信阳师范学院学报(自然科学版) Journal of Xinyang Teachers College
第 10 卷 第 1 期 1997 年 1 月 (Natural Science Edition) Vol. 10 No. 1 Jan. 1997

城镇体系空间结构的信息维分析

陈彦光 罗 静
(信阳师范学院地理系. 河南 ,信阳 464000) (信阳师范学院经济系. 河南 ,信阳 464000)

摘 要 本文提出城镇体系空间结构的信息维测算及其分析方法。以河南省北部为研究区 ,测算出该区域城镇体系空间分布的信息维数 ,然后结合多分维进行数值分析 ,揭示系统的结构特征并指出问题的实质所在 ,进而提出优化的设想

关键词 城镇体系 ;空间结构 ;分形 ;分维 ;信息熵 ;信息维 ;多分维 ;豫北地区
分类号 K928.5

地理学的明确主题是认识区域差异。有差异就有信息 ,信息是差异的再现 ,因此 ,区域差异必然导致地理空间信息流的产生 ,以区域性为其特征的地理学从而可以被认为是一门空间信息科学[1]。城市地理学也不例外 ,城镇体系的研究过程实则是地理信息的分析过程 ,为此需要知道系统的信息熵。但熵值的测算并非易事 ,有时根本难以测量。由于地物分布的无标度性(non - scaling) ,测出的信息量常常是一个变化的数值 ,不能作为系统分析的mb有效判据。幸运的是 ,现已证明申农(Shannon) 熵与豪斯道夫 (Hausdorff) 维数等价[2] ,于是可以运用分形几何学的方法 ,以分维(fractal dimension) 代替信息熵 ,使得信息分析与分形研究接口。对于城镇体系来说 ,可以借助尺度变换 ,将信息熵直接转化为信息维。本文试以河南省北部地区 (简称“豫北地区”) 为研究范围 ,提出城镇体系空间结构的信息维测算及其分析范例。

1 城镇体系的信息维模型

1.1 申农信息熵与信息量

首先给出申农熵公式。对于具有 N 种可能的独立结果的随机试验 $X = \{ X_i , i = 1 ,2 , \cdots, N\}$,它在试验前所包含的平均不确定性数量可表示为

$$H = -\sum_{i=1}^{N} P_i \log P_i \tag{1}$$

式中 H 为信息熵, P_i 为 X_i 出现的概率且满足

$$0 \le P_i \le 1 , (i = 1 ,2 , \cdots, N) ; \sum_{i=1}^{N} P_i = 1 \tag{2}$$

这样两个条件。假定试验后 ,结果是 X_{io} 出现 ,即后验概率分布为

$$P_i = \begin{cases} 1 , & i = i_0 \\ 0 , & i \ne i_0 , \end{cases} \quad (i = 1 ,2 , \cdots, N) \tag{3}$$

此时不存在后验不确定性 ,熵值为零 ,故该试验所提供的信息量 I 为

$$I = H - 0 = -\sum_{i=1}^{N} P_i \log P_i \tag{4}$$

可以看出 , I 与 H 数值相等 ,却是两个不同的的概念 : H 表示不确定性 , I 则表示消除这个不确定性所需要的信息量。

1.2 区域城镇分布的信息维

可以运用网格化方法求得地物分布的信息量 ,但绝大多数地理现象 (包括城镇) 的空间分布具有无标度的分形性质 ,故改变网格的尺寸 ,测得的信息量是一个变化的值 ,此时可以将其转化为信息维[3]。

