

非线性科学及其在医学中的应用

涂承媛, 曾衍钧

(北京工业大学生物医学工程中心, 北京 100022)

[摘要] 简要介绍了非线性科学中的分形维方法和信息熵方法及其用于房颤信息的处理和分析, 计算了未房颤时 P 波的及房颤时 f 波的格子维和信息熵, 特别是, 比较了未房颤时和发生房颤时的相关计算结果的显著差别, 从而很容易判断是否发生了房颤。

[关键词] 复杂大系统; 非线性; 分形理论; 分形维; 信息熵 (Shannon 熵); 房颤

[中图分类号] TP391; R540.4⁺1 **[文献标识码]** A **[文章编号]** 1009-1742 (2003) 08-0045-05

1 引言

非线性科学是一门新兴的先进的热门学科, 是诸多非线性理论的总称, 其研究对象涉及复杂大系统。在这类系统中, 整体不能简单地归结为部分之和, 其各个部分之间的相互作用常常是非线性的。

传统的科学方法是通过将复杂的研究对象分解为若干个细小的组成部分进行研究。这种传统方法已广泛用于各个领域 (自然科学, 社会科学, 工程技术, 等等), 取得了丰硕成果。遗憾的是, 传统方法的固有弱点在于“只见树木, 不见森林”, 缺乏对所研究对象的全面了解和综合分析, 因而用于研究复杂大系统是无法胜任的。

人体正是一个复杂大系统, 其各部分之间的强耦合相互作用是非线性的, 因而非线性科学在医学领域大有用武之地。

笔者介绍分形和分形维、信息熵等非线性理论及其在医学中的应用和笔者的一些有关研究成果。

2 分形与分形维

2.1 分形概述

分形 (fractal) 是数学家 B.B.Mandelbrot 于

1975 年最先创用的。迄今为止, 分形尚无严格的数学定义, B. B. Mandelbrot 曾建议把分形定义为: 局部以某种方式与整体相似的、即具备自相似性的几何形体。分形几何理论首先出现在地貌、地表形态、地质结构、材料结构等研究领域, 随后在宇宙星系的研究等方面也得到了广泛应用。

这里所说的自相似性既包括严格自相似性, 也包括统计自相似性。属于前一类的有 Koch 曲线、Peano 曲线、Cantor 栅栏 (Cantor 集) 等。现实世界的几何体大多不具备严格的自相似性, 却可以有统计意义上的自相似性, 例如, 曲曲弯弯的海岸线便具有以 Koch 曲线为基本单元的统计自相似性。人和动物体内的肺、血管、胆管等的复杂分支结构也具有这种统计自相似性。

人体内的分形结构产生于胚胎发育过程, 与生命同在。例如, 健康人的心率即便在静止状态下也有很大波动, 心率的时序曲线呈现出忽高忽低的波浪状, 看起来似乎是完全随机的, 但如果观察较长时段的这种变化图像就会发现, 从一批心率数据中还是能找出其规律性。例如, 观察一段几个小时的心率时序图即可见到, 各个较短时段的心率时序图像都同整个时段的总图像具有统计自相似性。

[收稿日期] 2003-03-14; **修回日期** 2003-04-28

[基金项目] 国家自然科学基金资助项目 (10072006)

[作者简介] 涂承媛 (1963-), 女, 北京市人, 北京工业大学副教授, 博士研究生

维数是几何对象的重要基本特征量，分形理论中的这一特征量称为分形维（分形维数，分数维）。

2.2 分形维

在传统的 Euclid 几何学中，维数是一个很重要的概念。其涉及的 Euclid 维都是整数：点为零维，线为 1 维，面为 2 维，体（空间）为 3 维；在 Einstein 相对论中，则讨论了时 - 空 4 维世界；热力学所研究的 n 维相空间，其维数 n 可以是大于 4 的整数；至于科学技术研究中涉及的 n 维因素空间，影响事件进行过程的因素每增加一个，即相当于维数增 1，这儿的 n 也是整数。

以上所述，涉及的都是规则的几何图形。然而对于广泛的复杂几何图形，例如，对 Koch 曲线、Sierpinski 地毯等，整数维的概念已不能适合所有场合，必须将维数的概念加以拓展，使之可以具有分数值，即引入分形维（或称分维数），例如，Koch 曲线的分形维是 1.261 8，Sierpinski 地毯的分形维是 1.892 8，Vicsek 图形的分形维是 1.465 0，等等^[1]。

