分形维数(Fractal dimension)及其测量方法

孙博玲
(哈尔滨学院, 哈尔滨, 150086)

摘　要　给出了相似维数、容量维数、盒子维数、信息维数、关联维数、广义维数分形维数等 6 种分形维数的定义方式, 并讨论了它们的应用范围。在此基础上, 概括出改变粗视化程度求维数、根据测度关系求维数、根据关联函数求维数、根据分布函数求维数、根据波谱求维数等 5 种测量维数的方法, 为不同领域的科研人员应用维数测量提供了有力工具。

关键词　分形维数; 粗视化; 无标度区
分类号　O144

Fractal Dimension and Its Metrical Methods/Sun Boling(Computer Department of Harbin College, Harbin 150086, P. R. China)//Journal of Northeast Forestry University. – 2004, 32(3). – 116~ 119

The definitions of six kinds of fractal dimension are introduced, including similarity of dimensions, capacity dimension, box dimension, information dimension, interrelation dimension and broad dimension, and their scopes of application are also discussed. On the basis of which five methods of measuring dimension are summed up in terms of megascopic degree, estimation, interrelation function, distribution function and spectrum function. This research provides a powerful tool for researching staff from different fields in dimension measurement.

Key words　Fractal dimention; Megascopic; Non– gradational area

虽然数学家豪斯道夫(F. Hausdorff)于 1919 年就提出了分数维的概念, 但对集合维数性质的研究直到 20 世纪 70 年代才引起学术界的广泛重视。1975 年, 美籍法国数学家曼德尔布罗特(B. B. Mardelbrot)出版了分形几何学的第一部著作《分形、形、机遇和维数》, 认为分形集这类奇异集合的性质不能用欧氏测度来刻划, 而维数恰是此类集合尺度变化下的不变量, 并主张用维数来刻划这类集合; 于是提出了分形维数的概念, 即刻划分形集合性质的维数叫分形维数。它来源于分数维的概念, 但由于现实的分形集往往从不同的侧面体现出分形性, 所以也就存在多种分形维数定义以及各自的测量方法。

1　分形维数

1.1　相似维数

先让我们看一个简单的事实。根据相似性, 如图 1 把各图形的边长 2 等分, 当然, 线段是一半长度的 2 个线段, 正方形则是每边为原来 1/2 的 4 个正方形, 而立方体则是 8 个。也就是说, 线段、正方形、立方体可被看成为分别由 2、4、8 个把全体分成 1/2 的相似形组成。2、4、8 数字还可以写成 2^1、2^2、2^3, 显然这里的指数与其图形的经验维数相一致。推而广之, 若某图形是由把全体缩小成 1/a 的 b 个相似形所组成, 由于 $b= a^D$, 则有

$$D= \log b/\log a。\qquad (1)$$

此 D 便是几何图形的维数, 由于它是通过相似变换得来的, 所以此维数一般称为相似维数。用 D_s 表示。显然 D_s 不必一定是整数, 所有经典分形集的维数都可由此式计算, 如 Cantor 集。$a= 3$, $b= 2$, $D_s= \log 2/\log 3= 0.630\ 9\cdots$。Koch 曲线 $a= 3$, $b= 4$, $D_s= \log 4/\log 3= 1.261\ 8\cdots$。

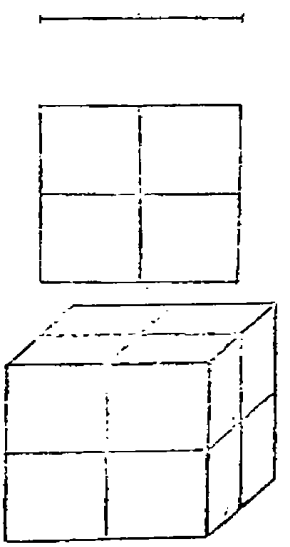


图 1　把线段、正方形、立方体的单位长度分成一半

虽然相似维数是把经验维数扩大为非整数值的划时代的量, 但它的适用范围毕竟是有限的。因为只有严格相似性规则的分数维图形, 才能定义为这个维数, 而实际上, 大多数分形图形都不是严格自相似的, 如某些随机图形, 为此有必要定义新的维数。