取一个地理区域 (方形或矩形) [4] ,将其划分为相同大小的若干子区域(图 1) ,每个子

收稿日期 1996 - 06 - 18

区域即网格的尺寸为 r,考察各个子区域中出现城镇的概率 P_i,由此可以计算出整个区域中城镇分布的信息量:

$$I(r) = -\sum_{i=1}^{N(r)} P_i \ln P_i \tag{5}$$

式中 $N(r)$ 为非空小区域的数目, $I(r)$ 为信息量。改变子区域的尺寸 r,可得一系列的 $I(r)$ 值。如果城镇体系的空间结构具有分形性质,则满足

$$I(r) \propto r^{-D_1} \tag{6}$$

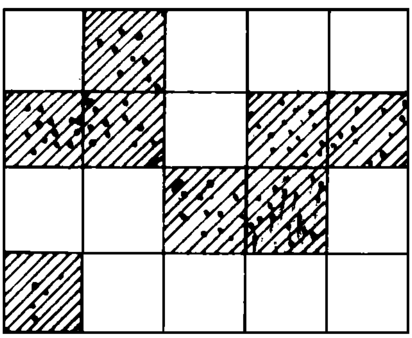

图 1 区域分形体的网格覆盖

定义的负幂律,式中 D_1 为信息维。

实际上,存在一个一般的分维(即广义维)公式

$$D_q = \frac{1}{q-1}\lim_{r\to 0}\frac{\ln \sum_{i=1}^{N(r)} P_i^q}{\ln r} \tag{7}$$

这里 q 为参数($q = -\infty, \cdots, -2, -1, 0, 1, 2, \cdots, \infty$)。当 $q = 0$ 时, D_0 为容量维即豪斯道夫维;当 $q = 1$ 时, D_1 为信息维;当 $q = 2$ 时, D_2 为关联维,且有 $D_0 > D_1 > D_2$。改变参数 q,可以定义无穷多种维数, D_q 可称之为第 q 阶信息维。通常所指的信息维是 D_1,即式(5)、(6)式定义的分维。观察可知,式(7)中隐含着瑞利(Renyi)熵公式

$$I_q(r) = \frac{1}{1-q}\ln \sum_i P_i^q \tag{8}$$

根据 $P_i^q = \exp(q \ln p_i)$,考虑 $q \to 1$ 时的极限可得

$$\lim_{q\to 1}\frac{1}{1-q}\ln \sum_i P_i^q = -\sum_i P_i \ln P_i \tag{9}$$

于是得出狭义的信息维公式

$$D_1 = \lim_{r\to 0}\frac{\sum_{i=1}^{N(r)} P_i \ln P_i}{\ln r} = -\lim_{r\to 0}\frac{I(r)}{\ln r} \tag{10}$$

信息维数的大小,可以反映区域城镇空间分布的聚散程度[3]。当所有的城镇聚集到一点时,系统为零维;当城镇均匀分布于地表时,系统为 2 维;通常情况下系统的维数介于 02 维之间;要素分布愈均衡,系统的信息维数愈高(即愈趋近于 2 维),这正好与系统的信息熵对应,要素愈均匀,信息熵值愈高。用分数维代替信息熵作地理系统的信息分析,其理论依据之一便在于此。

由于瘦分形(Thin Fractal)的豪斯道夫维数大于拓扑维数而小于所在空间的维数,而城镇体系分布的地理空间可视为 2 维,考虑到信息维 D_1 与容量维 D_0 的关系: $D_0 > D_1$,则应有 $1 < D_1 < 2$(体系要素数目必大于 1,故不可能为零维)。下面以豫北地区为例说明城镇体系信息维的测算和分析方法。

2　城镇体系空间结构的信息维测算

2.1　研究区范围的确定

本文将豫北地区定义为:以河南省的几何中心许昌市所在的 34 °N 线为界的河南省北部地区。基于分形体的自相似性质(Self-similarity),研究区的确切界线并不重要(不影响总体分析效果),但也并非随意可为:首先要明确研究客体的轴心位置所在(本文以郑、汴、洛为轴带)。更重要的是地球表面曲率的影响:研究区太大,则受地图投影变形的影响,分维测算不准;太小则样本不足,不能反映系统的特征。为测算信息维,现确定一个矩形区域,其范围大致是:33. 91 °N36. 11 °N,111. 07 °E115. 69 °E。区内含县级以上城镇总数为 $N = 92$(包括山东、河北、山西等省的若干城镇,它们不妨碍本区域的城镇分形分析)。

