

基于几何概率的聚类分析方法及其在遥感影像分类中的应用

黄利文¹⁾ 毛政元^{1),2)} 李二振²⁾ 汪小钦¹⁾ 吴升²⁾

¹⁾(福州大学空间数据挖掘与信息共享教育部重点实验室,福州 350002)

²⁾(福州大学福建省空间信息工程研究中心,福州 350002)

摘要 针对现有非监督分类方法不能自动确定最佳分类数、对包含噪声的大数据集适应性差的问题,提出了一种基于几何概率的聚类分析方法,即按照先分大类、后分小类、逐层细分的顺序来确定分类方案,其同一分类层次上不同子类进一步细分的步骤相同,但执行过程彼此相互独立。在每一分类层次上,以几何概率为理论基础,根据样本在特征空间中的分布结构确定类的数目、提取类的中心位置、搜索类的边界。通过TM遥感影像的分类实例及其与ERDAS中的监督、非监督分类方法进行对比的结果表明,基于几何概率的聚类分析方法能明显提高分类精度。

关键词 聚类分析 几何概率 遥感影像分类

中图法分类号: TP391.4 TP751 文献标识码: A 文章编号: 1006-8961(2007)04-0633-08

The Cluster Analysis Approaches Based on Geometric Probability and Its Application in the Classification of Remotely Sensed Images

HUANG Li-wen¹⁾, MAO Zheng-yuan^{1),2)}, LI Er-zhen²⁾, WANG Xiao-qin¹⁾, WU Sheng²⁾

¹⁾(Key Laboratory of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fuzhou 350002)

²⁾(Spatial Information Research Center of Fujian Province, Fuzhou University, Fuzhou 350002)

Abstract Current unsupervised classification methods unable to determine optimum classes and poorly suitable for large and noise-included datasets. A cluster analysis approach based on geometric probability has been put forth, which gradually generates a hierarchical classification scheme in top-down order. The step of further classification for different sub-classes at the same classification level is identical but the executive process is independent for each other. Determining class number, extracting class central position, and searching the boundaries between different classes are performed at each classification level according to the distribution structure of samples in the feature space under the theoretical support of geometric probability. The algorithm of cluster analysis based on geometric probability is contrasted to ERDAS built-in supervised and unsupervised classification algorithms via a case of classifying a TM (thematic mapper) remotely sensed image. It turns out that the one based on geometric probability can obviously improve the classification accuracy.

Keywords cluster analysis, geometric probability, classification of remotely sensed images

1 引言

聚类分析是一种被广泛应用的探索型数据分析

方法,经数十年的发展已形成了非常庞大的内容体系,而且实现的算法很多^[1-4]。按照聚类算法实现的思路,目前应用较多的算法主要是以下3种类型:

(1)采用系统聚类法的思想;(2)采用动态中心聚类

基金项目:国家重大基础研究前期研究专项(2003CCA02100);国家自然科学基金项目(40471113)

收稿日期:2005-10-15;改回日期:2006-01-16

第一作者简介:黄利文(1979~),男,硕士研究生(现在泉州师范学院理工学院公基部任教)。主要研究数理统计。E-mail: livern@126.com。通讯作者:毛政元,男,博士,教授,博士生导师。E-mail: zymao@fzu.edu.cn

法的思想;(3)采用软计算方法的思想,比如模糊集、模拟退火算法等。其中第1类方法的优点是可以得到一个聚类树状图,在不同的相似水平上截断能得到不同的聚类方案,其缺点是抗噪声能力差、内存开销大,不能处理大样本^[5]。动态中心聚类算法虽速度快、占用内存少,适合处理大样本,也具有一定的抗噪声能力,能够较好地克服系统聚类法存在的弱点,但需要预先指定聚类中心,对非球状集聚的复杂样本适应性差,聚类过程容易收敛于局部最优^[6]。软计算方法聚类虽然能够改良动态中心聚类法的某些弱点^[7-9],但一般要求预先指定分类数和一些参数,聚类结果容易受到主观因素的影响^[10-12]。

遥感影像分类是聚类分析重要的应用研究专题之一^[13,14]。由于传感器本身存在误差、天气变化的干扰、地物空间分布和季相变化的复杂性等原因,遥感影像中普遍存在同物异谱和异物同谱的问题。对遥感数据这种包含噪声的大容量复杂样本,采用现有的聚类分析算法直接得到的分类结果中错分的情况较多,一般低于监督分类的精度^[15,16]。本文提出一种基于几何概率的聚类分析方法^[17,18],试图解决现有聚类分析方法在处理包含噪声的大容量复杂样本方面存在的问题,提高分类的精度。该方法运用几何概率直接根据像元在特征空间(以像元的波段为坐标轴,然后以与波段对应的灰度值为坐标的抽象空间)中的分布特征决定分类数,按照先分大类、后分小类、逐层细分的顺序确定分类方案。