1.2　容量维数

假定要考虑的图形是 d 维欧几里得空间 R^d 中的有界集合。用半径为 ε 的 d 维球包覆其集合时, 假定 $N(\varepsilon)$ 是球的个数的最小值。容量维数 D_c 可用下式来定义

$$D_c \equiv \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}。\qquad (2)$$

此定义与 Hausdorff 维数很相似(http://fractal.com.cn)。在 Hausdorff 维数中, 虽然把球的大小作为比 ε 还小的任意球, 但如果把它限定在 1 个大小为特殊情况下则为容量维数。

作者简介: 孙博玲, 女, 1972 年 11 月生, 哈尔滨学院计算机系, 讲师。

收稿日期: 2004 年 4 月 6 日。
责任编辑: 戴芳天。

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

D_C 虽常与 D_H 相一致,但有时也取不同的值,一般的关系是

$$D_C \geqslant D_H 。 \tag{3}$$

维数的Hausdorff 定义和容量定义,在数学上都是很严密的。但要广泛用于自然科学,有时也有不适之处。比方说,不论哪个定义都把包覆球的半径的极限考虑为 0,这在实际测定中是不能达到的。所以有必要将维数定义改成更实用一些的。

1.3　盒子维数

容量维数的定义,提示了一种测量分形的方法。取边长为 ε 的小盒子,把分形覆盖起来。由于分形内部有各种层次的空洞和缝隙,有些小盒子会是空的。数有多少盒子不是空的,把这个数目记为 $N(\varepsilon)$。然后缩小盒子的尺寸 ε,所数得的 $N(\varepsilon)$自然要增大。根据前面的定义,只要在双对数坐标纸上画出 $\ln N(\varepsilon)$ 对 $\ln\varepsilon$ 的曲线,其直线部分的斜率就是此分形对象的盒子孙维数 D_0。

这种看起来很简便的“数盒子”方法,有着理论和实践两方面的局限性。对于实际计算,只有分维小于二维或在二维附近,而相空间维数也不高时,它才是可行的;维数增高后,计算量迅速上升,以致很难得到收敛的结果。从理论上看,一个小盒子不管是包含了分形的一个点或是一批点,都算是非空的,可在 $N(\varepsilon)$中占有一席之地,这就完全不能反映分形内部的不均匀性。不过分维的定义就是如此;只有修改维数的定义,才能改进描述的细致程度。

1.4　信息维数

让我们把小盒子编号。如果知道分形中的点落入第 i 只盒子的概率是 P_i,那就可以写出用尺寸为 ε 的盒子进行测算所得的信息量。

$$I = -\sum_{i=1}^{N(\varepsilon)} P_i \ln P_i 。$$

然后就可以用 I 代替 $N(c)$,定义“信息维数” D_i:

$$D_i = \lim_{\varepsilon \to 0} \frac{\sum_{i=1}^{N(\varepsilon)} P_i \ln P_i}{\ln \varepsilon} 。 \tag{4}$$

假如落入每只盒子的概率都相同,即 $P_i = 1/N(\varepsilon)$,求和记号下面的每一项都和编号 I 无关,则 $I = \ln N(\varepsilon)$,于是就回到了前面盒子维数 D_0 的定义,一般说来

$$d \leqslant D_1 \leqslant D_0 \tag{5}$$

信息维数不是新概念。早在 20 世纪 50 年代匈牙利数学家瑞奈(A. Renyi)就从概率论角度建立了上面的定义,不过当时没有把它和分形联系起来[2]。从计算角度看,信息维数 D_1 的算法比盒子维数 D_0 还要费事一些(要先求 P_i),因而也是不切合实际的。

1.5　关联维数

关联函数是最基本的统计量之一,从这一函数型也可求得分形维数。

如果把在空间随机分布的某量坐标 X 处中的密度记为 $\rho(x)$,则关联函数 $C(\varepsilon)$ 可用下式定义

$$C(\varepsilon) \equiv < \rho(x)\rho(x+\varepsilon) > 。 \tag{6}$$

这里< …> 表示平均。根据情况,平均可以是全体平均,也可以是空间平均。如果分布各个方向均等,只能用两点间的距离 $\varepsilon = |\varepsilon|$ 的函数来表示关联函数。