2.2 信息维数的测算

现设研究区的边长为 1 单位(长边和宽边各取不同的单位),分别将各边划分成 k 等分,则研究区被分成 k^2 个小区域,这时有

$$r = 1/k \tag{11}$$

r 为小区域的尺寸,统计落入每个小矩形区中的城镇数目 N_{ij},令

$$P_{ij} = N_{ij}/N, (i, j = 1, 2, \cdots, N) \tag{12}$$

这里 P_{ij} 可近似地视为城镇落入某个小矩形区的概率,于是得整个区域的信息量为

$$I_{(r)} = \sum_{i=1}^{k}\sum_{j=1}^{k} P_{ij}\ln(1/P_{ij}) \tag{13}$$

式中的双求和只对 $P_{ij} \neq 0$ 的小矩形区进行。改变 K 值、从而改变 r,得不同的信息量值 $I(r)$(表 1),代入信息维公式(10)得

$$D_1 = \lim_{k\to\infty} \frac{\sum_{i}^{k}\sum_{j}^{k} P_{ij}\ln(1/P_{ij})}{\ln K} \tag{14}$$

由此可以计算信息维数。

表 1 对研究区网格化所得的统计数据

K	$N(r)$	$P_{ij} = N_{ij}/N$ ($N = 92$)	$I(r)$
2	4	21/92, 22/92, 23/92, 26/92	1.383
3	9	5/92, 8/92, 10/92, 3×11/92, 12/92, 15/92	2.163
4	16	2/92, 2×4/92, 5×5/92, 3×6/92, 2×7/92, 2×8/92, 9/92	2.726
5	24	3×2/92, 8×3/92, 6×4/92, 4×5/92, 3×6/92	3.128
6	34	6×1/92, 9×2/92, 13×3/92, 3×4/92, 4×5/92	3.426
7	45	16×1/92, 16×2/92, 9×3/92, 3×4/92, 5/92	3.690
8	54	25×1/92, 21×2/92, 7×3/92, 4/92	3.894
9	63	39×1/92, 19×2/92, 5×3/92	4.056
10	68	48×1/92, 16×2/92, 4×3/92	4.137
11	73	58×1/92, 11×2/92, 4×3/92	4.213
…	…	… … … …	…

注:本表据《中国地图册》(中国地图出版社,1992 年第 8 版)测算,地图比例尺为 1∶3500000.

实际上,测算分维应结合 $lg - lg$ 坐标图进行。依次取 $K = 2, 3, \cdots, 16, \cdots$,得一系列的点对 $(I(r), \ln(1/r))$,然后以 $I(r)$ 为纵坐标,以 $\ln(1/r)$ 为横坐标,画出双对数坐标图(图 2)。在此图中,典型的信息维曲线由具有统计意义的三段构成:

其中中间直线部分的斜率即为所求的信息维 D_1 值,相应的范围为无标度区(scaling range),两个拐点 E、F 所对应的标度值 γ_E、γ_F 分别为无标度区的上、下限。计算结果为 $D_1 = 1.648$。

显然,信息维数在此给出的城镇体系的信维息量非常有限,仅此很难说明什么实质性问题,为此利用表 1 进而推求多分形(multifractals)维数。多分维 $f(a)$ 谱与广义维等价,

