分形维数计算方法的研究

聂笃宪, 曾文曲, 文有为
(广东 工业大学 计算机学院, 广东 广州 510090)

摘 要: 分形维数作为科学研究的重要工具之一, 它是描述自然界和非线性系统中不光滑和不规则几何体的有效工具, 其计算方法已经有多种, 应用领域也是十分广泛。然而, 各种方法各有不同, 文中就此对常用分形维数计算方法进行了系统的综合与研究, 主要包括圆规法、明科斯基方法、变换方法、盒子计算方法、周长- 面积法、裂缝岛屿方法、分形布朗模型法, 对每种方法的含义和模型及相关的应用领域进行了阐述, 并给出了其方法的计算机实现算法。

关键词: 分形; 分形维数; 计算方法
中图分类号: TP301.6 文献标识码: A 文章编号: 1005- 3751(2004)09- 0017- 03

Fractal Dimension Study of Calculating Method

NIE Du-xian, ZENG Wen-qu, WEN You-wei
(Faculty of Computer, Guangdong University of Technology, Guangzhou 510090, China)

Abstract: As one of science study tools, fractal dimension is usually an effective way that used to describe the non- smooth and non- regular geometry objects in the nature and non- linear systems. There are all kinds of methods about calculating it. Its application is greatly wide, but there are distinctive differences between these methods. According to these, this paper studies and integrates these methods systematically, including Compass Method, Minkowski Method, Variation Method, Box counting Method, Perimeter Area Method, Slit island Method, Fractal Brown model Method. This paper also expatiated the meanings, basic models and the application in each field. At the same time, algorithms of these methods implemented by the computer are found in the paper.

Key words: fractal; fractal dimension; calculating methods

0 引 言

分形几何学是 Mandelbrot 在 20 世纪 70 年代创立的。“分形(fractal)”一词也是他提出的, 它来源于拉丁语“fraetus”, 含义是“不规则”和“破碎”。分形科学是非线性科学的主要分支之一, 它在自然科学各个学科中, 甚至在经济和社会活动中, 都有着广泛的应用, 如材料学、地质、矿物、图像处理等领域。

Mandelbrot 指出分形具有三个要素[1]: 形状(form)、机遇(chance)和维数(dimension)。分形的形状是支离破碎、参差不齐的不规则形状; 机遇是描述对象的差异性; 分形维数可以是分数, 它也是分形几何对象的一个重要特征量。为了准确描述分形对象, 早在 1919 年 F. Hausdorff 提出了 Hausdorff 维[2], 它是基于集合的“覆盖”和“豪斯多夫测度”思想。Hausdorff 维具有对任何集合 $\boldsymbol{F}$ 都有定义和便于从理论上研究的优点, 而其缺点主要是无法应用于工程上, 因为在现实生活中, 不存在无限小。并且, 在工程上往往需要计算分形维数, 因此, 有必要对分形维数的计算方法进行研究。

1 分形维数的计算方法

1.1 圆规法(又称分规法 CM)

设 C 是分形曲线, 则 Mandelbrot[3] 给出:

$$L_E = L_H \delta^{1-D} \tag{1}$$

式中: L_E 表示 C 的欧式长度, L_H 表示 C 的 Hausdorff 长度, δ 表示 C 的标度, D 是 C 的分形维数。如果 δ 为分规间距, 则 L_E 可看作是间距为 δ 的分规测量 C 所得的长度, $\frac{L_E}{\delta}$ 则为以 δ 为步长, 测量 C 时所得到的步数, 记作:

$$N(\delta) = \frac{L_E}{\delta} \tag{2}$$

变形(2)式得: $\frac{L_E}{\delta} = L_H \delta^{-D}$ (3)

对(3)式两边先取对数再取极限有: $D = \lim_{\delta \to 0} \frac{\ln N(\delta)}{-\ln \delta}$ (4)

这里 D 就是分规法计算出来的分形维数, 在不同的步长 δ_i(标度)下, 测量分形曲线 C 得到不同的 $N_i(\delta_i)$, 在双对数系数中, 拟合数据$(-\ln\delta_i, \ln N_i(\delta_i))$所得直线斜率即为 D。