如果关联函数是幂型,则两点间的距离便不存在特征长度,关联总是以同样比例衰减。假如关系为

$$C(\varepsilon) \propto \varepsilon^{-a} , \tag{7}$$

则有

$$a = d - D 。 \tag{8}$$

式中: d ——空间维数;

D ——分形维数。

1983 年, P. Grassberger 和 J. Procassia 给出了关联维数的定义[1]:

$$D_2 = \lim_{\varepsilon \to 0} \frac{\ln C(\varepsilon)}{\ln \varepsilon} 。 \tag{9}$$

式中

$$C(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} H(\varepsilon - |x_i - x_j|) 。 \tag{10}$$

1.6　广义维数

H. G. E. Hentschel 等提出了广义维数的概念,其定义是

$$D_q = -\lim_{\varepsilon \to 0} \frac{S_q(\varepsilon)}{\ln \varepsilon} 。 \tag{11}$$

式中

$$S_q(\varepsilon) = \frac{1}{1-q} \ln \left[\sum_{i=1}^{N(\varepsilon)} P_i^q \right] \tag{12}$$

是 q 阶 Renyi 信息, D_q 叫作 q 阶广义维数,有时也叫 Renyi 信息维数。

式(11)与式(12)中的 q 取正负整数。显然,当 $q = 0$ 时,可证 $D_{q|q=0} = D_0$;当 $q = 1$ 时,利用罗必塔法取极限可得 $D_1 = \lim_{\varepsilon \to 0} D_q$;且 $D_{q|q=2} = D_2$,即 $q = 2$ 时,广义维数 D_q 就是关键维数 D_2。

2　分形维数的基本测量方法

2.1　改变粗视化程度求维数的方法

这是基于盒子维数和信息维数的定义设计的一种测量分形维数的方法[3]。即用圆和球、线段和正方形、立方体等具有特征长度的基本图形去近似分形图形的一种方法。

例如,用长度为 ε 的线段集合近似图 2 海岸线的复杂曲线。做法如下:首先把曲线的一端作为起点,然后以此点为中心画一个半径为 ε 的圆,把此圆与曲线最初相交点和起点用直线连接起来,把此交点重新看作起点,以后反复进行同样的操作。像这样用长度 ε 的折线去近似海岸线时,把必要的线段总数作为 $N(\varepsilon)$。若基准长度 ε 变小,则 $N(\varepsilon)$增加。如果在双对数坐标上画出 $\ln N(\varepsilon)$ 对 $\ln\varepsilon$ 的曲线有直线部分,其斜率就是此海岸线的分形维数。

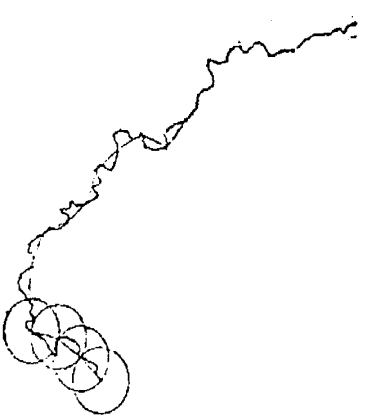


图 2　用折线近似海岸线

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

与此相似,把空间分割成边长为 ε 的细胞,然后来数所要考虑的形状中的那部分所含的细胞数 $N(\varepsilon)$。例如求平面上点的分布的分形维数,首先,用间隔为 ε 的格子把平面分割成边长为 ε 的正方形,而且,数出在此平面上至少包含一个点的正方形个数,并把此数记为 $N(\varepsilon)$(见图 3)。如果对 ε 进行各种变化时,当