分形维的种类很多，例如，相似维，容量维（Kolmogorov 维），关联维等等^[2~4]，常用的前二者的定义如下：

1) 相似维 一般地，当把某个图形的长度（或标度）放大 $1/r$ 倍时（ $r > 1$ ），得到 $N(r)$ 个和原图形相似的小图形。如果 $N = r^{-D}$ ，则指数 D 就具有维数的意义，称为相似维^[5]，记作

$$D_s = \frac{\ln N(r)}{\ln (1/r)} \quad (1)$$

例如，将直线段、正方形、立方体的边二等分（即 $r = 1/2$ ），如图 1 所示，则这时该直线段、正方形、立方体可看成为分别由 $2^1, 2^2, 2^3$ 个严格相似的小图形组成。这里的相似维 1, 2, 3 分别与有关图形的 Euclid 维（即经验维）是一致的。

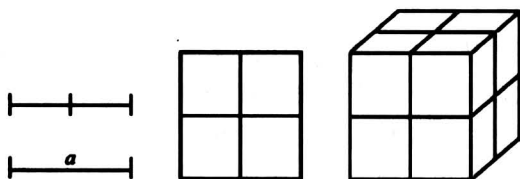


图 1 线、面、体的划分与其维数

Fig.1 Partition of a line, a square, a cubic and their dimension values

2) 容量维 设所要考虑的图形是 d 维 Euclid

空间 R^d 中的有界子集，并用半径为 ϵ 的 d 维小球来覆盖该图形，以 $N(\epsilon)$ 表示这类小球的个数的最小值，则容量维 D_c 为^[6]

$$D_c = \lim_{\epsilon \rightarrow 0} \frac{\ln N(\epsilon)}{\ln (1/\epsilon)} \quad (2)$$

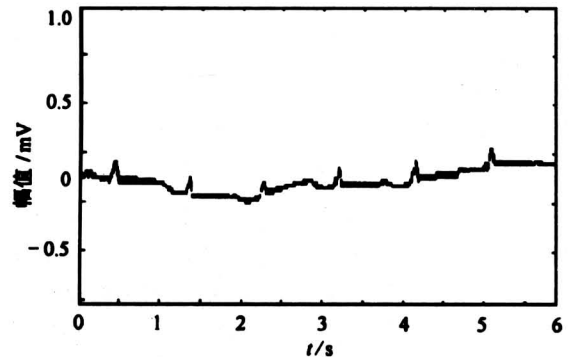
实际操作中，常将上述覆盖用的 d 维小球改为 d 维小盒，由式 (2) 得到的便是盒子维（也称盒维数）。

当研究 2 维平面上的图形时，用于覆盖的可以是 2 维小格子，这时由式 (2) 得到的是格子维 D_g 。

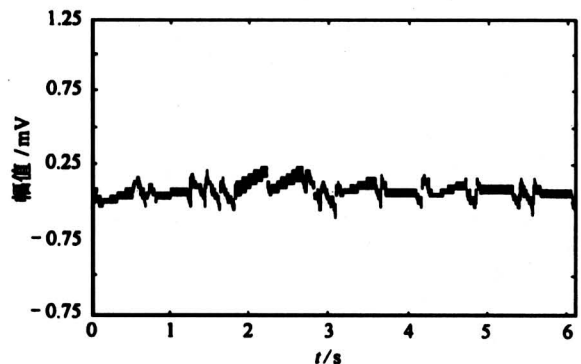
笔者采用简单有效的格子维对未房颤时和发生房颤时的 ECG 曲线进行划分，随之求得分形维。

3 房颤的分形维分析

ECG 的时域波形一般由 P 波、QRS - T 复合波等组成。房颤发生时，其中的 P 波演变为 f 波（房颤波）。利用频数直方图技术将 ECG 图像中的 QRS - T 波剔除掉，留下的便是代表心房活动的 P 波（未房颤时）或 f 波（房颤时）^[7]，如图 2 所示。



a. P 波



b. f 波

图 2 P 波和 f 波的图像

Fig.2 The curves of P wave and f wave

3.1 数据的归一化处理

图 2 所示的 P 波曲线、f 波曲线的自变量单位是秒 (s)，因变量单位是毫伏 (mV)，两者的量纲不同，于是，分别提取每一心动周期的 P 波数据、f 波数据，并对之进行归一化处理，使其转化为无量纲的物理量，而后求出分形维。

