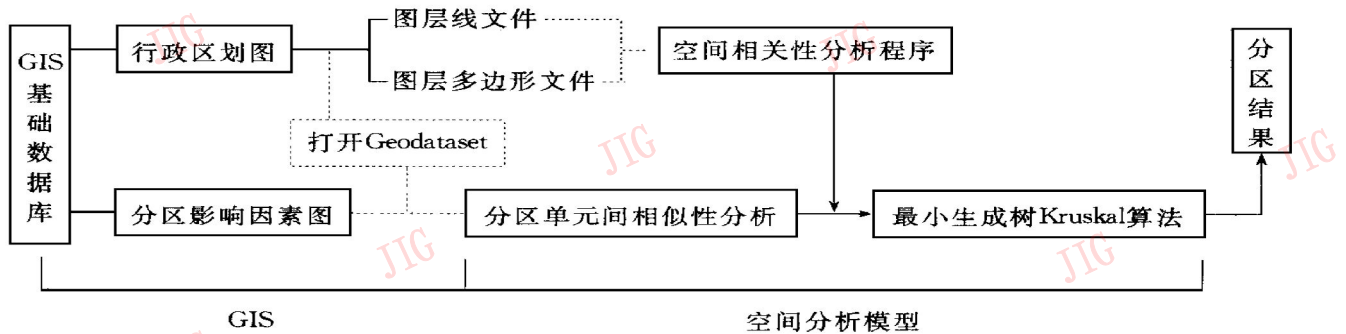


图1 空间分析法与图论树算法结合下区划方法技术框图

## 2 农业种植利用分区实例研究

### 2.1 分区指标的选择和数据提取

根据专家的意见, 同时结合龙游县当地实际情况, 以乡 (镇) 为地域单元, 从农业生产条件和农业经济状况两方面, 选择了土壤有机质含量 (%)、酸碱度、坡地占土地总面积比重 (%)、年平均气温 (°C)、有效灌溉面积占耕地面积比 (%)、农村人均收入 (元/年)、种植业产值占农业产值比重 (%) 等 7 项区划指标进行农业区划.

提取分区指标单元数据分为从 GIS 数据库中提取和从统计资料中提取两种方法: (1) 采用 GIS 手段从 GIS 的指标图层中提取单元数据文件, 如土

壤有机质、土壤 pH 值、坡地占土地总面积比重等数据; (2) 将收集到的统计资料用键盘输入到数据文件中, 如农村人均年收入、种植业产值比重、有效灌溉面积比重等指标.

### 2.2 构造邻接矩阵

以龙游县的三十二个村镇为分区单元, 对该县种植利用分区进行区划, 首先构造邻接矩阵  $|E|$ :

$$a_{ij} = \begin{cases} 1, & < V_i, V_j > \in |E|; \\ 0, & \text{其它.} \end{cases}$$

式中 1 代表分区单元向量, 0 代表不相邻.

工作中是利用 ARC/INFO 的 clean 命令所建立的拓扑关系来自动生存邻接矩阵, 下面为 ARC/INFO 中邻接关系表示 (见图 2).