收稿日期: 2003- 12- 11
基金项目: 广东省基金资助项目((2003)32475)
作者简介: 聂笃宪(1974—), 男, 湖南邵阳人, 硕士研究生, 研究方向为图像处理。

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1.2 明科斯基方法

当一个表面的侧面 T 被半径为 r 的圆盘所覆盖时, 可以获得一条"香肠", 取名为"明科斯基香肠", 这个香肠有面积 $A(Tr)$[4], 则

$$D = \lim_{r\to 0} \frac{\ln A(Tr)}{-\ln r} \tag{5}$$

这里 D 是侧面 T 的分形维, 在不同的半径 r_i, 得到不同的 $A_i(Tr_i)$, 在双对数系数中, 拟合数据 $(-\ln r_i, \ln A_i(Tr_i))$ 所得直线斜率即为 D。

这个方法常常用在图像分析中, 但它要求侧面具有和整体不同的局部形态学, 其次, r_{max} 与整个侧面的长度 (L) 比较要小 $(r_{max} < L/6)$。否则, 计算结果误差会很大。

1.3 变换方法

它是由 Tricot 引入的, 是基于范围 η 的摆动, 对于曲线 T, 参数形式为 $z(t)$, 其中 t 在 $[a, b]$ 内变化, 定义摆动为:

$$\mathrm{osc}_\eta(t) = \sup_{|t-t'|\le\eta}(z(t') - z(t)) - \inf_{|t-t'|\le\eta}(z(t') - z(t)) \tag{6}$$

摆动 η- osc 是和给定的 t 有关, 盒子区域面积可由如下变换来代替[4]:

$$\mathrm{var}_\eta(z) = \int_a^b \mathrm{osc}_\eta(t)\,\mathrm{d}t \tag{7}$$

分形维数: $D = \lim_{\eta\to 0} \frac{\ln \mathrm{var}_\eta(z)}{-\ln\eta}$ (8)

这个方法存在问题: a. 当 $t < a+\eta$ 或者 $t > b-\eta$ 摆值没有定义; b. η_{max} 的选择依赖侧面的不规则性; c. 选择计算 η - osc 摆动的点数目不能太小 (不小于 1000)。改进的方法称为 ANAM 方法, 它是基于自相关最小化值, 通过对整个范围 η 的侧面局部不同取平均值, 这里引入 K_η^a,

$$K_\eta^a = \frac{1}{b-a}\int_{x=a}^{x=b}\Big[\frac{1}{\eta^2}\int_{t_1=0}^{\eta}\int_{t_2=0}^{\eta} |\, z(t+t_1) - z(t-t_2)\,|^a \mathrm{d}t_1\mathrm{d}t_2\Big]^{\frac{1}{a}}\mathrm{d}t \tag{9}$$

此时维数为: $D = \lim_{\eta\to 0} -\frac{\ln K_\eta^a}{\ln\eta}$ (10)

这里 D 是曲线 T 的分形维数, 对于不同的 η, 会有不同的 K_η^a 值, 在双对数系数中, 拟合数据 $(-\ln\eta, \ln K_\eta^a)$ 所得直线斜率即为 D。这个方法广泛用于裂缝模型建立的研究, 但是在图像离散化的过程中, 容易受到噪声等因素的影响。

1.4 盒子计算方法

设 $A \subset R^n$ 是一个非空集合, 在欧式距离下, 用边长为 $\frac{1}{T_n}$ 的小盒子紧邻地去包含 A, 设 $N_n(A)$ 表示包含 A 所需要的最小盒子数, 则

$$D = \lim_{n\to\infty} \frac{\ln N_n(A)}{-\ln T_n} \tag{11}$$

即为集合 A 的盒维数。对二维平面上的集合, 盒维数的计算方法是: 逐渐增大 T_n, 分别计算出相应的 $N_n(A)$ 的值, 得到 $(-\ln T_n, \ln N_n(A))$ 数据对, 再利用最小均方差方法求出 $\ln N_n(A)$ 相对 $\ln T_n$ 的斜率, 这就是盒维数。此方法使用较为广泛, 但计算结果不是十分精确[5]。