2 聚类过程

基于几何概率的聚类分析采用先分大类、后分小类的分类思想,是一个自顶向下逐步求精的过程,其在第1分类层次上,分类过程只需进行1次,假设在第1分类层次上样本被分为 c 个子类,则在第2分类层次上,就需要针对 c 个子类中的每一个进行1次独立的再分类,这样同一分类过程共重复 c 次,依此类推,每一个分类层次上需要分类的次数等于上一个分类层次上得到的子类数。在某一分类层次上针对某一子类分类的过程中,若发现该子类不宜进一步细分,则该子类下的分类终止;如果在某一分类层次上所有的子类都不宜进一步细分,则分类完成。

2.1 第1层次的分类

各个分类层次上的每一次分类都包括选择数据

源、类的判识、确定类的分界线和归类4个步骤,以下以遥感影像第1分类层次为例来详细介绍每一个步骤的具体过程。

2.1.1 选择数据源

遥感影像通常由若干波段组成,每一个波段记录在该波段范围内地物反射或发射电磁波的强弱,基于几何概率的聚类分析在每次聚类时仅选取其中两个波段的数据,由于采用了先分大类、后分小类的分类思想,数据选取具有高度的灵活性,即在针对不同层次或同一层次不同子类细分的过程中,既可以选取相同的两个波段,也可以选取不同的两个波段,还可以选取一个相同、另一个不同的两个波段。具体究竟选择哪两个波段,与特定分类层次上的任务有关,可以采用以下两种方法决定:(1)根据经验或地物光谱特征;(2)通过对不同波段组合所得分类结果的比较来确定优选组合,如果所有的波段两两组合的分类结果都不佳,还可以通过对已有的两个波段或多个波段的数据进行加工得到含有多个波段信息的合成波段,并将其作为分类数据。

本文采用2002年冬季厦门九龙江口的TM遥感影像为试验数据,图幅大小为 1498×1281 pixels,对应区域的典型地物包括水体、居民地、裸露岩石、农田、植被和道路,试验时,采用第4和第5波段的数据作为第1分类层次的数据源。

2.1.2 类的判识

本步骤的主要任务是确定类的数目并提取每一类的中心点。将2.1节中选取的每一个像元视为样本中的一个样品,并将两个波段的灰度值视为样品在2维空间中的坐标,而由TM遥感影像的第4和第5波段构成的2维样本数据则可以视为分布于平面上边长为255级灰度的正方形区域内的一个点集 S ,基于几何概率聚类的核心思想是通过分析与2维样本数据对应的平面点集的空间分布规律来确定类的数目并提取每一类的中心。

由于TM图像的灰度值为整数,因此 S 中的点只可能分布在边长为255级灰度的正方形区域上坐标值取整数的位置上(总数为 256×256),密度统计时,只需针对坐标值取整数的每一个位置,先统计 S 中落在其上的点数,同时显示正方形区域内至少对应 S 中一个点的位置(如图1(a)所示),然后以各位置上的点数作为第3维数据,通过插值生成的等值线(如图1(b)所示)显示 S 中的点在正方形区域上的分布规律。