采用的归一化方法如下：

1) 时间值的归一化 以 T 表示一个心动周期的时间值 (s)，令均匀采样点的序号数 i 依次取值为 $0, 1, 2, \dots, N-1$ ，共 N 个采样点， $(N-1)$ 个采样间隔，各采样点的时间值 $t(i)$ 的计算式为

$$t(i) = \frac{iT}{N-1}, i = 0, 1, 2, \dots, N-1. \quad (3)$$

而 $t(i)$ 的归一化取值 $n(i)$ 为

$$n(i) = \frac{t(i)}{T} = \frac{i}{N-1}, i = 0, 1, 2, \dots, N-1. \quad (4)$$

2) 幅值的归一化 令该心动周期内的幅值 $y(i)$ 的最大值和最小值分别为 y_{\max} 和 y_{\min} ，则

$$y_{\max} = \max \{y(i)\}, i = 0, 1, 2, \dots, N-1,$$

$$y_{\min} = \min \{y(i)\}, i = 0, 1, 2, \dots, N-1. \quad (5)$$

于是，归一化幅值 $Y(i)$ 为

$$Y(i) = (y(i) - y_{\min}) / (y_{\max} - y_{\min}). \quad (6)$$

经此处理后， $Y(i), n(i)$ 都化为 $[0, 1]$ 区间内的无量纲值。

3.2 求分形维的方法

利用 3.1 节的方法，分别对每一心动周期的 P 波、f 波的相关数据进行归一化处理，得到归一化后的 P 波数据或 f 波数据，据此绘出归一化 P 波曲线、归一化 f 波曲线。

先讨论如何计算 P 波曲线的分形维。

P 波曲线是处于 2 维 Euclid 平面上，于是，用边长为 a 的小方格将归一化 P 波曲线所在的平面进行划分，以 $N(a)$ 表示包含该曲线的小方格数目。通过改变方格子的边长 a 的值，可得到一系列的数据对 $\{a, N(a)\}$ ，再据此算出相应的一系列的数据对 $\{\ln a, \ln N(a)\}$ ，并对之进行最小二乘法拟合，便得出

$$\ln N(a) = \ln k - D \ln a. \quad (7)$$

式 (7) 的指数形式为

$$N(a) = ka^{-D}. \quad (8)$$

显然， D 相当于维数，即所要求的分形维 (确切地说，是格子维)。

f 波的分形维的求法也是如此。

3.3 计算结果

房颤的分形维分析中所使用的 ECG 数据均来自 MIT/BIH 心电数据库，未发生房颤时的和房颤时的数据各半，共 32 例 ECG 曲线，每例各取其 50 个心动周期进行分析，即利用上述方法分别求出每例 ECG 曲线的 50 个 D 值，再算出这 50 个 D 的均值 D_{mean} 。

表 1 所列的 16 例是未房颤时的和 16 例发生房颤时的分维数均值 D_{mean} 。由表 1 可见，房颤时 f 波的 D_{mean} 值小于未房颤时 P 波的 D_{mean} 值。

表 1 P 波、f 波的分形维

Table 1 The fractal-dimension values of P waves and f waves

P 波				f 波			
编号	D_{mean}	编号	D_{mean}	编号	D_{mean}	编号	D_{mean}
1	1.840 2	9	1.843 6	1	1.673 4	9	1.653 4
2	1.852 4	10	1.872 4	2	1.668 9	10	1.672 6
3	1.852 8	11	1.833 9	3	1.687 5	11	1.661 5
4	1.848 9	12	1.881 7	4	1.691 8	12	1.690 3
5	1.868 4	13	1.846 2	5	1.631 3	13	1.651 7
6	1.885 8	14	1.864 3	6	1.690 9	14	1.673 1
7	1.869 1	15	1.855 0	7	1.676 6	15	1.632 9
8	1.839 4	16	1.843 1	8	1.696 6	16	1.671 1

在利用此 16 (例) \times 50 (周期) P 波数据和 16 (例) \times 50 (周期) f 波数据、即总共 $(16 + 16) \times 50 = 1\ 600$ 个数据进行房颤诊断时，若根据 P 波某个周期的数据算出的分形维 D 的值显示为房颤属性，则这样的结果为假阳性；而若根据 f 波某个周期的数据算出的分形维 D 的值显示为非房颤属性，则这样的结果为假阴性。相应于这 1 600 个数据，计算结果的分布情况见表 2。

表 2 算出的 D 值的分布

Table 2 The distribution of D values

数据总数	阳性数		阴性数		正确率* /%
	真阳性	假阳性	真阴性	假阴性	
1 600	646	121	679	154	82.8

* 正确率 = $(646 + 679) / 1\ 600$

4 信息熵^[8~12]