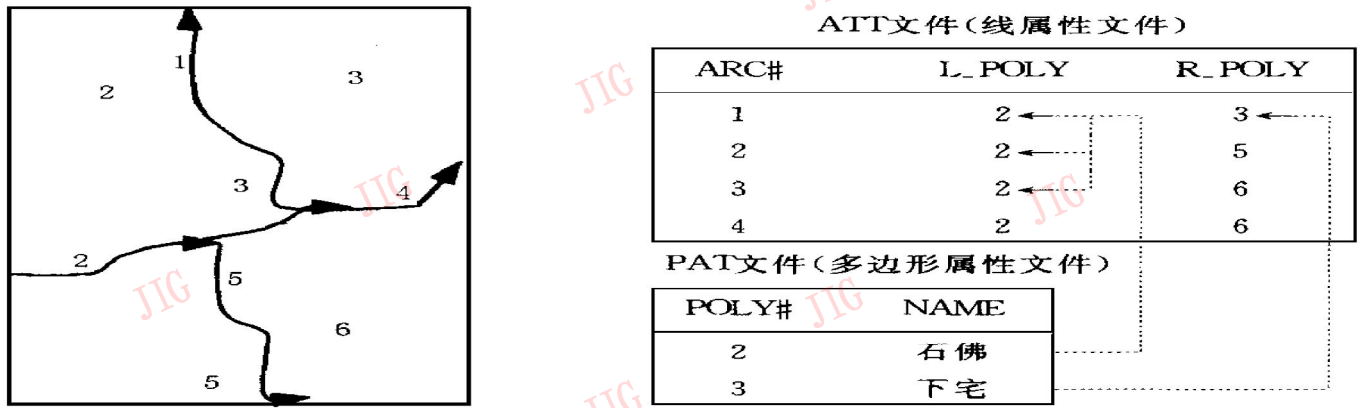


图2 ARC/INFO 中的邻接关系

### 2.3 构造权矩阵

在图论几何表示中, 每条边的长度可作为各端点间的权值, 这在网络的最佳路径问题中经常使用, 而种植利用分区中的权值就是各单元间的内在相似性, 通常构造这种相似性矩阵可用欧氏距离、相关系数、相似系数等来表示. 本研究采用相似系数法来构造权矩阵  $D$ :

$$d_{ij} = \sum_{k=1}^m \frac{x_{ki} - \bar{x}_i}{S_j} \times \frac{x_{kj} - \bar{x}_j}{S_j} \quad i, j = 1, 2, \dots, n \quad (7)$$

$$S_i^2 = \sum_{k=1}^m x_{ki}^2 \quad S_j^2 = \sum_{k=1}^m x_{kj}^2$$

式中,  $n$  为分区单元数,  $m$  为分区指数数.

### 2.4 求最小生成树

将权矩阵  $D$  和邻接矩阵  $E$  相乘, 使不邻接的分区单元赋权值为零, 运算后得到龙游县种植利用相邻分区单元的权值共 71 个.

图论中生成树是指包含图的所有顶点且连通的无圈图, 某树的树枝上的权之和称为该树的权. 求最小生成树就是寻找到最小权的树. 在利用最小生成树进行农业区划时, 首先将种植利用分区所研究的区域作为图  $G$ , 然后求图的最小生成树, 就是求区域总体相似系数最小的分区单元连接图. 目前求最小生成树的算法有枚举法、Kruskal 法、Prim 法等. 本研究中为便于算法的计算机编程, 选用了 Kruskal 法, 该算法的基本思路是在不构成回路的条件下“择优录取”权小的边<sup>[5,6]</sup>.

### 2.5 最小生成树分割

根据区内分异小、区间分异大的原则, 选择适当的  $A$  值, 将最小生成树分割成  $m$  个子树. 分割时要符合区间分异大的原则, 应选择各分区单元之间权值(相似系数)为最小的连接边, 并将此作为最小生成树

的分割点, 使得分割成的子树相似性最小, 分异最大. 通过对龙游县农业种植利用分区最小生成树分析, 权值(相似系数)最小的连接边依次为  $d(2, 6) = 0.108$ ,  $d(5, 11) = 0.256$ ,  $d(21, 24) = 0.368$ . 在进行区划时, 若将  $A$  值选择为 0.4, 则龙游县被分成四个农业种植利用区, 其中  $d(2, 6)$  的分割是不必要的, 因为它将塔石村单独划为一个分区, 其原因是龙游县北部 95% 以上的地形是低于 100m 的丘陵和平原, 只有塔石村大量分布着几百米高的山地, 最高达到 900 多米, 因此在最小生成树运算中出现  $d(2, 6)$  分割点. 由于区划的基本点是对研究区进行成片分割, 每个区都包括数个单元, 因此将塔石村(区划单元之一)单独划成一个分区是不符合分区逻辑的, 所以此断点不取. 最后选择  $A$  为 {0.2, 0.4} 的分割范围值, 得  $d(5, 11)$  和  $d(21, 24)$  两个分割点(见图 3 黑色箭头所示), 将龙游县分成三个农业种植利用区(见图 4).