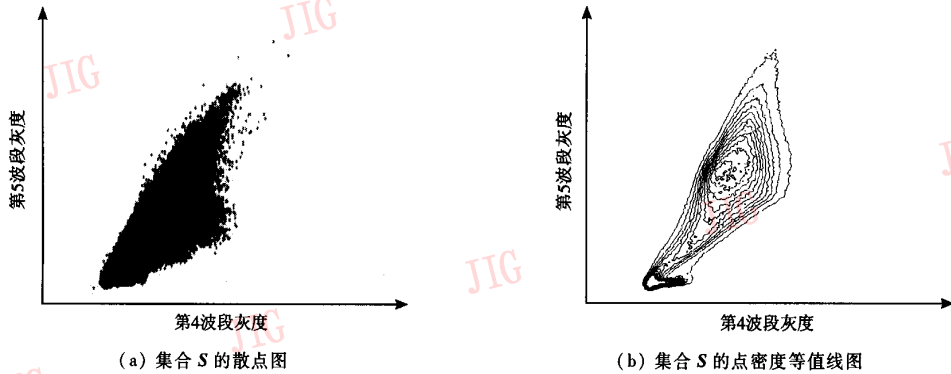


图 1 集合 S 的散点图和点密度等值线图

Fig.1 The scattered points diagram and the point density contour diagram of S

观察图 1(b)可以直观地看出 S 中的点在正方形区域上并不是均匀分布(若是均匀分布,则聚类分析无意义),而是有两个明显的一级集聚中心,由于二者之间存在重叠部分,且不是球状集聚,因此采用动态中心法或者采用以距离(或相似程度)为归类指数的各种聚类方法都不可能取得好的分类结果,而基于几何概率的聚类分析却能通过计算找到这两个集聚中心,并进而确定二者合理的分界线。

记点集 S 中的点数为 n,将 S 中的 n 个点两两相连得到一个线段集合

$$L(\theta_i, i = 1, 2, 3, \dots)$$

其中, θ_i 表示 L 中第 i 条线段与纵轴正向的夹角,且有

$$0 < \theta_i \leq \pi, i = 1, 2, 3, \dots$$

针对 L 构造下列函数

$$H(\theta, \Delta\theta) = \frac{N}{T} \times \frac{1}{P(\theta, \Delta\theta)}$$

式中, T 表示 L 中线段的总数,且

$$T = C_n^2 = n(n-1)/2$$

N 表示 L 中出现在区间

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2} \right]$$

上的线段数,其可以由统计求得。函数 P 的含义是在随机条件下,边长为给定值 a 的正方形区域内线段出现在区间

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2} \right]$$

上的概率,函数 H 的含义是集合 L 中线段落入区间

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2} \right]$$

的频率与完全随机情况下线段落入同一区间的概率

之比,当 S 中的点为随机分布或单中心集聚分布时,函数 H 的图形为一条没有明显起伏的曲线,当 S 中的点为多中心集聚分布时,函数 H 的图形将存在若干峰值,峰值的高度代表集聚的程度,峰值的数目与集聚的中心数及其分布有关^[1]。

由 2.1.1 节可知, $n = 1498 \times 1281, a = 255, \theta_i (i = 1, 2, \dots, T)$ 的值可根据点集 S 中各点的坐标逐一计算得到,函数 P 的值可利用几何概率的计算方法解算^[18]。将 $n, a, \theta_i (i = 1, 2, \dots, T)$ 和函数 P 的值代入函数 H,按预先设定的 $\Delta\theta$ 和步长 h 计算函数 H 的值并画出图形。本例中取 $\Delta\theta = \pi/10, h = 0.01$,画出的图形如图 2 所示,该图只有一个明显的峰值,此峰值对应的集聚方向 $\theta = 2.55\text{rad}$ (即集聚方向与 y 轴正向所成夹角的弧度,下同)。

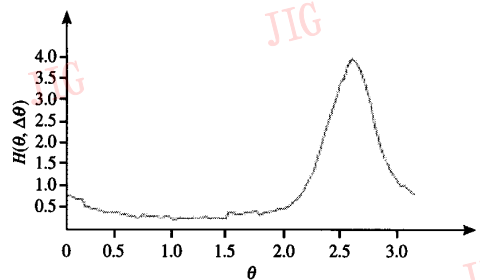


图 2 函数 H 的图形 ($\Delta\theta = \pi/10, h = 0.01$)

Fig.2 The diagram of function $H(\Delta\theta = \pi/10, h = 0.01)$

在计算函数 H 值的过程中, $\Delta\theta$ 及步长 h 的取值要适中,一般 $\Delta\theta$ 应在 $[\pi/36, \pi/10]$ 之间取值, h 应在区间 $[0.01, 0.1]$ 上取值,否则,函数 H 的图形不能确切地指示数据集在某个特定层次上的集聚方向,导致结果失真。图 3 显示了角度区间宽度 $\Delta\theta$ 及步长 h 取不同值时函数 H 的图形,由于 $\Delta\theta$ 和 h 的值不合适,