4.1 信息熵的概念

熵是描述复杂大系统的一个很有用的物理量，熵的大小是无序度的一种度量，适于用来研究随机

过程。从 Clausius 的热力学熵，到 Boltzmann 的统计熵，再到 Shannom 的信息熵，熵分析法已广泛应用于物理、化学、生物学、医学、经济学等众多自然科学和社会科学领域。

信息熵（又名 Shannom 熵）是 C.E. Shannom 于 1948 年在其论文“通信的数学理论（A mathematical theory of communication）”中提出来的，其定义为

$$H(p_j) = - \sum_{j=1}^N p_j \lg p_j, j = 1, 2, \dots, N, \quad (9)$$

式中， p_j 是第 j 个消息的出现概率， $H(p_j)$ 是信息熵（即信源发出的平均信息量）。

信息熵是分析随机过程的通信系统所含的信息量、衡量其不确定性（即随机性）的重要方法。

4.2 信息通信系统与医学

医学诊断治疗过程和护理过程实质上可以看成是对病人所患疾病进行相关信息的采集、传输、存储、处理的过程。

在临床医疗实践中，当医务人员首次接触病人时，即开始对病人的信息进行采集和处理。医生通过各种诊断手段，运用自身的专业知识和技能，对病人所患疾病进行分析综合，进行诊断，拟定治疗方案并实行之。护士通过各种途径对病人身体状况进行判断，做出护理诊断，依据医嘱进行各项护理活动，执行护理程序。整个医疗活动过程既有病人（信源）的生理和病理信息的发送过程，又有医生护理人员（信宿）对病人信息的接收过程，而具体的医疗活动行为可视为广义上的信息流通的信道。在抽象层次上，可以把临床诊断过程中的信息流视为一种通信系统，其框图如图 3 所示^[11,12]。

在这个系统中，信源信息 X 的发生并不是一次性单向传递过程，其总量是在医务人员主动行为的影响下形成的，实际是一个不断增加的值。或者说，它同时也是时间的函数，现实中表现为医生病案记录中所包含的病人全部信息，其中，既包括阳性结果，也包括阴性结果；既包括有效信息，也包括冗余信息。信宿收到的信息 Y 中既有信源信息 X （这是主体），也有噪声 Z ，必须从 Y 中将 Z 剔除，从而得到与 X 相应的有效信息，据以计算信息熵 H 。

4.3 房颤的信息熵分析

将信息熵理论用于房颤信息的处理和分析。

房颤的 ECG 检测过程是一种医学信息的流通

传递系统，如图 3 所示。图 3 中，信源发出的信息是生理、病理信号，而信宿收到的是从 ECG 信号获得的并剔除了传输噪声的 P 波、f 波信息。

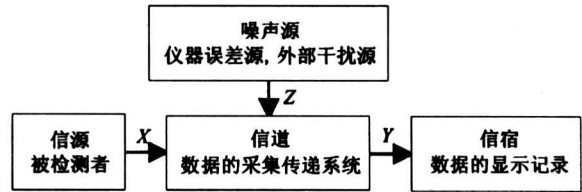


图 3 医学诊断过程的通信系统框图

Fig.3 The medical diagnosis as a communication system and its block diagram

在根据式 (9) 计算信息熵时，必须知道各个 p_j 的值 ($j = 1, 2, \dots, N$)。求 p_j 的方法是：由很窄的狭长条块将每一心动周期的、经过归一化处理的 P 波曲线和 f 波曲线予以分割，再根据各条块中含有的图形像素数 $d(j), j = 1, 2, \dots, N$ 来计算概率

$$p_j = \frac{d(j)}{\sum_j d(j)}, j = 1, 2, \dots, N. \quad (10)$$

对于 P 波、f 波，算得的信息熵如表 3 所示。

房颤的信息熵分析中所使用的 ECG 数据均来自 MIT/BIH 心电数据库，未发生房颤的和房颤的各 16 例。对每例 ECG 曲线，均取其 50 个心动周期，分别计算出每个周期的信息熵，得到 50 个 H 值，而后求其平均值 H_{mean} 。表 3 所列的是 16 例未房颤时的和 16 例发生房颤时的 H_{mean} 。由表 3 可见，房颤时 f 波的 H_{mean} 值显著大于未房颤时 P 波的 H_{mean} 。