1.5 周长– 面积法

设 C 是封闭的分形曲线, 则由分形几何的理论[1,6], 有:

$$L_H^{\frac{1}{D}} = a_0 A^{\frac{1}{2}} \tag{12}$$

式中: L_D 表示 C 的Hausdorff 长度, A 表示 C 所包围的欧式面积, D 表示 C 的分形维数, a_0 是无量纲常数, 称为形状因子, 由 (1) 式有:

$$L_E^{\frac{1}{D}} = L_H^{\frac{1}{D}}\delta^{\frac{1-D}{D}} = a_0 A^{\frac{1}{2}}\delta^{\frac{1-D}{D}} = a_0\delta^{\frac{1}{D}}A^{\frac{1}{2}}\delta^{-1} \tag{13}$$

即有: $\left(\frac{L_E}{\delta}\right)^{\frac{1}{D}} = a_0\left(\frac{A^{\frac{1}{2}}}{\delta}\right)$ (14)

对 (14) 式两边取对数, 再取极限有:

$$D = \lim_{\delta\to 0} \frac{\ln\left(\frac{L_E}{\delta}\right)}{\ln a_0 + \ln\left(\frac{A^{\frac{1}{2}}}{\delta}\right)} \tag{15}$$

D 就是周长– 面积法计算的分形维数。对于给定的几何图形, a_0 和 D 是常数, 所以在不同的尺度 $\delta_i (i = 1, 2, \cdots, n)$ 下, 测量分形曲线 C 得 n 个周长值 L_{E_i} 和 n 个面积值 A_i, 在双对数坐标系中, 对数据 $\left(\ln\left(\frac{A_i^{\frac{1}{2}}}{\delta_i}\right), \ln\left(\frac{L_{E_i}}{\delta_i}\right)\right)$ 作线性回归即可得 D。

1.6 裂缝岛屿方法(SIM)

这个方法[7,8] 首先是由 Mandelbrot 提出来测量裂缝表面的分形维数, 它是基于分形几何自相似的假设, 也就是用一个平面与裂缝表面相交, 交叉曲线的分形维数 $D_f = D'_f - 1$, 这里表面的维数为 D'_f。通过实验以后, 在裂缝样本的表面就镀上有镍或铬的电解质, 并经过适度打磨, 这样就形成了灰色"岛屿" 与银色的"湖" 共存的表面 (可以想象是用平行平面切割以后的结果一样)。在自然界的分形几何中, Mandelbrot 扩展了这种湖的周长和岛屿的面积关系, 这种方法的理论基础是:

$$C(r) = \frac{\left(\sum P_i(r)\right)^{\frac{1}{D_f}}}{\left(\sum A_i(r)\right)^{\frac{1}{2}}} \tag{16}$$

这里 $C(r)$ 是一个不依赖于岛屿的大小、形状和标度的常数, $P_i(r)$ 为周长, r 是标度, $A_i(r)$ 是第 I 个岛屿的面积, D'_f 是裂缝表面的维数。分形维数通过 $\ln A_i(r)/\ln P_i(r)$ 线性回归分析而得到。这样计算时可能有多个维数值, 最后裂缝表面的维数应取其平均值; 使用这个方法的前提条件是, 裂缝部分岛屿的自相似性, 而不是自仿射性, D'_f 必须是一个常数, 对形变裂缝, 如塑料的形变, 可能有多个 D 值。

1.7 分形布朗模型法

若 $f(x)$ 是分形布朗函数, 则其概率分布满足:

$$P_r\left\{\frac{f(x+\triangle x)}{\|\triangle x\|^H} < y\right\} = F(y) \tag{17}$$