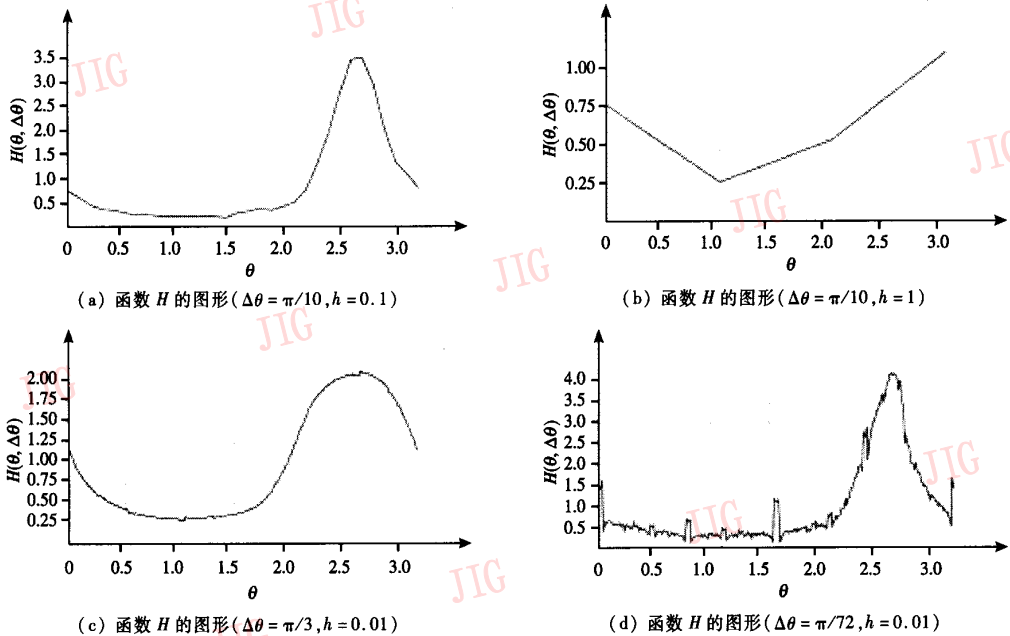


图 3 不同 $\Delta\theta, h$ 取值下函数 H 的图形

Fig. 3 The diagram of function H for different $\Delta\theta$ and h

图形比较粗糙,因此准确的峰值不易确定。

点集 S 的集聚中心按以下步骤提取:(1)以一宽度为 d_1 的平行条带沿着集聚方向在特征空间(点集 S 分布的正方形区域)内对点集 S 以步长 h_1 进行扫描,当平行条带中点的密度达到最大时,就记录平行条带所在的位置,并记此时平行条带在特征空间的区域为 D ;(2)以另一宽度为 d_2 的平行条带沿着集聚方向的垂直方向对区域 D 以步长 h_2 进行扫描,并将每次扫描时条带中点的密度显示在图上;(3)根据图上的峰值判断中心点的个数,并计算出中心点的坐标。提取集聚中心的过程中,扫描宽度

和扫描步长的取值要适中,太小,则计算量大、计算过程耗时长;太大,则结果会出现误差。当取扫描宽度为正方形边长的 $1/50 \sim 1/40$ 、扫描步长为正方形边长的 $1/150 \sim 1/120$ 时,精度和速度都能满足要求。本例中取 $d_1 = 4, h_1 = 1, d_2 = 4, h_2 = 1$,两个集聚中心的坐标分别为 $(9, 5), (49, 65)$,图 4 显示了特征空间中集聚方向和高密度集聚带内的集聚方向垂直的点密度变化。当存在多个集聚方向时,则要先逐一求出每个集聚方向上存在的集聚中心,经匹配并消去重复记录的集聚中心后才是需要的结果。本例中只有一个集聚方向,故无此步骤。

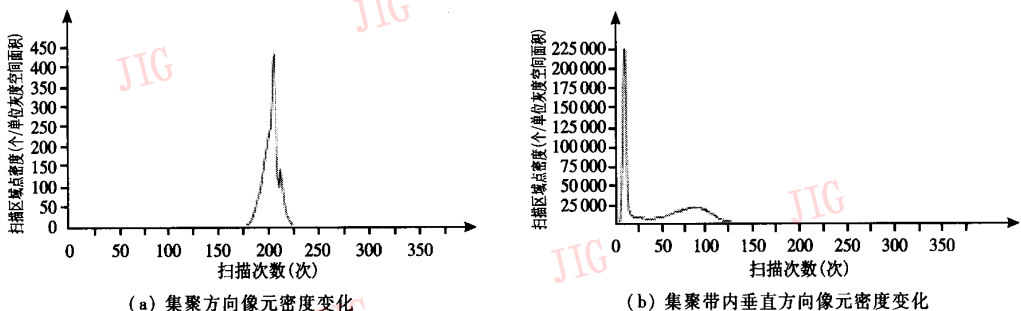


图 4 特征空间中像元密度的变化

Fig. 4 Variation of pixel density in the feature space

2.1.3 确定类的分界线

类的分界线直接影响分类精度,是基于几何概

率聚类的重要步骤。由于同类地物像元的灰度值大致符合正态分布,点集 S 中点的密度从集聚中心向

外递减,根据贝叶斯分类理论,可将两集聚中心连线之间密度最小处作为两类的分界点。类的分界线的确定是先将2.2节中得到的两个集聚中心(9,5),(49,65)连成一条直线,然后在该线段上找到点密度最小处,再以此处为起始点,沿着集聚中心连线的垂直方向开始向两侧搜索局部密度谷点,当遇到特征空间的边界时停止,最后连接每次搜索得到的各局部密度谷点即得到类的分界线(图5)。

