

最优分割法在空间分析中的应用

沈建法

(华东师范大学地理系)

空间分异规律(包括空间分类、区划)的研究是地理学的传统领域。国内外学者应用各种聚类分析方法进行这方面的研究已取得了许多成果,用单指标分组方法作空间分析也有先例。但大多数变量聚类方法难以求得最优解。本文用最优分割法对社会经济指标作单序列最优分组,分析中国社会、经济空间分异的基本特征。

本文主要数据来源是《中国经济统计年鉴1983》、《中国百科年鉴1983》。研究范围暂不包括台湾、香港、澳门等地区。

一、最优分割原理与程序框图

单序列最优分割问题的提法是:设有一有序序列 X_1, X_2, \dots, X_n ,要分成 K 类。记 \bar{X}_{ij}, D_{ij} 分别为 X_j, X_{i+1}, \dots, X_j 的均值和方差,即:

$$\bar{X}_{ij} = \frac{1}{j-i+1} \sum_{L=i}^j X_L;$$

$$D_{ij} = \sum_{L=i}^j (X_L - \bar{X}_{ij})^2;$$

$$1 \leq j \leq n \quad 1 \leq i \leq j$$

所谓最优分割就是要找出 $K-1$ 个分割点 t_2, \dots, t_k ,把序列分割成 K 段,使各段方差之和为最小,即:

$$\min D = \sum_{i=1}^K D_{t_i, t_{i+1}-1} \dots \dots \textcircled{1}$$

这里 $t_1 = 1, t_{K+1} - 1 = n; t_i < t_{i+1}, i = 2, 3, \dots, K$ 。

对于上述问题, Fisher 首先提出了一种类似于动态规划的递推算法。记 E_m, i 为前 m 个数据作最优 i 分割时的总方差, T_m, i 为其最后一个分割点(即第 $i-1$ 组的最后一个数据号),则有如下递推公式:

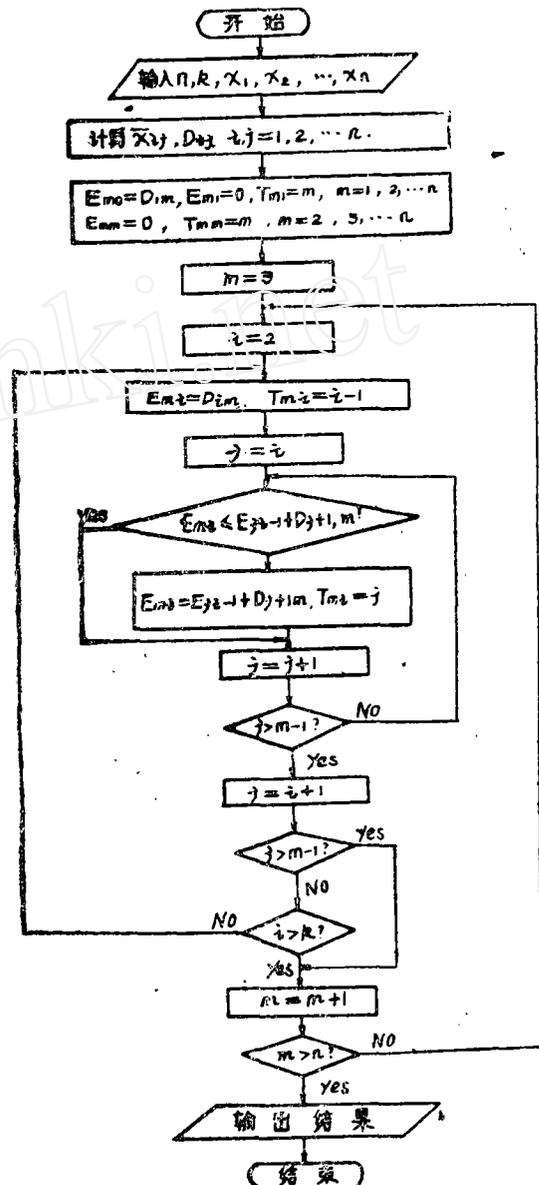


图1 最优分割法程序框图

$$\begin{cases} E_{m,i} = \min_{i-1 \leq j^* \leq m-1} \{ E_{j,i-1} + D_{j+1,m} \} \\ T_{m,i} = j^* & \begin{matrix} i = 2, 3, \dots, K^* \\ m = 3, 4, \dots, n \end{matrix} \end{cases} \dots\dots\dots ②$$

$$R \leq R_{\alpha}(n, k) = \frac{100}{1 + F_{\alpha}(k-1, n-k) \frac{n-1}{n-k}} \dots\dots\dots ③$$