式中 x 为 N 维欧式空间 R^n 的任一点; $F(y)$ 为高斯分布概

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

率函数; H 是描述自相似性或表面不规则度的因子, 它与分形维数 D, 以及函数的拓扑维数 N 满足:

$$D = N + 1 - H \tag{18}$$

当 $F(y)$ 取为 $N(0, \delta^2)$ 时, 由(17) 式可得:

$$E \mid f(x + \Delta x) - f(x) \mid^2 = C \parallel \Delta x \parallel^{2H} \tag{19}$$

式中 C 是常数, 对(19) 式两边取对数有:

$$\ln E \mid f(x + \Delta x) - f(x) \mid^2 = 2H \ln \parallel \Delta x \parallel + \ln C \tag{20}$$

通过最小二乘法拟合数据点对{ $\ln E \mid f(x + \Delta x) - f(x) \mid^2, \ln \parallel \Delta x \parallel$ }, 则由拟合直线的斜率即可求得 H。若对图像而言, 则 $f(x, y)$ 是整个图像, 拓扑维数 $N = 2$, 则图像的分形维数可以表示为: $D = 3 - H$。此方法常用于人工景物与自然景物的识别, 但计算量大, 多应与小波变换一起使用[6]。

2　维数方法的仿真

以上介绍了分形维数计算方法, 下面应用它们来计算图像的分形维数。首先介绍盒子计算方法, 为了方便计算机处理, 这里使用 *BMP* 文件格式图像, 输入图像的宽为 W, 高为 H, 均以像素为计量单位。盒子的边长 L 定义为 $\min(W, H)$ 的 T_n 等分对 $\min(W, H)$ 的归一化值, 即 $L = 1/T_n$, 考虑到图像分辨率问题, T_n 的取值为不超过 $\min(W, H)$ 的整数。设 F 代表整个图像组成的集合, 按选定盒子边长 $L = 1/T_n$ 分割 F 为 I 列 J 行紧密相邻的二维子图像阵列, 每个子图像是边长为 $\mathrm{int}(\min(W, H)/T_n)$ 个像素的正方形区域, 构成的集合可记为 $A_{ij}(i, j \in Z-, i < I, j < J)$, 满足 $A_{ij} \cap A_{mn} = \varnothing (i = m$, 与 $j = n$ 不同时成立), $\cup A_{ij} \supset F$, 所有子图像构成了 F 的一个覆盖。$\mathrm{Nn}(A)$ 为含有图像的所有盒子数, 算法中盒维数的计算公式为:

$$D = \frac{\ln N_n(A) - \ln N_{n-1}(A)}{\ln T_n - \ln T_{n-1}} \tag{21}$$

算法的步骤如下:

(1) 读入图像文件, 得图像的宽 W 和高 H, 均以像素为单位;

(2) 比较 W 和 H, 得出 $\min(W, H)$;

(3) 取分形盒子边长 $1/T_n$;

(4) 按照所选取的盒子边长分割图像为 IxJ 个子图;

(5) 统计整个图中含有图像边缘的盒子数 Nn;

(6) 保存 Nn 和 T_n;

(7) 如果 $\dfrac{\ln N_n(A) - \ln N_{n-1}(A)}{\ln T_n - \ln T_{n-1}} \leq \delta$(阈值), 进入下一步, 否则增加 T_n, 返回步骤(3);

(8) 根据(21) 式计算维数 D。

明科斯基方法算法: *a*. 设置半径 $R(R < L/6)$ 的初值; *b*. 选择曲线的起始点 $P1$; *c*. 以 $P1$ 为圆心, R 为半径画圆, 记圆与曲线的交点为 $P2$; *d*. 以 $P2$ 为圆心, R 为半径画圆, 重复此过程直到曲线结束, 保存 R 和圆个数 $N(R)$ 的值; *e*. 减少 R 的值, 重复 *a*. 至 *b*. 步直到 R 小于某个阈值; *f*. 在双对数系数中, 拟合数据($\ln R$, $\ln N(R)$) 得分形维数 D。其他方法的算法步骤可以参照盒子计算方法。图 1、图 2 和表 1 是对 Lena 图计算的结果。