$$N(\varepsilon) \propto \varepsilon^{-D} \tag{13}$$

的关系得到满足,这些点的分布即为 D 维数。这一方法不仅适用于点分布和曲线形状,而且还可以解析像河流这种含有大量分岔的图形,所以是个利用价值较高的方法。

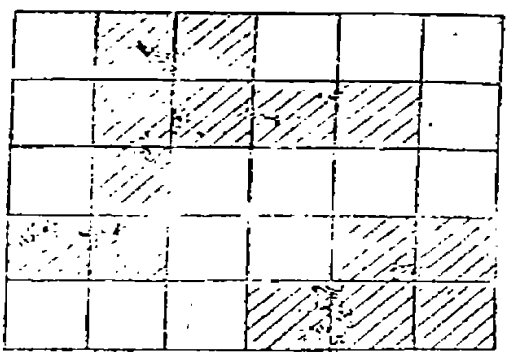


图 3 根据正方形粗视化

若把这一方法引进信息量的量度,便可得到分形的信息维数,这是对概率点分布有效的方法。

2.2 根据测度关系求维数的方法

这是根据分形维数具有非整数维数的测度性质来设计求维数的一种方法。

若把立方体一边的长度扩大到 2 倍,那么 2 维测度的表面积即为 2^2 倍,3 维测度的体积即为 2^3 倍。因此,若把单位长度扩大到 2 倍,并假定它能成为具有 2^D 倍的量,那么此量也可称之为 D 维数的。

一般若假定长度为 L,面积为 S,体积为 V 时,则可得出如下关系式

$$L \propto S^{1/2} \propto V^{1/3}。 \tag{14}$$

这一关系式的意义是,若把 L 扩大到 k 倍,那么 $S^{1/2}$ 和 $V^{1/3}$ 也都扩大到 k 倍。若把具有 D 维度的量假定为 X,则可把(14)式变成一般公式

$$L \propto S^{1/2} \propto V^{1/3} \propto X^{1/D}。 \tag{14'}$$

于是我们可以利用这个关系式很容易求出某些分形对象的维数。

例如在测定岛屿海岸线的分形维数时,假定面积为 S,海岸线长度为 X。因岛的面积明显具有 2 维测度,所以根据 $S^{1/2} \propto X^{1/D}$ 即可求得海岸线的分形维数 D。因为视 X 的维数为未知数,所以需要一定的技巧。最便利和常用的方法是使空间量子化,把面积 S 和长度 X 都作为自然数。

首先,尽量用细格子把所考虑的平面分割成小正方形的集合体。然后把那些既使包含一小点岛的正方形涂黑(图 4)。把黑正方形的个数记为 S_N,把与白正方形相接的黑正方形的个数记为 X_N。如果单位正方形的大小足够小的话,则可以为 $S \propto S_N$, $X \propto X_N$ 是成立的。对许多不同面积的岛可用同一方法去求出 S_N 和 X_N,如果存在能够满足下式的 D,

$$S_N^{1/2} \propto X_N^{1/D}。 \tag{15}$$

那么海岸线的分形维数就是 D。此时单位正方形的大小越小误差也就越小,与用粗视化方法最不同之处是,不变化单位正方形的大小,而预先尽量把它固定得小一点。

此问题也可以换一种方式来求。考虑部分长海岸线的一部分时,所要调查的该段海岸线两端的直线距离 L 和海岸线长度 X_N 的关系。若把所要调查的海岸线部分的大小进行各种变化,则可得到大量 L 和 X_N 的组合。此时,在 L 与 X_N 之间,如果下述关系式成立

$$L \propto X_N^{1/D}, \tag{16}$$

这个 D 也同样是海岸线的分形维数。

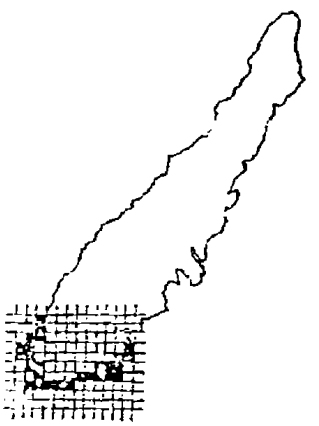


图 4 尽可能用小的正方形使岛的形状粗视化

2.3 根据关联函数求维数的方法

此方法是依据关联维数的定义设计的,多用于对实验数据的处理。

考虑实验中测得的一个数据序列 $x_1, x_2, x_3, \cdots, x_i, \cdots x_i$ 是第 i 时刻测得的值。例如,第 i 秒时布朗粒子的模座标。由于不知道实际的相空间维数有多高,我们先用这些数据支起一个 m 维的空间。造这个 m 维的"嵌入空间"的办法很多。例如取 $m=10$,把 $x_1, x_2, x_3, \cdots x_{10}$ 作为十维空间中的一个矢量 y_1。然后右移一步,把 $x_2, x_3, x_4, \cdots x_{11}$ 作为十维空间中的第二个矢量 y_2。这样构造出一大批矢量 $y_1, y_2, y_3, \cdots y_k$。现在随便给定一个数 ε,然后检查一遍有多少对点 (y_1, y_j) 之间的距离小于 ε,把距离小于 ε 的"点对"中所占比例记作 $C(\varepsilon)$。用计算机很容昌从原始数据中算出 $c(\varepsilon)$。