初始条件如下:

$$\begin{cases} E_{m,0} = D_{1,m}, E_{m,1} = 0, T_{m,1} = m \\ m = 1, 2, \dots, n \\ E_{m,m} = 0, T_{m,m} = m - 1, m = 2, 3, \dots, n \end{cases}$$

图1为最优分割法计算机程序框图。计算机输出结果包括:

- ①不同最优分割的总方差分布图、表,据此可确定最佳分割段数和进行显著性检验。
- ②不同最优分割各组的数据个数、均值及组内方差。
- ③原样品(空间单元)在不同分割时的划分等级表。
- ④用网格法输出的空间等级分布图。

二、空间分异的显著性检验与空间凝聚指数

方差分析是检验组间差异显著性的常用方法。由此可导出空间分异显著性检验式如下:

这里R为组内方差占总方差的百分比; $R_{\alpha}(n, k)$ 为在 α 信度下n个地区分成k组时R值的显著性上限值; $F_{\alpha}(k-1, n-k)$ 为F分布置信限。

当③式成立时则空间分异显著,否则不显著。

空间分异显著只是表明各空间类型具有显著的差异性,它们在空间分布上凝聚抑或分散的程度则需用空间凝聚指数(Spatial Coherence Index)来衡量。我们提出下列公式来计算空间凝聚指数:

$$C = \sum_{i=1}^k \frac{n_i}{n} \cdot \frac{1}{k_i} = \frac{1}{n} \sum_{i=1}^k \frac{n_i}{k_i} \dots\dots\dots ④$$

这里n为总的单元数, n_i 为属于i类的单元数, K_i 为i类在空间上占据的块数。这一指标的基本出发点是:同类单元越是集中分布,空间凝聚度就越高,即凝聚度同同类单元分布块数成反比,各

表1 22个指标不同分割的组内方差占总方差的比例(%)

指标号	指标名	分割数R	1	2	3	4*	5 Δ	6 Δ
①	人口密度		100	26.1	7.4	3.3	1.7	1.0
②	城市化水平		100	25.5	8.8	4.4	2.0	0.4
③	单位面积工农业总产值		100	5.2	0.8	0.3	0.2	0.1
④	人均工农业总产值		100	21.8	6.5	2.0	1.0	0.7
⑤	工业产值比例		100	48.0	16.7	4.0	2.6	1.5
⑥	人均社会商品零售额		100	18.1	9.3	3.3	1.4	0.8
⑦	每万人科研人员数		100	19.7	10.4	2.9	1.7	0.7
⑧	人口自然增长率		100	47.5	22.2	13.2	5.0	3.2
⑨	工农业总产值增长速度		100	41.2	22.1	12.2	6.9	1.8
⑩	人均产粮		100	41.6	19.2	10.0	5.7	3.3
⑪	粮食亩产		100	21.4	10.7	5.5	3.7	2.4
⑫	农业中种植业产值比例		100	31.2	7.5	4.3	2.5	1.8
⑬	农业中林业产值比例		100	23.6	12.6	7.6	4.1	2.5
⑭	农业中畜牧业产值比例		100	21.7	9.2	4.7	2.5	1.4
⑮	人均产煤		100	20.7	8.3	3.4	1.5	0.6
⑯	人均发电量		100	30.1	11.4	3.6	2.3	1.6
⑰	工业中重工业产值比例		100	35.4	12.9	7.9	4.4	2.6
⑱	大型企业比例		100	32.8	11.5	6.3	3.7	2.0
⑲	小型企业比例		100	22.7	9.0	5.4	2.7	1.5
⑳	铁路密度		100	21.5	11.4	6.8	3.4	1.8
㉑	公路密度		100	40.9	18.3	10.4	4.4	2.8
㉒	单位工农业产值运输量		100	26.5	11.4	6.7	4.2	2.7
	平均		100	28.3	11.7	5.8	3.1	1.8