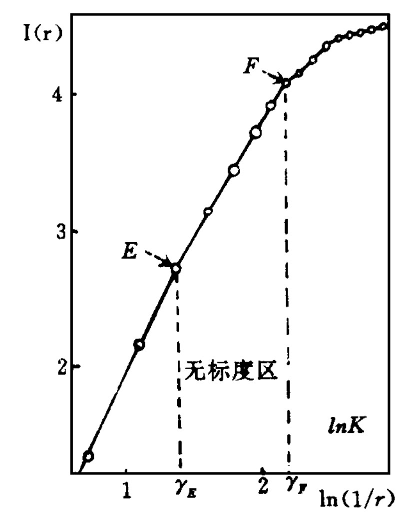

图 2 豫北地区城镇分布的信息维

但物理意义更明显。定义当 $r \to 0$ 时第 i 个格子即小区域,见图 1)所含的几率 P_i 标度为

$$P_i r^a \qquad (15)$$

这里 a 为密集指数,且

$$N_{(a)} \mathrm{d} a r^{-f(a)} \mathrm{d} a \qquad (16)$$

$N_{(a)} \mathrm{d} a$ 是落在 $a \to a + \mathrm{d}a$ 内的格子数,$f(a)$ 为 a 在 $a \to a + \mathrm{d} a$ 内的分形子集的分集。由 D_q 与 $f(a)$ 的勒让得(Legendre)变换

$$(q-1) D_q = qa(q) - f[a(q)] \qquad (17)$$

$$a(q) = \frac{d}{dq}(q-1) D_q \qquad (18)$$

可得

$$f\,[a(0)] = D_0,\ a(1) = f[a(1)],$$

$$a_{min} = D_{+\infty},\ a_{max} = D_{-\infty}$$

引用 μ——权重[5],由几(概)率分布可直接计算 $f(a)$ 谱。μ—权重公式为

$$\mu_i = P_i^{\,q} / \sum_i P_i^{\,q} \qquad (19)$$

于是得

$$a(q) = \lim_{r \to 0} \frac{1}{\ln \gamma_i} \sum \mu_i \ln P_i \qquad (20)$$

$$f[a(q)] = \lim_{r \to 0} \frac{1}{\ln \gamma_i} \sum \mu_i \ln \mu_i \qquad (21)$$

对于简单的几何分形,只有

$$D_q = f[a(0)] = D_0 \qquad (-\infty < q < \infty)$$

即分维恰为豪斯道夫维,而对于城镇体系的多重分形,可得非平凡的结果。

可以看出,当 $q = 0$ 时,$D_0 = f[a(0)]$为容量维;当 $q = 1$ 时,$D_1 = f[a(1)] = a(1)$为信息维(此时 $\mu_i = P_i$)。现进一步令 q 为各种实数,对 $f[a(q)]$取值,发现当 q 取负值时,计算结果出现紊乱,即当 $q \to -\infty$时,所得的数值异常(表 2)。

表 2　多分维及异常的 $f(a)$ 谱

q	…	-5	…	-2	-1	0	1	2	…	5	…
$a(a)$	…	1.061	…	1.671	1.793	1.712	1.648	1.503	…	1.270	…
$f(a)$	…	4.081	…	1.918	1.716	1.715	1.648	1.432	…	0.685	…
D_q	…	1.564	…	1.753	1.755	1.715	1.648	1.574	…	1.416	…

3　城镇体系空间结构的分数维分析

3.1 信息维数分析

由上节的测算结果,首先可以看出,信息维的无标度区比较狭窄(图 2),原因可能是:其一,研究区的范围较小;其二,城镇的截取下限太高。总之是样本不足。但是,如果扩大研究区的范围,将会受到地图投影变形的较大影响,从而测算误差较大;如果降低城镇体系要素的下限(如下降至乡镇级),又难以保证测算时城镇的定位精度。更重要的是,研究区的总体范围和城镇规模的截取下限均与研究主题有关。一般不能随意扩大。无标度区狭窄的更大原因应该是豫北城镇体系的空间结构不佳,这是我们所要分析的重点。

信息维可以反映整个豫北地区城镇分布的空间特征。通过前面的测算工作可知:第一,在双对数坐标图上(图 2),存在明显的无标度区,表明整个豫北城镇的空间分布是自相似的,但无标度区比较狭窄,暗示着城镇体系的空间演化受到无序因素的较大干扰;第二,信息维值为 1.65 左右,说明城镇分布的均衡程度适中,而且暗示着影响城镇体系空间局势的控制变量不是很多;第三,豫北地区的城镇分布具有多重分形结构,但当 $q < 0$ 时,$f(a)$等主要参量不再收敛(表 2),这里可能有一个临界值 q_c,在 $q_c = -1$ 附近出现了标度间断,即城镇出现的概率 P_i 与线度 r 之间不再存在标度关系,可见研究区城镇的多分形结构只存在于 $q > -1$ 的阶数。换言之,在城镇分布密集(因而测度概率也集中)的地域分形结构较好,而在城镇分布分散的地域则形态比较单调。