如果 ε 取得太大,当然一切"点对"的距离都不会超过它,因为,$C(\varepsilon)=1$,取对数后有 $\ln C(\varepsilon)=0$。这样的 ε 当然反映不了系统的内部性质。适当缩小 ε,可能在一段 ε 区内有

$$C(\varepsilon) = \varepsilon^{\gamma}。 \tag{16'}$$

和分形维数的定义相比,可见 γ 是一种维数。实际上 γ 是对关联维数 D_2 的很好逼近。如果 ε 取得太小,实验中一切偶然噪声便会表现出来。由于噪声在任何一维上都起作用,m 维空间中就会测得 $\gamma = m$。

这样,我们看到实际系统的尺度变换,在大小两头受到限制。ε 太大时,$\gamma = 0$;ε 太小时,$\gamma = m$,都与所研究的系统的本质没有关系。只有中间 $C(\varepsilon) = \varepsilon^{\gamma}$ 的一段,才对应无标度区。有没有无标度区,不取决于主观愿望,而决定于客观事实。上述维数测量方法的好处在于:只要对一批 m 值画出 $\ln C(\varepsilon) - \ln\varepsilon$ 曲线,看看是否在斜率等于 0 和 m 的两段直线之间还有一段斜率 γ 介于 0 和 m 之间的直线,它就自动指出无标度区的范围,并且表明实验数据的噪声背景有多大(大致

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

相当斜率以 γ 变到 m 处的 ε 值)。如果斜率直接从 0 变到 m, 往往是嵌入空间的维数 m 选得不够大。如果调整 m 之后还看不到无标度区, 则证明分形维数不存在。

2.4 分布函数求维数的方法

月面照片上的各种不同大小的月坑, 如果只看照片, 比例尺是完全看不出来的, 也就是说月坑的大小分布并没有特征长度。考虑这种大小分布时, 从其分布函数的类型即可求得分形堆数[4]。

把月坑直径记为 ε, 把直径大于 ε 的月坑存在概率记为 $P(\varepsilon)$。若把直径的分布密度记为 $P(s)$, 则有

$$P(\varepsilon) = \int_{\varepsilon}^{\infty} P(s)\mathrm{d}s。 \tag{17}$$

若想变换比例尺而分布类型不变, 对任意的 $\lambda > 0$,

$$P(\varepsilon) \propto P(\lambda\varepsilon) \tag{18}$$

必须成立。能常满足(18) 式的 ε 函数型, 只限于下面的幂型

$$P(\varepsilon) \propto \varepsilon^{-D}。 \tag{19}$$

当考虑用粗视化看不见小于 ε 的月坑时, 则能看见月坑的数目与 $P(\varepsilon)$ 成比例。变换粗视化程度, 在看不见小于 2ε 的月坑时, 能看见的月坑数与 $P(2\varepsilon)$ 成比例, 此数是用大小 ε 粗视化时的 2^{-D} 倍。一般若把用大小 ε 粗视化时看见的个数假定为 $N(\varepsilon)$, 因 $N(\varepsilon)$ 与 $P(\varepsilon)$ 成比例, 这里出现的 D 则与用粗视化的分形维数的定义(13) 相一致。

2.5 根据波谱求维数的方法

根据观测对空间或时间的随机变量的统计性质进行调查时, 往往可以较简单地得出用波数分解变动的波谱。而从波谱的角度看, 所谓改变观测的粗视化程度就是改善截止频率 f_c。这里所说的截止频率, 是指把较此更细的振动成分舍去的界限频率。因此, 如果说某变动是分形, 那么也就等于说既使变换截止频率 f_c 也不改变波谱的形状。具有这种性质的波谱 $s(f)$ 只限于下述幂型:

$$s(f) \propto f^{-\beta}。 \tag{20}$$

此幂的指数 β 与分形维数的关系可由下述事实得知。例如, 因某电路的电压 $V(t)$ 的杂音而产生变动的图表, 而其电路电压 $V(t)$ 可视为是时间的函数。当这种变动波谱为(20) 式时, 若把曲线图表的分形维数记作 D, 则

$$\beta = 5 - 2D, \tag{21}$$

这一关系即可成立。

3 结束语

分形理论还处于发展阶段, 不同学科对分形维数的应用各不相同。还没找到对任何事物都适用的分形维数的定义。尽管如此, 分形理论对解决以往难以解决的问题提供了一套很好的方法, 所以在自然科学和社会科学的许多领域都得到广泛的应用。