注: * 0.005信度下所有指标显著
 Δ 0.001信度下所有指标显著

表 2 最优五分割结果

等级	指标号	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮	⑯	⑰	⑱	⑳	㉑	㉒	
地区																							
北京	2	1	3	2	1	1	1	2	3	5	2	5	4	3	3	2	3	1	5	1	1	4	
天津	2	1	2	2	1	2	2	2	3	5	4	5	5	5	5	2	4	2	4	2	2	5	
河北	4	5	5	5	3	5	5	3	2	3	3	2	4	3	4	4	1	3	4	2	4	3	
山西	4	3	5	4	2	4	4	3	3	4	4	4	1	5	1	3	1	4	2	4	3	1	
内蒙古	5	3	5	5	3	4	4	3	3	4	5	4	1	3	3	4	2	4	1	5	5	2	
辽宁	4	2	4	3	1	3	3	3	3	3	1	3	4	4	3	3	2	2	3	3	3	3	
吉林	5	2	5	4	2	3	3	4	1	1	2	1	3	5	3	3	2	4	2	3	4	3	
黑龙江	5	2	5	4	2	3	3	4	3	2	4	1	2	5	3	3	1	3	2	4	4	3	
上海	1	1	1	1	1	1	2	3	4	5	1	5	5	4	5	1	4	2	4	2	2	5	
浙江	2	4	4	3	2	4	5	4	2	2	1	3	5	5	5	4	4	5	1	4	3	5	
安徽	3	4	4	4	2	4	5	3	3	1	4	4	4	4	5	4	5	5	1	4	3	5	
福建	4	4	5	5	3	4	5	2	2	4	2	4	1	5	5	5	5	5	1	4	2	5	
江西	4	4	5	5	4	5	5	3	4	2	2	3	2	5	4	5	3	5	1	4	3	4	
山东	3	3	4	4	3	4	5	3	2	2	2	4	5	4	4	4	4	4	2	4	4	5	
河南	3	5	5	4	5	5	3	2	3	3	1	4	5	3	5	3	5	4	1	4	3	4	
湖北	4	3	5	4	2	4	4	4	3	3	2	2	3	5	5	4	3	4	1	4	3	5	
湖南	4	4	5	5	4	5	5	3	4	2	1	2	3	4	4	5	3	4	1	4	2	4	
广东	4	4	5	5	2	3	5	2	3	4	2	4	1	4	5	5	5	4	1	5	2	5	
广西	4	4	5	5	4	5	5	2	4	3	2	3	2	4	5	5	4	4	1	5	4	5	
四川	4	4	5	5	4	5	5	3	2	5	3	3	2	4	4	5	2	3	2	4	4	4	
云南	5	5	5	5	4	5	5	3	4	3	3	3	1	3	4	5	3	4	2	5	4	4	
西藏	4	4	5	5	4	5	5	3	3	4	3	3	1	3	4	5	3	4	2	5	4	4	
宁夏	4	4	5	5	2	5	4	4	3	4	4	2	2	5	4	4	3	2	3	4	3	4	
青海	5	4	5	5	3	4	3	5	4	5	4	4	4	4	2	4	5	2	3	5	5	4	
海南	5	3	5	5	3	3	4	2	2	3	4	1	2	5	2	3	1	3	3	5	4	1	
新疆	5	3	5	5	4	3	3	5	2	4	4	1	3	4	3	5	3	5	2	5	5	4	