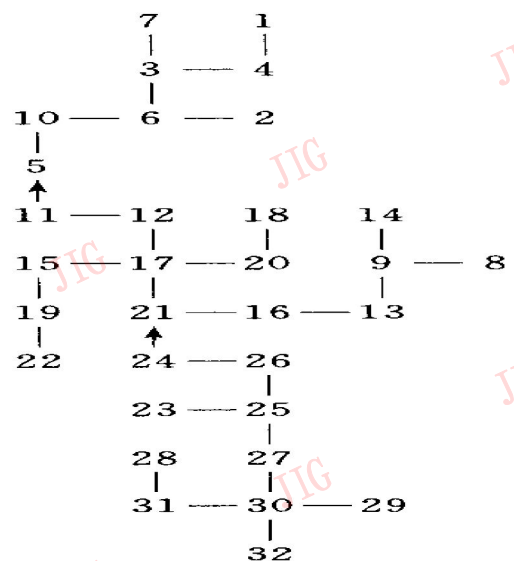


图3 农业种植利用分区最小生成树

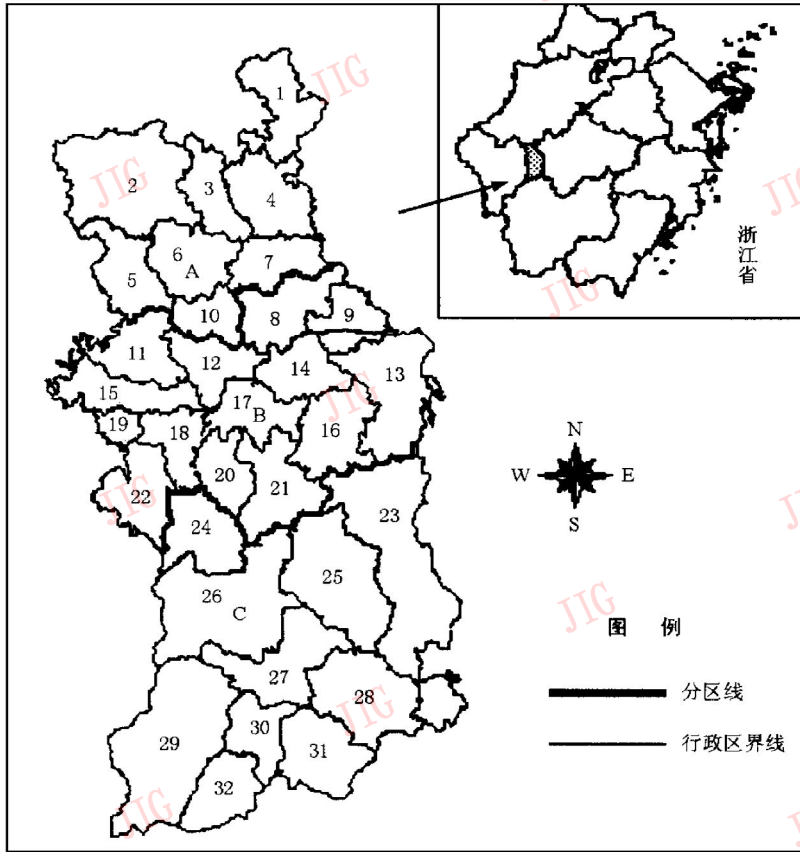


图 4 用最小生成树法区划的龙游县农业种植利用分区

### 3 结 论

(1) 龙游县农业局在综合分析影响本县农业发展的土地、气候、水资源因素的基础上, 根据地域分异规律, 将该县农业用地划分为三个种植区(见图 5 所示), 即龙北丘陵粮、油、桔种植区(图中 A); 龙中丘陵平原粮、果、棉、桑种植区(图中 B); 龙南山地茶、果、粮种植区(图中 C). 当地农业局在进行区划时, 没有完全依乡镇行政边界进行划分, 而是在保持一定行政区界完整性的同时, 采用河流和山脊等明显的地形分界线, 这是符合区划原则的. 由于区划人员对当地种植业发展条件与环境相当了解, 同时有一定的区划经验, 所以区划结果能较好的符合地域分异规律, 这也是与本文所用方法区划结果稍有差异的原因(见图 4、图 5). 将本研究所得的结果与当地采用传统方法区划的结果相比较, 三个分区界线基本相似. 说明本研究结果能较好地反映当地农业生产条件的地域差异, 方法是可行的.

