

基于多尺度空间聚类的经济区域划分研究*

陈小瑜

(泉州师范学院 资源与环境科学学院, 福建 泉州 362000)

摘要:传统的聚类方法能够自动实现类型的划分,但其只考虑空间对象的属性特征,忽略了空间对象之间位置的邻近性。在尺度空间理论上,利用多尺度空间聚类方法,同时兼顾空间对象间的空间相邻和属性相似这两个特征,使得在空间尺度由小变大的过程中,具有高度空间相互依赖关系的空间对象相互融合,得到不同空间尺度下的区域划分结果。并以福建省67个县市1990—2004年历年的人均GDP数据分析为例,进行了全省的经济区域划分应用试验。结果表明与实际的区域经济发展情况相符。该方法能够综合考虑空间位置邻近性和属性特征相似性的聚类要求,使得聚类结果更加合理,为制定区域经济持续发展战略提供依据。

关键词:多尺度;空间聚类;福建;经济区域划分

中图分类号:F127;O212.1

文献标志码:A

文章编号:1672-669X(2011)05-0081-04

空间聚类分析是空间模式识别和空间数据挖掘的重要手段之一^[1]。传统的聚类方法能够自动实现类型的划分^[2-3],但其只考虑空间对象的属性特征,忽略了空间对象之间位置的相互作用,即聚类后的结果在空间上是离散分布的。传统空间对象本质上具有地理位置和属性特征双重含义,二者结合才能完整地描述空间特征和空间差异^[4-5]。基于尺度空间理论的聚类方法能够同时兼顾空间相邻和属性相似这两个特征。基于此,本文以福建省为例,进行经济水平区域划分研究。

1 尺度空间理论

尺度空间理论:首先将每一空间单元视作空间的一个光点,于是一组空间数据集构成一幅空间图像。当模糊化这一图像时,每一小光点首先变为一个小光斑。进一步地模糊使得小光斑逐渐溶为大光斑。当尺度充分低时,整个图像融合为一个光斑。如果将每个光斑看成一个类,则上述过程便形成逐级的聚类结构层,层的高度由空间尺度的倒数表示^[6]。

中科院骆剑承^[7]等在尺度空间理论上,提出多尺度空间单元区域划分方法,在考虑空间单元属性信息的同时,增加了空间单元的相互依赖关系,使得在空间尺度由小变大过程中,具有高度空间相

互依赖关系的空间单元相互融合,得到不同空间尺度下的区域划分。美国学者王法辉^[8]借鉴了骆剑承等的研究方法,应用尺度空间聚类法分析芝加哥就业便捷度与谋杀犯罪之间的关系。

尺度空间理论的发展得益于计算机图像处理技术的进步,但大多数应用在遥感的数据处理方面,把尺度空间聚类法应用到社会经济分析中的研究在国内外的相关研究较少。

2 多尺度空间聚类算法^[4]

以相邻关系作为空间单元之间的相互关系来说明一定尺度下空间单元的聚类过程。假设一个空间数据集中包括 n 个空间单元 $S = \{S_1, S_2, \dots, S_n\}$ 及其相邻的 m 个空间集为 $S' = \{S'_1, S'_2, \dots, S'_m\}$ 。每个地理空间单元都拥有 d 维的表示个体自然或人文的属性数据 $X = \{x_1, x_2, \dots, x_d\}$,可以用一个指标来表示空间单元 i 与其周围相邻单元的相互作用指标,最简单的指标可用属性之间的相关系数来表示。

$$r_{ij} = \frac{\sum_{k=1}^d (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^d (x_{ik} - \bar{x}_i)^2} \cdot \sqrt{\sum_{k=1}^d (x_{jk} - \bar{x}_j)^2}} \quad (i = 1, \dots, n; j = 1, \dots, m) \quad (1)$$

* 收稿日期 2011-05-04 修回日期 2011-06-20 网络出版时间 2011-09-17 13:59:00

资助项目:泉州师范学院校科研自选项目资助(No. 2010KJ08)

作者简介:陈小瑜,女,讲师,硕士研究生,研究方向为GIS应用。

网络出版地址: http://www.cnki.net/kcms/detail/50.1165.N.20110917.1359.201105.81_017.html

式中 r_{ij} 的取值范围为 0 ~ 1。通过判断 i 与 S' 中最相关的单元 k

$$r_{ik} = \max_{k=1}^m r_{ij} \quad (2)$$

建立单元 i 和单元 k 之间的连接通路。再确定通路的连接方向 D_{ik} 。首先根据单元的属性数据来表示其代表量值 (Q_i, Q_k), 比如可以简单地用各属性分量的加权累加作为其代表量值, 然后通过比较 Q_i 和 Q_k 来确定连接方向 D_{ik} : 当 $Q_i < Q_k$ $D_{ik} = 1$, 反之 $D_{ik} = 0$ 。这样, 每一个空间单元与其相邻的单元只有唯一的一个有向连接。