相应于 1 600 个数据的计算结果见表 4。

表 3 P 波、f 波的信息熵

Table 3 The Shannom entropy of P waves and f waves

P 波				f 波			
编号	H_{mean}	编号	H_{mean}	编号	H_{mean}	编号	H_{mean}
1	5.736 2	9	5.695 5	1	8.355 7	9	8.056 9
2	6.177 6	10	5.901 1	2	8.284 3	10	8.212 6
3	5.702 0	11	5.952 2	3	8.183 3	11	8.278 0
4	6.156 4	12	6.063 9	4	8.197 8	12	8.097 6
5	5.969 7	13	5.966 4	5	8.370 7	13	8.338 1
6	6.077 9	14	6.059 7	6	8.409 5	14	8.457 3
7	5.966 4	15	5.829 0	7	8.432 5	15	8.376 6
8	6.059 7	16	5.898 9	8	8.208 5	16	8.224 4

表 4 算出的 H 值的分布Table 4 The distribution of H values

数据总数	阳性数		阴性数		正确率* /%
	真阳性	假阳性	真阴性	假阴性	
1 600	685	103	697	115	86.4

* 正确率 = (685 + 697) / 1 600

6 结语

利用分形维分析法对 P 波、f 波的信息进行了处理和分析，利用 2 维盒子维（即格子维）计算了 16 例未房颤时每—心动周期的（每例取 50 个周期）、16 例房颤时每—心动周期（每例也取 50 个周期）的分形维 D ，据以计算相应的 50 个 D 的均值 D_{mean} 。

未房颤时与房颤时的值之间的明显差异表明，用分形维分析很容易判断是否发生了房颤。

利用信息熵（shannon 熵）对 P 波、f 波的信息进行处理和分析，用上述计算 D_{mean} 的同样方法，可以得到相应的 32 个均值 H_{mean} 。

未房颤时与房颤时的 H_{mean} 值间的显著差异表明，用信息熵分析法很容易判断是否发生了房颤。

参考文献

- [1] 张济忠. 分形 [M]. 北京: 清华大学出版社, 1997. 80~94
- [2] Mannelqvist A, Groth M R. Comparison of fractal analyses methods and fractal dimension for pre-treated stainless steel surfaces and the correlation to adhesive joint strength [J]. Applied Physics A: Materials Science and Processing, 2001, 73(3): 47~355
- [3] Sivakumar B. Is a chaotic multi-fractal approach for rainfall possible? [J]. Hydrological Processes, 2001, 15(6): 943~955
- [4] Oscar C, Patricia M. New method for fuzzy estimation of the fractal dimension and its applications to time series analysis and pattern recognition [J]. Annual Conference of the North American Fuzzy Information Processing Society, NAFIPS, Sponsored by: NAFIPS IEEE, 2000, (Jul 13-Jul 15): 451~455
- [5] 汪富泉, 朱后强. 分形 [M], 济南: 山东教育出版社, 1996. 22~26
- [6] 高安秀树(日). 分维数 [M]. 北京: 地震出版社, 1989. 7~25
- [7] 涂承媛, 曾衍钧, 李树信. 心电图信号的频域分析 [J], 中国工程科学, 2002, 4(12): 66~70
- [8] 稽芸, 耿孝和, 吴布生. 临床诊断治疗过程信息熵测定原理研究 [J]. 系统工程理论与实践, 1999, 19(6): 102~106
- [9] 朱雪龙, 应用信息论基础 [M]. 北京: 清华大学出版社, 2001. 14~19
- [10] 童勤业, 刘刚强, 刘加友, 等. EEG 的信息熵分析 [J]. 生物物理学报, 1993, 9(4): 645~649
- [11] 尹殿云, 孙朝辉, 孟凡, 等. 信息熵在临床医学中的初步应用 [J]. 中国医院统计, 1997, 4(4): 227~228
- [12] 仪垂祥. 非线性科学及其在地学中的应用 [M]. 北京: 气象出版社, 1995. 53~5

Nonlinear Science and Its Medical Applications

Tu Chengyuan, Zeng Yanjun

(Biomedical Engineering Center, Beijing Polytechnic University, Beijing 100022, China)

[Abstract] The fractal-dimension theory and the Shannon entropy theory of nonlinear science are briefly described, and the methods to apply them to analyze AF's information are developed. The grid-dimension values and the Shannon entropy values of P-wave and those of f-wave are computed. The comparison of the remarkable difference between values of these two kinds of waves is given so as to easily determine whether AF happens or not.

[Key words] complicated and large system; nonlinear; fractal theory; fractal-dimension; Shannon entropy; AF(atrial fibrillation)