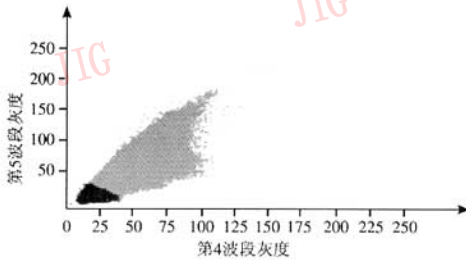


图5 类的分界线

Fig. 5 Classification boundary

本分类实例中只有两个集聚中心,确定分界线的过程相对比较简单。当存在3个以上的集聚中心时,则确定分界线的过程如下:先以各集聚中心为顶点生成Delaunay三角网;然后在三角网的每一个三角形内部,同时从3条边上的密度最低处开始向三角形内部搜索局部密度谷点,形成3个搜索方向,每个搜索方向每次前进一步,当任意两个搜索方向相遇时,则这两个搜索方向上的搜索停止,此时即可连接在这两个搜索方向上搜索得到的局部密度谷点形成一条分界线,而在另一个搜索方向上则继续搜索局部密度谷点,当遇到已经生成的分界线时停止,则通过连接在该搜索方向上搜索得到的局部密度谷点来形成另一条分界线,该分界线将先前形成的分界线一分为二,它们共同确定了三角形内部点的归属;在特征空间中,Delaunay三角网未覆盖的部分,对三角网中最外层三角形的外边界,可依次从每一边界上的密度最低处开始搜索局部密度谷点,当遇到特征空间的边界或已经生成的分界线时停止,连接搜索得到的局部密度谷点形成各分界线。

2.1.4 归类

对搜索得到的各边界线之间、边界线与特征空间边界之间进行相交剪断运算,并以得到的线段集为弧段重建拓扑关系,则得到的各个子区对特征空间的分割就代表一种分类方案,每个子区代表一类,

计算并统计落入各个子区的像元,即得到确定的分类方案。将本例第1分类层次的结果显示在遥感影像上,其中在特征空间位于分界线左下侧的像元用黑色显示,位于分界线右上侧的像元用白色显示(图6),二者的地理界线与水陆边界大体吻合,黑色区域基本为水体(以下称该类为 A_1),白色区域基本为陆地(以下称该类为 A_2)。

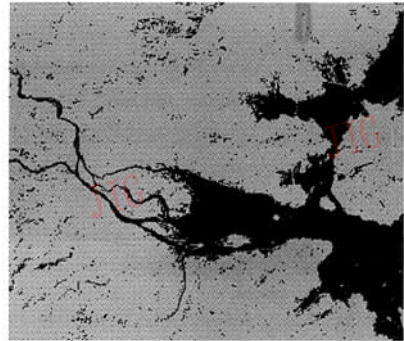


图6 根据分界线得到的类

Fig. 6 Classifications derived from the boundary

2.2 其他层次的分类

利用2.1节中的方法对第1层次得到的两类进行细分。其中对 A_1 第2层次的细分是先采用TM第1波段和第2波段作为数据源,函数 H 指示的集聚方向为2.0,扫描得到的集聚中心分别为(51,20)和(71,35), A_1 被分为水体(以下称为 B_1)和阴影;然后采用TM第2波段和第3波段的数据对 B_1 进行第3层次的细分,函数 H 指示的集聚方向为2.8,扫描得到的集聚中心分别为(31,28)和(34,35),这样 B_1 被分为清洁水体(简称清水)和浑浊水体(简称浊水)。