表 3 最优五分割各组(等级)均值和空间凝聚指数

指标名	1	2	3	4	5	单位	空间凝聚指数
人口密度	1926	6.7	425	215	46	人/km ²	0.948
城市化水平	65	41	29	20	14	%	0.471
单位面积工农业总产值	116.0	22.3	16.5	5.2	1.1	万元/km ²	0.931
人均工农业总产值	6.0	3.1	1.5	1.0	0.6	千元/人	0.800
工业产值比例	89.5	68.7	61.0	55.3	18.5	%	0.382
人均社会商品零售额	824	546	321	231	160	元/人	0.490
每万人科研人员数	387	238	109	80	50	人/万元	0.914
人口自然增长率	21.2	14.2	10.2	6.8	0.8	%	0.793
工农业总产值增长速度	18.4	12.7	10.1	5.3	-5.4	%	0.423
人均产粮	1.3	0.9	0.8	0.6	0.4	公斤/人	0.339
粮食亩产	618	514	400	299	234	斤/亩	0.333
农业中种植业产值比例	72	66	59	52	31	%	0.351
农业中林业产值比例	8.2	6.1	4.2	2.3	0.5	%	0.411
农业中畜牧业产值比例	55	40	22	17	12	%	0.434
人均产煤	6.2	2.5	1.1	0.5	0.1	吨/人	0.629
人均发电量	1884	979	537	319	211	度/人	0.833
工业中重工业产值比例	75.5	60.4	52.1	43.9	36.7	%	0.529
大型企业比例	2.5	1.2	0.7	0.4	0.1	%	0.454
小型企业比例	99.0	98.4	97.2	94.6	92.7	%	0.440
铁路密度	512	373	214	99	35	km/万km ²	0.552
公路密度	4797	3033	2057	1276	314	km/万km ²	0.724
单位工农业产值运输量	57	36	23	13	6	吨/万元	0.586

类凝聚度的加权平均就得空间凝聚指数。这一指数在区间 [0, 1] 间取值, 当所有同类单元在空间上连续分布时, 空间凝聚指数为 1。

对 1983 年, 我国大陆 29 个省、市、自治区 22 个社会经济指标的空间序列分别进行最优分割。表 1 为这 22 个指标不同分割的组内方差占总方差的比例。经检验, 在信度为 0.005 时, 大多数指标的最优三分割显著, 所有指标的最优四分割、五分割、六分割显著, 当信度为 0.001 时, 最优四分割仅人口自然增长率、工农业总产值增长速度二个指标不显著, 最优五分割、六分割全部显著。从平均来说, 最优三分割损失信息 11.7%, 最优四分割损失信息 5.3%, 而最优五分割仅损失信息 3.1%。表 2 为 22 个指标的最优五分割结果, 表明 29 个省、市、自治区划归的不同等级。表 3 为 22 个指标最优五分割的各组均值。

根据最优五分割结果作出的等级分布图, 用式④计算了空间凝聚指数, 也列在表 3 中。可见有些指标如人口密度、人均工农业总产值、单位面积工

农业总产值、每万人科研人员数、人口自然增长率、人均产煤量、公路密度等空间凝聚指数较大, 一般在 0.6 以上, 说明这些指标的空间分异规律性较强 (参见图 2、3、4)。另一些指标如人均产粮、粮食亩产、农业中种植业产值比例空间凝聚指数较低, 在 0.4 以下, 表现为不同等级的地域单元互相混杂, 空间分异一致性较差。



图 4 人均工农业产值分布图

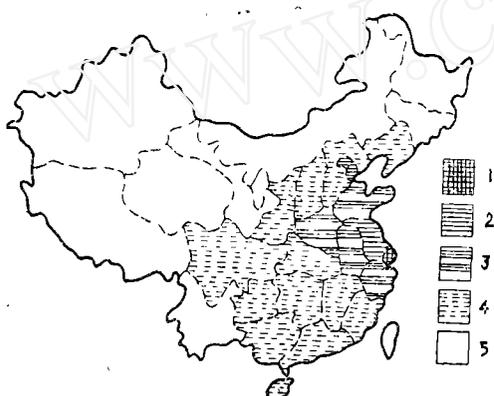


图 2 人口密度等级分布图

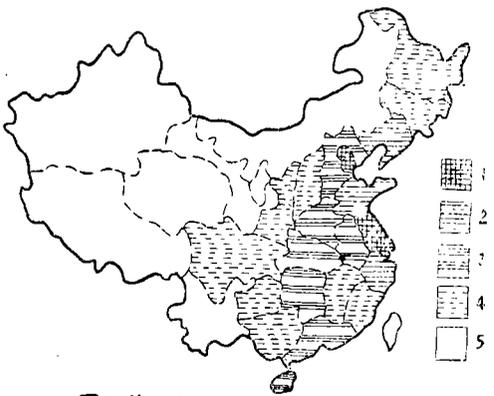


图 3 单位面积工农业产值分布图

三、中国经济空间分异模式与经济地带

根据我国 22 个社会经济指标最优五分割结果作出的等级分布图, 大致可归纳出如下三种空间分异模式:

模式一: 东—西分布型。这一分布型最为普遍, 共有 8 个指标: 人口密度 (图 2)、单位面积工农业总产值 (图 3)、人均工农业总产值 (图 4)、铁路密度、公路密度、人均产粮、粮食亩产、农业中畜牧业产值比例等。形成这一分布型的主要原因可归结为自然地理环境, 社会经济发展过程中的差异性。

模式二: 南—北型。具有这一分布型的有 5 个指标: 城市化水平 (图 5)、工业产值占工农业总产值比例、人均产煤、单位工农业总产值运输量。北部地区资源较丰富, 东北、华北为经济发达地区, 其西部则为人口稀疏、农业不发达、解放以来工业有较大发展的地区, 且北部重工业比重较大, 因此形成了这种北高南低的分布模式。

* 这两个指标最优五分割结果表明大多数地区均划归第五等级, 故采用最优八分割结果, 图 3、图 4 中原一、二、三、四等已划归一等。图 3、图 4 的空间凝聚指数分别为 0.655、0.661。

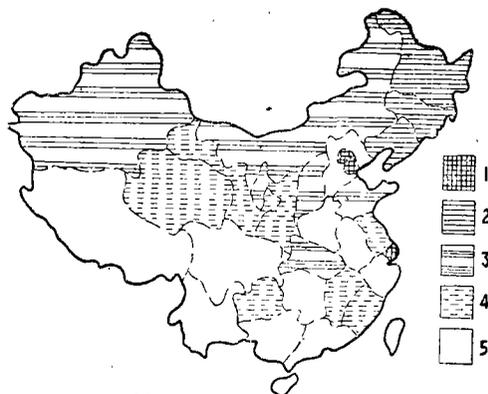


图5 城市化水平分布图

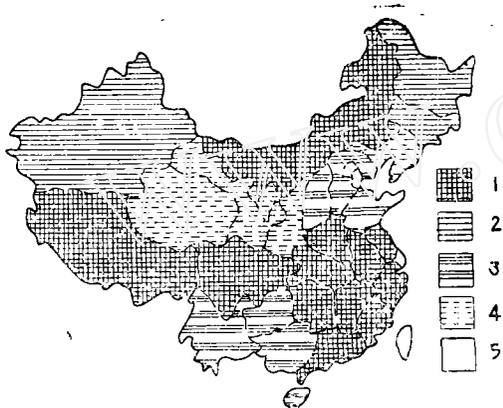


图6 小型企业比例分布

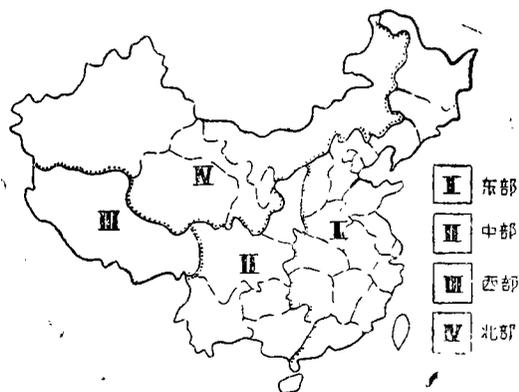


图7 经济地带的划分

模式三：扇型。这一分布型的基本特点是等级较高或较低的单元位于中西部甘肃、陕西一带，其它等级的单元则大致呈扇形分布于四周，而且空间凝聚指数较小，均在0.53以下。具有这一分布型的指标有人均社会商品零售额、工业中重工业产值比例、大型企业比例、小型企业比例（图6）等4个指标。由图6可见，东南部经济发达地区及西藏、内蒙古等不发达地区小型企业比例高，青海、陕西、甘肃、宁夏等内地不发达地区及工业高度发达的辽宁、北京、天津、上海小型企业比例低。