根据以上规则, 对所有空间单元建立与其相邻单元的连接关系, 形成完整的路径传递系统。然后按照下列步骤对相互有单向连接关系的单元进行融合归并:

1) 寻找一个初始空间单元 S_0 , 该单元没有其他单元指向它, 只有它指向别的单元;

2) 从 S_0 开始根据向外连接方向查找下一个单元 (S_1, S_2, \dots, S_i), 直到单元没有向外连接的单元 S_i 为止;

3) 把 0 ~ t 的所有单向连接归并为一个类别, 以最后终止单元 S_i 为其类别标志。并对以上已搜索的单元加上标志;

4) 对其他未搜索到的单元按以上步骤进行搜索, 直到全部单元都归并到一个类别。

这样得到第 1 相邻尺度下的融合聚类结果。各类别中心单元为没有向外连接的终止单元 (S_{i1}, S_{i2}, \dots)。该尺度下的各中心单元作为下一尺度下的单一空间单元, 按照以上融合规则进行下一步融合聚类。直到全部空间单元融合为一个类别。

3 应用实例

3.1 研究区概况和数据来源

本文以福建省作为研究区域。福建地处我国东南沿海, 是全国经济较发达的省份之一。海峡西岸经济区的提出与实施具有十分重大的战略意义, 为福建省的城市发展带来了新的机遇, 但面对“长三角”和“珠三角”的强烈“袭夺”, 福建省亟待确定各个城市的腹地范围, 以避免在发展空间上出现恶性竞争, 同时需要协调以城市为主体的各个区域之间的关系, 从而使海峡西岸经济区形成一个整体, 并与周边地区相抗衡^[9]。为此, 本文以福建省 67 个县市(金门除外, 见图 1)1990—2004 年历年的人均 GDP 指标为分析数据, 对全省的 67 个县市进行经济区域

划分。

3.2 聚类分析

3.2.1 空间邻接矩阵 对每个县市, 找出与其有相邻关系的县市群, 本文采用空间邻接矩阵来刻画各县市间的空间邻接关系。

对福建省以县市为单位的行政区图建立拓扑, 提取弧段属性表中的左多边形 ($Lpoly$) 和右多边形 ($Rpoly$) 两个属性, 其属性值为代表各县市的多边形的计算机标编码, 如表 1(部分内容从略), 将含有代表外多边形的数值“1”的记录去除掉, 数据如表 2(部分内容从略)。

表 1 福建省 67 个县市的编码

县市名	编码	县市名	编码	县市名	编码
浦城县	2	永安市	35	长汀县	39
武夷山市	3	武平县	48	上杭县	49
光泽县	4	建宁县	19	德化县	40
松溪县	5	明溪县	27	连城县	41
建阳市	6	永春县	47	永定县	55
寿宁县	7	大田县	36	漳平市	43
邵武市	8	将乐县	20	仙游县	45
政和县	9	古田县	22	安溪县	50
顺昌县	15	南平市	23	长泰县	59
屏南县	16	闽清县	29	柘荣县	12

表 2 福建省 67 个县市的行政拓扑关系

$Lpoly$	$Rpoly$	$Lpoly$	$Rpoly$	$Lpoly$	$Rpoly$
3	2	9	7	16	9
4	3	11	7	16	13
5	2	12	10	16	14
6	2	12	11	17	10
6	3	13	6	17	11
6	3	13	9	17	12
6	4	14	7	18	8
6	5	14	9	19	8
8	4	14	11	19	18
8	6	15	6	20	8
9	5	15	8	20	15
9	6	15	13	20	18

通过 Matlab 编程^[10], 建立空间邻接矩阵, 由于平潭县和东山县是岛屿, 与其他多边形没有公共的弧段, 但在本研究中并不能把它们孤立看待, 于是分别将平潭县定义为与其最临近的福清市邻接, 东山县定义为与其最临近的云霄县、诏安县邻接, 所以程序只计算其他 65 个县市的邻接矩阵。空间邻接矩阵的部分数据如表 3, 其中“1”代表相邻; “0”代表不