对 A_2 第2层次的细分先采用第3波段和3,4两波段的合成数据(植被指数)作为数据源,函数 H 指示的集聚方向为0.31,扫描得到的集聚中心分别为(32,164)和(39,138), A_2 被分为植被(以下称为 B_2)和非植被(以下称为 B_3);然后利用第4波段和第5波段的TM数据分别对 B_2 和 B_3 进行第3层次的细分,其中 B_2 被分为浓密植被(简称密植被)和稀疏植被(简称稀植被)(函数 H 指示的集聚方向为2.71,扫描得到的集聚中心分别为(37,30)和(54,67)); B_3 被分为居民地和旱地(函数 H 指示的集聚方向为2.74,扫描得到的集聚中心分别为(26,24)和(43,64))。

3 聚类结果及对比分析

图 7(a)为待分类 TM 数据第 5 波段、第 4 波段和第 3 波段合成的假彩色遥感影像,用本文介绍的算法聚类后,最终分成 7 类,分别是居民地、阴影、清洁水体、浑浊水体、浓密植被、稀疏植被和旱地(图 7(b))。

作为对比,利用 ERDAS (earth resources data analysis system)对同一图像分别进行监督分类(图 7(c))和非监督分类(图 7(d)),初始分类为 21 类,最终合并为 7 类,然后从遥感影像中随机抽取 250pixels,经检验,监督分类、非监督分类和基于几何概率聚类 3 种分类方法对被抽到的 250pixels 的分类结果如表 1 所示(表中加粗的数字表示对应地物类中归类正确的像元数),最后再根据表 1 中的数字计算每一种

分类方法的分类精度、Kappa 系数、总分类精度、总 Kappa 系数,计算结果列于表 2 中。分类精度和 Kappa 系数是评价遥感影像分类精度常用的定量指标,其中分类精度等于抽取的像元中归类正确的像元数与被抽取的像元总数之比,可以针对单个地物类型或全部地物类型进行计算,后者即总分类精度;Kappa 系数代表被评价分类比完全随机分类产生错误减少的比例,某种地物类型的 Kappa 系数 K_i 和总 Kappa 系数 K 的计算公式分别为