(2) 采用空间分析与图论树算法相结合的方法进行农业分区, 可在计算机上自动实现, 方法上具有普遍性. 在建立有关空间数据库的基础上, 自动分区

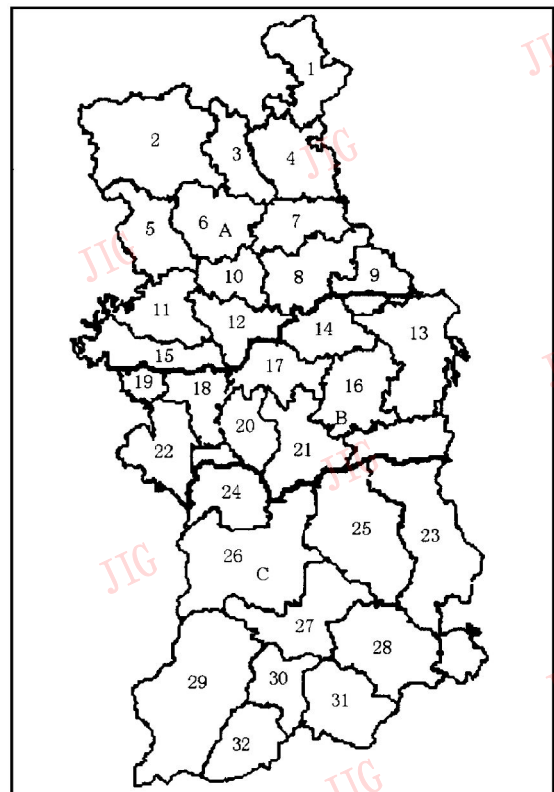


图 5 龙游县当地区划人员作的农业种植利用分区

模型运算可以在不到一个小时完成, 比普通定性方法效率明显提高, 这在分区结构调整中具有很大优

越性.同时也较好地解决了单纯利用数学模型所存在的两个问题.

(3) 由于各地的自然经济差异是模型所不能全部涵盖的,所以本研究所提出的自动分区方法还不能完全代替区划人员的工作.但不可否认,它是当前区划工作有力的辅助支持工具,同时也是区划研究的新方向.

### 参 考 文 献

- 1 周立三. 中国农业区划的理论与实践. 合肥: 中国科技大学出版社, 1993, 284~ 286.
- 2 姚建衢. 农业地域类型划分的聚类分析. 地理科学, 1988, 8(2): 147~ 155.
- 3 蔡继详, 李道远. 湖南省柑桔适种气候区的二级综合评判. 经济地理, 1987, 7(2): 110~ 115.
- 4 倪建华, 陈绥阳, 孙九林主编. 区域资源开发模型系统. 北京: 中国科学技术出版社, 1992.
- 5 范逢曦. 图论方法及应用. 太原: 山西科学教育出版社, 1987.
- 6 卢开澄, 卢华明. 图论及其应用. 北京: 清华大学出版社, 1995.



**史舟** 1970年生, 博士, 浙江大学农业遥感与信息技术应用研究所. 主要研究方向为遥感与地理信息系统技术在农业资源研究方面的应用.



**吴宏海** 1971年生, 讲师, 浙江省资源与环境信息系统重点实验室. 主要研究方向为地理信息系统技术开发与应用.

**王人潮** 浙江大学教授, 博士生导师. 致力于农业遥感与信息技术应用的科教工作. 曾获国家和省(部)级科技进步奖9项.