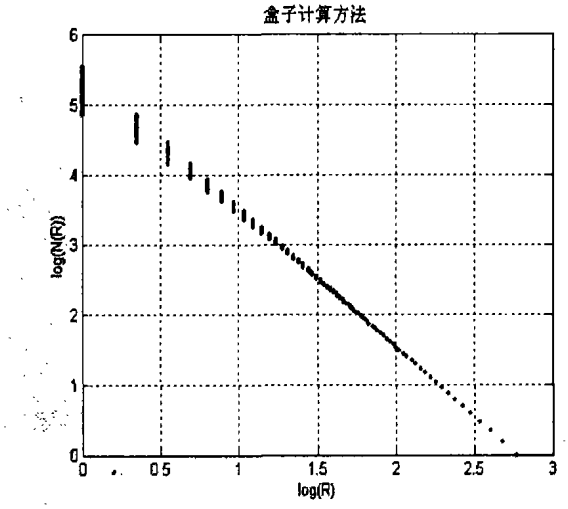


图 1　盒子算法的线性回归 D 值图

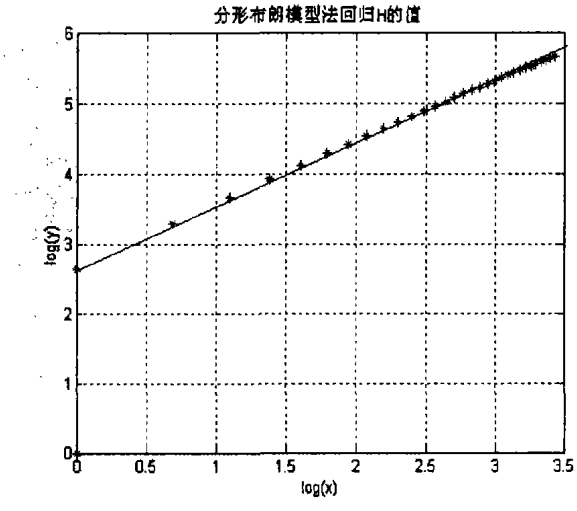


图 2　分形布朗模型法线性回归 H 值图

表 1　使用各种方法计算 Lena 图的分形维数结果

计算方法	分形维 D 值
分规法	2.01731
明科斯基方法	2.01564
变换方法	2.08967
盒子计算方法	2.01665
周长- 面积法	2.01651
裂缝岛屿方法	2.01646
分形布朗模型法	2.01652

3　小　结

计算分形维数是建立各种模型如裂缝模型、人工景物和自然景物识别等的基础, 文中主要系统地综合了各种计算分形维数的方法, 并指出了各种方法的具体不同, 同时给出了实现的算法。

(下转第 22 页)

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

```
generate_ attribute_ combinations(k):
  if k = 1
    create AC1  //AC1 表只有一个属性(设为 F1), 其中的记录为 RDB 的所有属性
  else
    select ACk-1.* , AC1.F1 as Fk from ACk-1, AC1 where ACk-1.Fk-1 < AC1.F1 into table ACk// 用 连接 的方法产生 k 属性组合表 ACk
  endif
  return ACk
generate_ frequent_ attribute_ valuse_ sets(ACk):
  create table Lk(lf1 c, lf2 c, …, lfk c, supp n)// 建立频繁属性值表 Lk 的存储结构
  select ACk
  scan// 扫描 ACk 中的所有记录
    scatter to Acv[k]// 将 ACk 中的当前记录存入数组 Acv 中, 该组数据实际为 RDB 中所有属性的某个 k 组合
    insert into Lk;
    select &Acv[1], &Acv[2], …, &Acv[k], count(*) from RDB;
    group by &Acv[1], &Acv[2], …, &Acv[k] having count(*) ≥ min_ sup
    // 通过聚集操作,得到 RDB 中当前 k 属性组合的频繁值集, 结果存入 Lk 中
    if returncount = 0  //" 剪枝" 操作, 即如果聚集操作返回的当前属性组合下的频繁值集为空, 则从 ACk 中删除当记录, 以减少 ACk+1 的属性组合数
      select ACk
      delete
  endif
  select ACk
endscan
return Lk
```