$$K_i = \frac{N x_{ii} - x_{i+} x_{+i}}{N x_{i+} - x_{i+} x_{+i}}$$

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} x_{+i})}$$

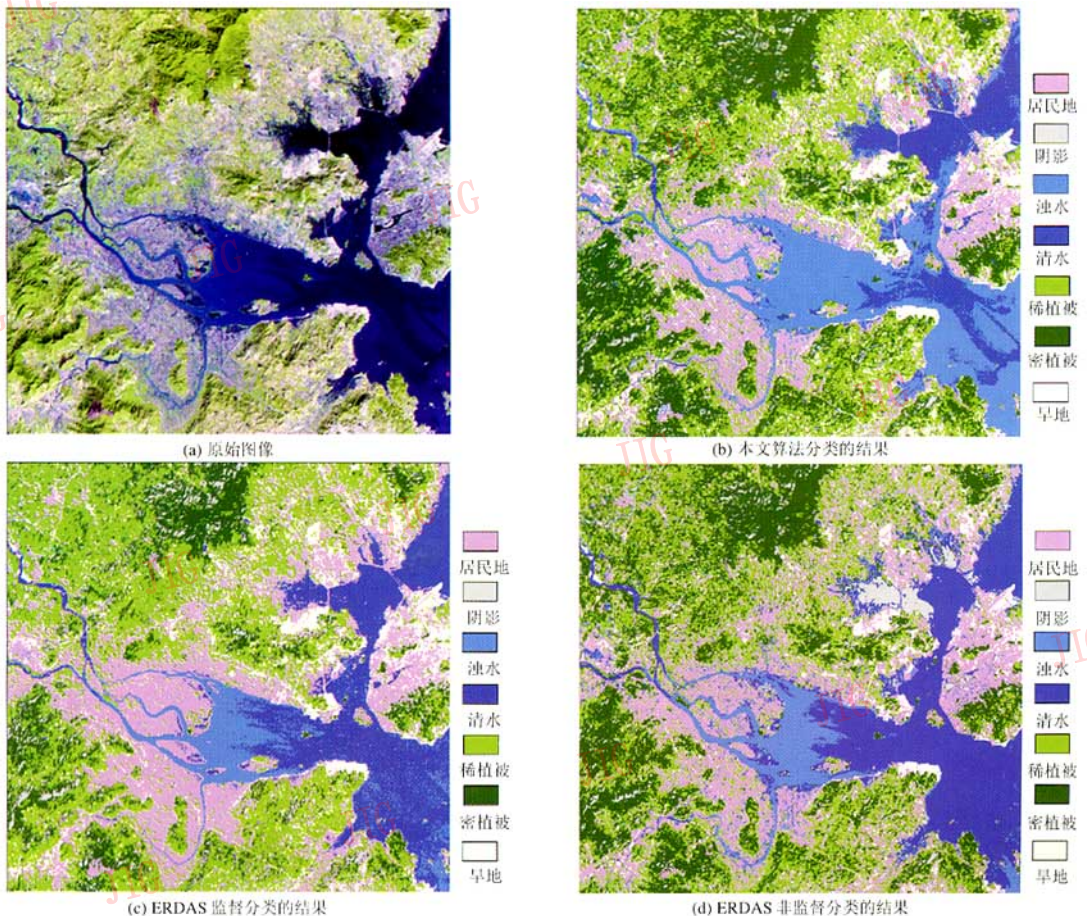


图 7 原始影像及不同方法分类结果

Fig. 7 The original image and classification results derived from different classification methods

其中, K_i 是第 i 行中地物的 Kappa 系数, K 是总 Kappa 系数, r 是误差矩阵的行数, x_{ii} 是第 i 行第 i 列 (主对角线) 上的值, $x_{i.}$ 和 $x_{.i}$ 分别是第 i 行的和与第 i 列的和, N 是样点总数 (即被抽到的全部像元数)。

表 1 3 种方法对随机抽取的 250pixels 的分类结果比较

Tab.1 The class labeling of 250 randomly selected pixels of the three classification methods

类别	ERDAS 监督分类方法							ERDAS 非监督分类方法							基于几何概率的聚类分析方法									
	居民地	阴影	浊水	密植被	疏植被	清水	旱地	总数	居民地	阴影	浊水	密植被	疏植被	清水	旱地	总数	居民地	阴影	浊水	密植被	疏植被	清水	旱地	总数
居民地	40	0	15	3	11	3	4	76	43	0	6	0	12	0	14	75	40	0	7	0	6	0	3	56
阴影	0	1	0	1	0	0	0	2	0	1	0	0	0	8	0	9	0	1	0	1	0	0	0	2
浊水	0	0	14	0	0	1	0	41	0	0	16	0	1	1	0	18	0	0	31	0	2	5	0	38
密植被	0	1	0	19	1	0	0	21	0	1	0	28	6	0	0	35	0	1	0	37	3	0	0	41
疏植被	8	0	2	18	52	0	3	83	6	0	2	13	43	0	1	65	7	0	2	4	53	0	2	68
清水	0	0	10	0	0	29	0	39	0	0	17	0	2	24	0	43	0	0	1	0	0	27	0	28
旱地	1	0	0	0	0	0	12	13	0	0	0	0	0	0	4	4	2	0	0	0	0	0	14	16

表 2 3 种分类方法对随机抽取的 250pixels 的分类精度

Tab.2 The classification accuracy of the three classification methods for 250 randomly selected pixels

	分类精度 (%)			Kappa 系数			结果评价		
	监督分类	非监督分类	本文方法	监督分类	非监督分类	本文方法	监督分类	非监督分类	本文方法
居民地	52.63	57.33	71.43	0.4108	0.4693	0.6446	总分类精度为 67.20%, 总 Kappa 系数为 0.5936	总分类精度为 64%, Kappa 系数为 0.5586	总分类精度为 81.6%, 总 Kappa 系数为 0.7742
阴影	50.00	11.11	50.00	0.4960	0.1039	0.4960			
浊水	93.33	88.89	81.58	0.9203	0.8671	0.7797			
密植被	90.48	80.00	90.24	0.8861	0.7608	0.8833			
疏植被	62.65	66.15	77.94	0.4980	0.5451	0.7035			
清水	74.36	55.81	96.43	0.7046	0.4909	0.9589			
旱地	92.31	100.00	87.50	0.9167	1.0000	0.8647			

4 结 论

基于几何概率的聚类分析采用先分大类、后分小类的顺序完成分类过程, 针对同一分类层次上不同子类, 其进一步细分的过程基本相同, 且相互独立, 可以采用样本不同的属性值。确定类的数目和集聚中心是基于几何概率进行聚类分析的核心内容, 其依据的是样本在特征空间中的分布结构, 运用几何概率能够直观地发现这种结构, 为确定类的数目和集聚中心提供客观依据。类的边界搜索是基于几何概率聚类的重要步骤, 其依据是特征空间中相邻集聚中心两两之间样品分布的密度。