3.2 子系统维数分析

为进一步查明问题所在,现对研究区内各城镇子系统进行考察,测算局部地域的分维。将研究范围限定在陇海线的郑州(郑)、开封(汴)、洛阳(洛)一带,分别以郑、汴、洛为中心城市,确定三个城镇子体系 $U_i(i=1,2,3)$,不必满足 $U_i \cap U_j = \varnothing(i \neq j, i=1,2,3)$,也就是说,三个子系统交集未必为空,允许要素重叠,这不影响分析结论,然后用下式求出各子系统城镇空间分布的豪斯道夫维数[6]

$$R_s = KS^{1/D} \tag{22}$$

式中 D 为分维,K 为比例系数,S 为所考虑的城镇个数,R_s 为城镇体系空间分布的平均半径,且 R_s 被定义为

$$R_s \equiv \left\langle \left(\frac{1}{S}\sum_{i=1}^{s} r_i^2\right)^{\frac{1}{2}} \right\rangle \tag{23}$$

式中 r_i 为某城镇(第 i 号要素)到中心城市(或首位城市)的直线距离(称重心距),$i=1, 2, \cdots, n$, n 为体系内要素总数,$s \leq n$,$\langle\ \rangle$ 表示平均。结合双对数坐标图(图 3),算出各系统的分维数(表 3)。

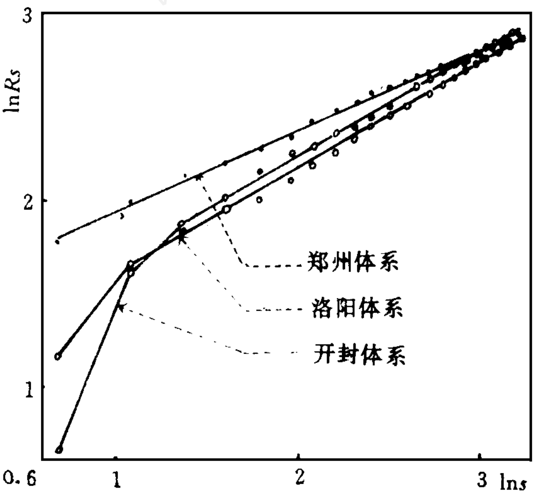

表 3 郑、汴、洛三体系空间结构的分维值

分维		郑州体系	开封体系	洛阳体系
计算值(平均)	D	2.387	1.385	1.570
	r	0.998	0.996	0.960
量测值(局部)		2.341	1.705	1.704

注:(1) 本表所依据的地图册同表 1;(2) 计算值为平均数;(3) 量测值取成直线的无标度区段。

从图 3 及表 3 可以发现,郑、汴、洛三个体系的空间分布分形结构均存在一些问题:其一,郑州体系的点子在 log - log 坐标图中分布较好,但其维数过高(>2),超过嵌入空间的维数,故不符合标准瘦分形定义,可视为具有潜在的分形结构;其二,开封体系和洛阳体系点子在 log - log 坐标图中分布不佳,暗示着城镇体系的空间发育不够协调,存在无序因素的干扰;其三,从测量值看,开封、洛阳体系的豪斯道夫维数(局部分维)与表 2 给出的容量维大约一致,但郑州体系的维数与它们不相匹配,同时也难以被环境系统所包容[3].