相邻。程序的算法如下：

$A = zeros(65)$;产生 65 × 65 的空间邻接矩阵 ,并初始化为 0

$$a = \begin{bmatrix} 3 & 2 \\ 4 & 3 \\ 5 & 2 \end{bmatrix}$$

将表 2 的数据赋予矩阵 a

For $i = 1 \sim n$ n 为表 2 的记录数

$$B(a(i,1), a(i,2)) = 1;$$

$$B(a(i,2), a(i,1)) = 1;$$

End

Dlmwrite('邻接矩阵.txt', A); 将矩阵 A 保存成文本格式

表 3 福建省 67 个县市的邻接矩阵

	2	3	4	5	6	7	8	9	10	...
2	0	1	0	1	1	0	0	0	0	...
3	1	0	1	0	1	0	0	0	0	...
4	0	1	0	0	1	0	1	0	0	...
5	1	0	0	0	1	0	0	1	0	...
6	1	1	1	1	0	0	1	1	0	...
7	0	0	0	0	0	0	0	1	0	...
8	0	0	1	0	1	0	0	0	0	...
9	0	0	0	1	1	1	0	0	0	...
10	0	0	0	0	0	0	0	0	0	...

3.2.2 相关系数 根据(1)式,求该县(市)与相邻各县(市)的发展相关系数。采用年递增速度(后一年减前一年,成为 14 维的属性数据)来求相关系数,使得相关系数值差异较大。部分数据如表 4。

表 4 67 个县市与其相邻县市的发展相关系数

	2	3	4	5	6	7	8	9	10	...
2	0	0.759	0	0.583	0.829	0	0	0	0	...
3	0.759	0	0.655	0	0.493	0	0	0	0	...
4	0	0.655	0	0	0.829	0	0.535	0	0	...
5	0.583	0	0	0	0.563	0	0	0.713	0	...
6	0.829	0.493	0.829	0.563	0	0	0.647	0.887	0	...
7	0	0	0	0	0	0	0	0.86	0	...
8	0	0	0.535	0	0.647	0	0	0	0	...
9	0	0	0	0.713	0.887	0.86	0	0	0	...
10	0	0	0	0	0	0	0	0	0	...

3.2.3 建立连接 根据相关性最大原则建立县市间的连接,然后根据代表量值(Q)之间的差异确定连接方向。本文用 15 年各县市的人均 GDP 的平均值来表示 Q 。各县市的 Q 值及其排序见表 5。

表 5 福建省 67 个县市 15 年人均 GDP 平均值排序

县市名	排序	县市名	排序	县市名	排序	县市名	排序
厦门市	1	沙县	18	尤溪	35	大田	52
福州市	2	罗源	19	云霄	36	永泰	53
石狮	3	南靖	20	莆田市	37	松溪	54
晋江	4	闽清	21	华安	38	连城	55
泉州市	5	邵武	22	建阳	39	古田	56
三明市	6	南安	23	建宁	40	平和	57
漳州市	7	闽侯	24	福安	41	上杭	58
龙岩市	8	连江	25	建瓯	42	屏南	59
永安	9	德化	26	清流	43	武平	60
福清	10	漳平	27	柘荣	44	寿宁	61
东山	11	将乐	28	诏安	45	宁化	62
长泰	12	永春	29	平潭	46	仙游	63
长乐	13	明溪	30	光泽	47	周宁	64
南平	14	宁德市	31	安溪	48	长汀	65
惠安	15	漳浦	32	霞浦	49	浦城	66
龙海	16	顺昌	33	永定	50	政和	67
泰宁	17	武夷山	34	福鼎	51		

建立了全体县市群的连接后,根据单向搜索,获得该尺度下的区域划分结果。在上一层尺度的区域划分基础上,用融合前各单元的属性值进行新一层尺度的单元融合。

3.2.4 聚类结果 图 1 为 4 个不同层次尺度的聚类结果:图 1(a)是初始状态,即福建省 67 个县市;图 1(b)是通过相邻吸引力的聚类结果;从图 1(c)开始可以发现区域聚类结构与实际地区级行政区有一定的重叠,为 I 级尺度聚类;图 1(d)整个福建省被划分 6 个类别,为 II 级尺度聚类。结果见表 6。

4 结论

传统空间单元的区域划分通常仅以属性数据作为划分依据,而对单元之间空间依赖关系考虑不周。在尺度空间理论基础上,利用多尺度空间聚类方法,不光考虑了空间单元的属性值之间的差异性,还考