由表 2 可知, 基于几何概率的聚类方法除旱地和浊水的分类精度稍低、浓密植被的分类精度大致与监督分类持平外, 其余各地物类型的分类精度均

明显高于遥感图像处理软件 ERDAS 中已经实现的监督、非监督分类方法。

基于几何概率的聚类方法保持了非监督分类所具有的全自动特征, 不需根据先验知识选定训练样本。与其他非监督分类方法相比, 该算法能更好地适应非球状集聚的复杂样本, 自动抑制大容量样本中的噪声, 确定最佳分类层次和每一分类层次上的最佳分类数, 得到正确的分类结果。本文提出的算法对 2 维数据集的聚类分析具有通用性, 通过扩展使其适应高维数据集将是下一步的研究目标。

参考文献 (References)

- Jain A K, Murty M N, Flynn P J. Data clustering: a review [J]. ACM Computing Surveys, 1999, 31(3): 265 ~ 323.
- Estivill-Castro V, Lee I. AUTOCLUST +: Automatic clustering of point-data sets in the presence of obstacles [A]. In: Proceedings of

- International Workshop on Temporal, Spatial and Spatial-temporal Data Mining[C], Lyon, France, 2000: 133 ~ 146.
- 3 Wang S T, Jiang H F, Lu H J. A new integrated clustering algorithm GFC and switching regressions[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2002, 16(4): 433 ~ 446.
 - 4 HUANG Kai-yi. The use of a newly developed algorithm of divisive hierarchical clustering for remote sensing image analysis [J]. International Journal of Remote Sensing, 2002, 23(16): 3149 ~ 3168.
 - 5 Everitt B S. Unsolved problems in cluster analysis[J]. Biometrics, 1979, 35(1): 169 ~ 181.
 - 6 Murray A T, Estivill-Castro V. Cluster discovery techniques for exploratory spatial data analysis [J]. International Journal of Geographical Information Science, 1998, 12(5): 431 ~ 443.
 - 7 Bandyopadhyay S, Maulik U, Pakhira M K. Clustering using simulated annealing with probabilistic redistribution[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2001, 15(2): 269 ~ 285.
 - 8 Dimitriadou E, Weingessel A, Hornik K. A combination scheme for fuzzy clustering[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2002, 16(7): 901 ~ 912.
 - 9 GAO Xin-bo. Fuzzy Cluster Analysis and Its Application [M]. Xi'an: Xi'an Electronic Science and Technology University Press, 2004. [高新波. 模糊聚类分析及其应用 [M]. 西安: 西安电子科技大学出版社, 2004.]
 - 10 YU Jian, CHENG Qian-sheng. The upper bound of the optimal number of clusters in fuzzy clustering [J]. Science in China (Series E), 2002, 32(2): 274 ~ 280. [于剑, 程乾生. 模糊聚类方法中的最佳聚类数的搜索范围 [J]. 中国科学 (E 辑), 2002, 32(2): 274 ~ 280.]
 - 11 Hardy A. On the number of clusters [J]. Computational Statistics & Data Analysis, 1996, 23(1): 83 ~ 96.
 - 12 Pakhira M K, Bandyopadhyay S, Maulik U. Validity index for crisp and fuzzy clusters [J]. Pattern Recognition, 2004, 37(3): 487 ~ 501.
 - 13 ZHANG Xiao-can, HUANG Zhi-cai, ZHAO Yuan-hong. The Processing of Remotely Sensed Digital Images [M]. Hangzhou: Zhejiang University Press, 1997. [章孝灿, 黄智才, 赵元洪. 遥感数字图像处理 [M]. 杭州: 浙江大学出版社, 1997.]
 - 14 Pakhira M K, Bandyopadhyay S, Maulik U. A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification [J]. Fuzzy Sets and Systems, 2005, 155(2): 191 ~ 214.
 - 15 JI M. Using fuzzy sets to improve cluster labelling in unsupervised classification [J]. International Journal of Remote Sensing, 2003, 24(4): 657 ~ 671.
 - 16 Duda T, Canty M. Unsupervised classification of satellite imagery: choosing a good algorithm [J]. International Journal of Remote Sensing, 2002, 23(11): 2193 ~ 2212.
 - 17 Seneta E, Parshall K H, Jongmans F. Nineteenth-century developments in geometric probability: J. J. Sylvester, M. W. Crofton, J. -É. Barbier and J. Bertrand [J]. Archive for History of Exact Sciences, 2001, 55(6): 501 ~ 524.
 - 18 MAO Zheng-yuan, LI Lin. The Measurement of Spatial Patterns and Its Application [M]. Beijing: Science Press, 2004. [毛政元, 李霖. 空间模式的测度及其应用 [M]. 北京: 科学出版社, 2004.]