各子系统的分维对比分析可以解释豫北地区多分形的标度间断:其原因可能是系统发育的自相似演化不够和谐及其与环境不太协调,以致分形性质发生某种程度的退化。豫北地区范围较广,各城镇系统(如郑、汴、洛体系)的发展演化过去各自独立,缺乏综合协调机制,没有形成一个有机整体,于是无序因素的干扰导致了多分维谱线的"断裂"。

3.3 系统优化建议

综上所述,可得如下结论:豫北地区的城镇体系业已发育了自相似的组织形态,且出现了多重分形构造,这是一种好的趋势,因为分形是大自然的优化结构,分形体能够有效地占据(地理)空间[3]。但是这种分形结构很不完善,需要利用综合协调技术予以改造,主要措施是:第一,科学地设置新城镇,降低郑州体系的分维,使之与开封体系和洛阳体系的分维匹配,并能被地理环境所包容——即城镇体系的分维要小于其支持系统(地理环境)的维数。第二,求出豫北地区城镇规模分布的最佳哲夫(Zipf)维数(q),使得区域城镇规模分布趋于合理从而形成满意的等级结构。因为等级结构影响系统要素的相关作用从而影响其空间分布。第三,利用经济杠杆和立法手段,引导城镇进行劳动地域分工[7],设法将资源投入边际收入高的城镇,达到系统要素边际收益的空间均衡[8],使得城镇之间形成和谐的相互关系,以便空间结构健康发展。

顺便说明的是:具体操作需要迭代函数系统(IFS)和地理信息系统(GIS)等技术手段的支持,有关细节本文不拟详述。

4　结语

本文旨在提供一种研究城镇体系空间结构的方法及其基本分析思路,故未对豫北地区城镇体系展开深入的探讨。但由前述分析可知,运用分形理论考察城镇系统结构是行之有效的。此方法从城镇体系空间结构的信息维测算开始,然后辅之以多分维以及子系统的(局部)豪斯道夫维数,逐步剖析,可以揭示藉其他理论方法无法揭示的问题,并能暗示系统优化的改良方案。因此,信息维分析将必成为城镇体系信息熵分析的替代工具。

*致谢:东北师范大学地理系丁四保教授为本项研究提供了理论指导和经费支持,在此谨表谢意。

参　考　文　献

1　陈涛. 地理学是一门空间信息科学. 信阳师范学院学报,1994,7(2):217220

2　Ya B Ryabko. Problemy Pevedaci Informatsii, 1986, 20(3):1626

3　陈涛. 豫北地区城镇体系的分形研究. 硕士学位论文,东北师范大学地理系,1995

4　高安秀树,沈步明等译. 分数维. 北京:地震出版社,1989,1518

5　Chhabra A, Jensen R V. *Direct determination of the f(a) singularity spectrum*. Phys Rev lett, 1989,62:13271330

6　陈涛. 城镇体系随机聚集的分形研究. 科技通报,1995,11(2):98101

7　陈才等. 区域经济地理学原理. 北京:中国科学技术出版社,1991,5981

8　陈涛. 区域资源空间分配的数学原理. 信阳师范学院学报,1995,8(2):202207

Analysis of Information Dimension on the Spatial Structure of Urban Systems

Chen Yanguang

(Dept of Geog, Xinyang Teachers College, Xinyang, Henan; China 464000)

Luo Jing

(Dept of Econ, Xinyang Teachers College, Xinyang, Henan; China 464000)

Abstract　A method of using information dimensions to analyse the spatial structure of systems of towns is put forward in this paper. The urban system in North Henan Province is taken as an example to illustrate how to measure the information dimensions and the $f(a)$ singularity spectrum of multifractals. An analysis of numerical values is made to bring to light the problems existing in the system and some preliminary plans of improving the system structure are presented to show the train of thought of applying fractal geometry to the problems of urban geography.

Key words　Urban system; Spatial structure; Fractals; Fractal dimension; Information entropy; Information dimension; Multifractals; North Henan Province

作者简介　陈彦光,原名陈涛,男,1965 年生。从事"地理分形与地理信息分析"研究,代表性论文有《城镇体系随机聚集的分形研究》等。

责任编辑　郭红建