在上述中国经济空间分异特征分析的基础上，以人均工农业总产值、单位面积工农业总产值为依据，结合考虑人口分布状况，初步划分出如下四个经济地带（见图7）。

东部经济发达地带：包括黑龙江、吉林、辽宁、北京、天津、河北、山西、山东、河南、安徽、江苏、湖南、湖北、江西、浙江、上海、福建、广东等18个省市。这一地带人口稠密、人均工农业产值、单位面积工农业产值均为最大。其中又以东部沿海的辽、京、津、鲁、苏、沪、浙最为发达，豫、皖、湘、赣、闽相对较落后。

中部经济欠发达地带：包括陕西、四川、贵州、广西、云南5个省区。这一地带人口密度、单位面积工农业总产值较高，人均工农业总产值较低，为人口密度高的经济不发达地区。又以桂、黔、滇相对较落后。

北部经济不发达地带：包括内蒙古、宁夏、甘肃、青海、新疆5个省区。人口密度、单位面积工农业总产值、人均工农业总产值均较低。为人口密度低的经济不发达地区。

西藏经济最不发达地带：仅包括西藏一个区。人口最为稀疏、经济最不发达。

表4为这四个经济地带的基本情况。

（下转154页）

才,最重要的是到社会中进行调查研究,了解社会当前与今后发展的需要,特别是四化建设和科技发展对培养人才的需要。由于四化的发展和经济与社会改革,客观要求与人们的认识也在不断变化,故调查需要逐步深入,不断提高,必须反复进行。随着问题的深入,调查的内容重点也不断的改变和深入。分校地理专业方向的确定就是经过几次反复深入调查的结果,目前专业虽然定下来,但仍存在许多问题需要进一步明确落实,如培养方向的具体落实,选修组方向的确定,如何把在校的教学与将来的工作更紧密联系起来,即学习与生产科研实践更紧密结合等问题都需要进一步调查解决。

(2) 应克服学科之间的偏见和旧的传统教育

的束缚。我们这些从事这项工作的人都比较熟悉过去情况,习惯于过去一套,传统的教育思想影响较深,专业知识过于狭窄,从思想到专业知识都存在不适应形势发展的状况。面临着思想上和知识结构上转向问题,需要有一个重新学习,重新认识的过程。

(3) 方向定了后,就需要立即组织力量予以实现,并且在实践中逐步充实与提高。教育改革象其他改革一样是个非常复杂的问题,没有现成的经验可循,也不可能一改即成,需要不断摸索前进。所以只要方向看准了,基本条件具备了,就要大胆地上,克服困难创造条件先把专业办起来,再逐步解决存在的问题,不断地加以完善。

全国有20个城市认真抓好改革 去年工业总产值均超过百亿元

新华社北京3月6日电 据新华社驻各地记者的调查,1986年全国有20个城市的工业总产值超过100亿元。这20个城市是:上海、北京、天津、苏州、无锡、沈阳、广州、武汉、南京、重庆、成都、大连、青岛、杭州、宁波、常州、扬州、佛山、南通、烟台。其中以上海市为最多,工业总产值达到900亿元。

据计算,这20个城市的工业总产值合计约4,000亿元,约占全国工业生产总产值的1/3以上。

认真抓好改革是这些城市工业生产取得进展的重要原因。这些城市中约有40%的全民所有制企业推行了厂长负责制;以这些城市的企业为骨干,以产品为“龙头”的企业联合体已遍及全国各地。

(摘自湖南日报1987年3月10日)

(上接159页)

表4 经济地带概况

项目 地带	面积 (%)	人口 (%)	人口密度 (人/km ²)	工农业总产值 (%)	单位面积产值 (万元/mk ²)	人均产值 (元/人)
东部	29.2	71.5	282	81.9	26.52	1011
中部	16.4	22.4	146	13.7	7.88	540
北部	41.8	5.9	15	4.4	0.98	651
西藏	12.6	1.9	2	0.8	0.036	364
全国	100	100	107	100	9.44	883

四、结语

本文是用最优分割法进行空间分析的一个尝试,初步分析了中国经济空间分异的基本趋势,归纳出三个空间分异模式,划分出四个经济地带。这

些结果将有助于认识中国经济空间差异性 & 大经济区的划分。但这些结果仅是初步的,有待于进一步分析研究,阐明空间模式与经济地带的成因及其在我国经济发展中的影响与意义。