经对比测试, 在相同的计算机上利用本算法和文献[8]中的算法对同一数据集进行操作, 两算法生成了相同的频繁属性值集, 而后者的运行时间是前者的3~ 4 倍, 因此说明本算法在生成关系数据表中频繁属性值集的效率是较高的。

4 结论与展望

文中在分析了关系数据库中隐含的关联规则一般具有多值、多维性后, 借鉴相关文献的挖掘思想, 利用 SQL 对关系数据库操作便捷而高效的特性, 提出了一个在关系数据库中挖掘关联规则的有效算法。该算法不但对数据集中的数据类型具有较强的适应性, 而且在引入概念分层和约束条件后, 只需对算法做部分扩展就可以挖掘出关系数据库中多层型和基于某些约束条件的关联规则。

参考文献:

[1] Agrawal R, Imielinski T, Wami A S. Mining association rules between sets of items in larger databases[A]. In Proc. of ACM SIGMOD Conference on Management of Data[C]. Washington, DC:[s.n.], 1993. 207- 216.

[2] Srikant R, Agrawal R. Mining Quantitative Association Rules in Large Relational Tables[A]. Proc. 1996 ACM SIGMOD Int'l Conf. Very Lage DataBases[C]. Montreal, Canada:[s.n.], 1996. 1- 12.

[3] 王 凌, 李云峰, 逄焕利, 等. 数据集中多属性关联规则发现算法[J]. 吉林工学院学报, 2000, 21(4): 26- 28.

[4] 王选文, 丁 夷, 范九伦. 关联规则挖掘在人事系统中的应用[J]. 西安邮电学院学报, 2001, 6(1): 21- 23.

[5] Agrawal R, Shim K. Developing tightly- coupled data mining Applications on a Relational Database System[A]. In Proc of the 2nd Int'l Conference on Knowledge Discovery in Databases and Data Mining[C]. Portland, Oregon:[s.n.], 1996. 287 - 290.

[6] 杨炳儒, 孙海洪, 熊范纶. 利用标准 SQL 查询挖掘多值型关联规则及其评价[J]. 计算机研究与发展, 2002, 39(3): 307- 312.

[7] 董 淳, 王敏慧, 李孟恒, 等. 关系表中联系规则挖掘的设计和实现[J]. 计算机工程, 1999, 25(1): 14- 15.

[8] Agrawal R, Srikant R. Fast algorithms for mining association rules[A]. In Proc. of the 20th Int'l Conf. On Very Large Databases[C]. Santiago, Chile:[s.n.], 1994. 478- 499.

[9] Houtsma M, Swami A. Set- oriented mining of association rules[R]. Research report RJ 9567, Sam Jose, California: IBM Almaden Research Center, 1993. 32- 44.

(上接第 19 页)

参考文献:

[1] 谢和平, 薛秀谦. 分形应用中的数学基础与方法[M]. 北京: 科学出版社, 1998.

[2] 法尔科内 K. 分形几何- 数学基础及其应用[M]. 曾文曲等译. 沈阳: 东北大学出版社, 1991.

[3] Mandelbrot B B. The fractal geometry of nature[M]. San Francisco: Freeman, 1982.

[4] Charkaluk E, Bigerelle M. Fractals and fracture[J]. Engineering Fracture Mechanics, 1998, 61: 119- 139.

[5] Mandelbrot B. Les objects fractal: forme, hazard et dimesion [M]. Paris: Flammarion, 1995.

[6] 田 杰, 陈 杰. 基于小波变换及分形特征的目标检测与识别[J]. 北京理工大学学报, 2003, 2(1): 95- 99.

[7] Charkaluk E, Big M. Fractals and fractal scaling in fracture mechanics[J]. Engineering Fracture Mechanics, 1998, 36: 119 - 139.

[8] Tricot C, Ferland P, Baran G. Fractal analysis of woen surfaces [J]. Wear, 1994, 172: 27- 33.

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net