参 考 文 献

1 Grassberger P, Procaccia J. Characterization of strange Attractors. Phys Rev Lett, 1983, 50: 346~ 349

2 Renyi A. Probability Theory. Nolland: Amsterdan, 1970

3 郝柏林. 分形和分维. 科学, 1986(1): 55~ 56

4 高安秀树著. 分数维. 学步明译. 北京: 地震出版社, 1989

(上接 **109** 页)平最好。第四列即因素 ㊄(海藻酸钠质量分数) 的极差为 6.00, 它的水平改变对实验指标影响位居第三位, 它的三个水平所对应的数值分别为 70.33、71.67、65.57, 以第二水平对应的数值 71.67 为最大, 所以取它的第二水平最好。第二列即因素 ㊁(温度) 的极差为 5.00, 是 4 个因素中极差最小的, 说明它的水平改变时对实验指标的影响最小, 它的三个水平所对应的数值分别为 66.00、70.67、71.00, 以第三水平所对应的数值 71.00 最大, 但从各水平所对应的数值看 71.00 和 70.67 只差 0.33, 并且温度因素影响又最小, 而各水平值相差较大, 所以取它的第二水平既不影响效果又节约能源、利于生产操作。

表 3　正交试验极差分析

极差	iv	㊁	㊂	㊃
K_1	155	198	206	211
K_2	217	212	197	215
K_3	251	213	220	197
$k_1 = K_1/3$	51.67	66.00	68.67	70.33
$k_2 = K_2/3$	72.33	70.67	65.67	71.67
$k_3 = K_3/3$	83.67	71.00	73.33	65.67
极差	32.00	5.00	7.66	6.00
优选方案	iv_3	㊁$_2$	㊂$_3$	㊃$_2$

从以上分析可以得出各因素对实验指标(微囊合格率) 的影响按大小次序为 iv(转速) > ㊂(时间) > ㊃(海藻酸钠质量分数) > ㊁(温度); 因此, 最好的方案应当是 iv_3 ㊁$_2$ ㊂$_3$ ㊃$_2$, 即: 转速 3 200 r/min、温度 50℃、时间 40 min、海藻酸钠质量分数 1.0%。

这里分析出来的最好方案在已经做过的 9 次实验中没有出现, 与它比较接近的是 7 号实验。在 7 号实验中, 只有温度 ㊁不是处在最好水平, 但温度对微囊合格率的影响是三个因素中最小的。从实验做出的结果看出, 7 号实验中的微囊合格率是 87.00%, 是 9 次实验中最高的, 这说明找出的最好方案是符合实际的。

为了最终确定上面找出的实验方案 iv_3 ㊁$_2$ ㊂$_3$ ㊃$_2$ 是不是最好方案, 按这个方案又进行实验, 结果微囊合格率为 88.06%, 比 7 号实验的微囊合格率高出 1.06 个百分点。这既说明上面所确定的方案为最好方案, 同时又说明因素 ㊁(温度) 确实是对试样指标(微囊合格率) 影响较小的因素。

3 结论

通过实验及分析得出凝聚法制备微囊乳化的较好工艺条件为: 转速 3 200 r/min、乳化温度 50℃、乳化时间 40 min、海藻酸钠质量分数 1.0%、乳化剂 SSDG 用量 6.0%。

参 考 文 献

1 于钢, 钱学仁. 加工纸工艺. 哈尔滨: 东北林业大学出版社, 1997. 110

2 魏福祥, 王新辉, 杨晓宇, 等. 天然高分子海藻酸盐成膜研究. 日用化学工业, 1998(1): 22~ 24

3 陈魁. 试验设计与分析. 北京: 清华大学出版社, 1996. 94~ 113

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