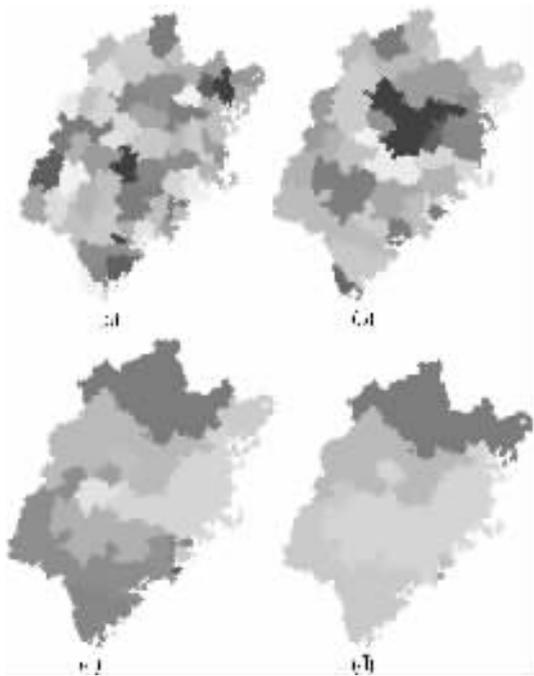


图 1 福建省区域经济多尺度空间聚类过程

表 6 福建省多尺度空间聚类结果

II 级尺度 聚类结果	中心单元	I 级尺度 聚类结果	中心单元	包含最小 单元数
中东部城市经济区	福州市	福州亚区	福州市	14
		晋江亚区	晋江市	2
		龙岩亚区	龙岩市	5
		泉州	泉州市	1
		永安	永安市	1
北部城市经济区	武夷山市	武夷山亚区	武夷山市	10
		福安亚区	福安市	4
中西部城市经济区	三明市	三明	三明市	1
		南平亚区	南平市	10
		沙县	沙县	1
惠安	惠安县	惠安	惠安县	1
南部城市经济区	厦门市	厦门亚区	厦门市	11
		东山亚区	东山县	5
		石狮	石狮市	1

参考文献:

[1] 邱凯昌. 空间数据挖掘与知识发现[M]. 武汉: 武汉测绘科技大学出版社, 2001.

虑了空间单元空间分布及其空间相互依赖性,使得在空间尺度由小变大过程中,具有高度空间相互依赖关系的空间单元相互融合,得到不同空间尺度下的聚类结果。该方法在区域空间聚类的应用,是以空间单元的相互依赖程度为基础,根据空间尺度的变化来寻找中心吸引点,然后对具有互相依赖关系的空间单元进行融合聚集,形成不同空间尺度下的空间区域的聚类结果。采用基本空间相关性和属性平均差异为基本模型,以福建省 15 年来各县市的人均 GDP 为指标,进行了社会经济发展水平空间聚类分析,结果与现实经济发展水平的区域划分基本吻合。

本文所采用的多尺度空间聚类方法,在应用上具有重要意义,但还存在一定的不足。为简化数据,只采用单一的 GDP 指标,要是能从多因素考虑,就可建立更为合理的空间相互作用模型,聚类结果可能更接近实际。

- [2] 刘晓,苏维词,邓吉祥. 重庆直辖 10 年区县经济差异及协调发展[J]. 重庆师范大学学报:自然科学版, 2009, 26(3): 35-39.
- [3] 任平,周介铭. 四川省经济发展水平区域差异综合评价研究[J]. 四川师范大学学报:自然科学版, 2007, 30(1): 102-105.
- [4] 郭仁忠. 空间分析[M]. 武汉: 武汉测绘科技大学出版社, 2000.
- [5] 陈小瑜,余明. 基于空间聚类分析的福建省各县市经济发展水平研究[J]. 热带地理, 2007, 27(4): 343-347.
- [6] Leung Y, Zhang J, Xu Z. Clustering by scale—space filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(12): 1396-1410.
- [7] 骆剑承,梁怡. 多尺度空间单元区域划分方法[J]. 地理学报, 2002, 57(2): 167-173.
- [8] 王法辉. 基于 GIS 的数量方法与应用[M]. 北京: 商务印书馆, 2009: 188-191.
- [9] 曹玮玮,孟凡星,白如山. 城市经济区与福建省区域经济的空间组织[J]. 福建师范大学学报:自然科学版, 2009, 25(3): 119-124.
- [10] 蒲俊. Maflab6.0 数学手册[M]. 上海: 浦东电子出版社, 2002.

(下转第 92 页)

A Study of Economic Regionalization Based on Multiscale Spatial Clustering

CHEN Xiao-yu

(Dept. of Resource and Environment , Quanzhou Normal College , Quanzhou Fujian 362000 , China)

Abstract : Automatic area division can be carried out by traditional cluster methods. But the attribute character was only considered in these methods , and the proximity between spatial objects was ignored. The method proposed using multi-scale spatial cluster was based on the theory of scale space. The proximity and attribute characters were combined in this method. So the great interdependent spatial objects were merged , and the division results with different spatial scale were obtained. The multiscale spatial clustering method was used. County—based statistical data of per capita GDP from 1990 to 2004 was adopted as a case. The experiment shows the result is the same as the actual condition. The proposed method considered both of the spatial proximity and attribute similarity. This approach has certain instructive significance to the sustainable region development of Fujian province.

Key words : multiscale ; spatial clustering ; Fujian province ; economic regionalization

(责任编辑 游中胜)