科学通报 2009 年 第 54 卷 第 3 期: 340 ~ 344
论 文
《中国科学》杂志社
SCIENCE IN CHINA PRESS

短时心率变异符号序列的熵分析方法

卞春华[①], 马千里[②], 司峻峰[①], 吴旭辉[①], 宁新宝[①], 王东进[③]

① 南京大学电子科学与工程系近代声学教育部重点实验室, 生物医学电子工程研究所, 南京 210093;
② 南京邮电大学地理与生物信息学院, 南京 210003;
③ 南京大学医学院附属鼓楼医院, 南京 210008
E-mail: bch@nju.edu.cn, maql@njupt.edu.cn

2008-09-18 收稿, 2008-10-23 接受
国家自然科学基金(批准号: 60501003, 60701002)和江苏省高校自然科学研究指导性计划(编号: 06KJD510138)资助项目

摘要 使用复杂性和非线性方法研究心率变异(heart rate variability, HRV)信号时序和结构中的规律, 对揭示心血管调节的潜在规律和生理本质具有重要的研究价值, 而研究适用于短时 HRV 信号的方法更具有重要的临床应用价值. 根据 HRV 信号方向信息的特点, 提出了符号序列的熵分析(sign series entropy analysis, SSEA)方法, 并将其应用于实际HRV信号的分析. 结果显示, SSEA 方法能从短时 HRV 信号中敏感地检测出生理和病理状态的变化, 并且受信号噪声和非平稳性的影响较小, 为临床 HRV 信号或其他复杂生理信号分析提供了一个有效的方法.

关键词
心率变异
复杂性
熵
老年化

心率变异(heart rate variability, HRV)指逐次心跳间期之间的微小时间差异, 其中蕴含着有关心血管调节的大量信息. 大量研究表明, 自主神经系统两支——交感神经和副交感神经与心率变异性之间存在着重要的联系: 当交感神经控制增强或副交感神经功能受损时, 心率变异性降低, 易促发心肌缺血、心功能紊乱等心脏疾病, 甚至导致死亡. 心率受诸多因素如运动、呼吸、血压、体温及精神紧张的影响而变化, 昼夜生物节律和肾素-血管紧张素系统等因素也参与调节. 这些因素使得HRV序列表现出高度的非平稳性和复杂性[1].

除了常规的线性分析手段外, 很多非线性动力学方法和复杂性测度也被用于HRV序列的分析, 这些研究对揭示心率变异和心血管调节的潜在规律和生理本质起到了促进作用. 主要方法有李亚普诺夫指数、关联维、多重分形和熵测度等[2~6]. 熵的概念最早出现于热力学, 具有宏观意义; 玻尔兹曼使熵具有分子水平的微观意义; 香农熵尽管是针对信息论提出, 但其具有广义熵的意义. 从广义上来讲, 玻尔兹曼热力学熵是香农信息熵的特殊情况. 熵不再局限于热力学和分子热运动, 而是不确定事件所普遍具有的性质. 信息熵被广泛用于描述信号的不规则性. 关于熵的研究一直是国际研究的前沿, 包含对其概念的理解和外延的拓展[7,8], 并被广泛应用于自然科学和社会科学研究各领域[9~11].

除了对原始时间序列进行分析外, 符号化是动力学系统研究的一个重要的手段. 其意义在于粗粒化掉一部分细节信息, 而保留感兴趣的部分. 通过不同程度的粗粒化, 舍去更小层次上的细节, 有利于突出更根本的行为. 适当的粗粒化, 有助于得出更严格的结论[12~15]. 其最大的优势在于对噪声的影响不敏感, 而其应用的关键部分在于如何划分合适的符号区域以适当地粗粒化信号. 文献[14]以整个时间序列的均值为基线, 划分几个符号区域, 其缺点在于易受信号非平稳性的影响. 而文献[6]则采用局部均值作为基线划分符号区域, 对文献[14]的方法有所改进, 但没有明确的生理意义, 并且符号区域划分的阈值参数根据经验确定, 没有客观的取值标准.

本文根据 HRV 信号的特点首次提出了用三种符号代表心率变异的方向, 并提出了适合其动力学分析的符号序列熵分析(sign series entropy analysis, SSEA)方法. 通过对国际标准数据库年轻、年老两组心率变异信号的分析, 测试所提出方法的有效性, 并将该方法应用于临床. 发现该方法不仅可以从短时

数据中有效检测出如老年化这样的生理状态的变化，也可以用于检测临床患者，如房颤患者(atrial fibrillation, AF)的病理状态所导致的 HRV 的改变.

1 HRV 符号序列的熵分析方法

对心率变异信号的研究，其研究对象主要是心跳间期时间序列. 心跳间期指连续两次心跳之间的时间间隔，一般通过自动检测算法检测连续采集的心电图(electrocardiogram, ECG)相邻 R 波的间隔获得，称为 RR 间期，我们将其记为 $RR(i)$，i=1,2,3,⋯N，表示第 i 次和第 i+1 次心跳之间的时间间隔.

心跳间期不是绝对均匀的，其变化有三种方式. 相应的，我们使用三种符号代表心跳间期(瞬时心率)的变化方向为

$$x(i)=\begin{cases}0, & RR(i+1)<RR(i),\\ 1, & RR(i+1)=RR(i),\\ 2, & RR(i+1)>RR(i),\end{cases}\quad i=1,2,3,\cdots N-1. \tag{1}$$

其中 $x(i)$=0 表示心跳间期减小，瞬时心率增大; $x(i)$=1 表示心跳间期(瞬时心率)基本持平；而 $x(i)$=2 则表示心跳间期增大，瞬时心率减小. 在本文的方法中，三种符号仅代表三种状态，其数值大小没有意义. 通过以上的符号化方法，心跳间期具体变化的幅度被粗粒化，而仅保留其方向信息. 具体符号化过程如图 1 所示.

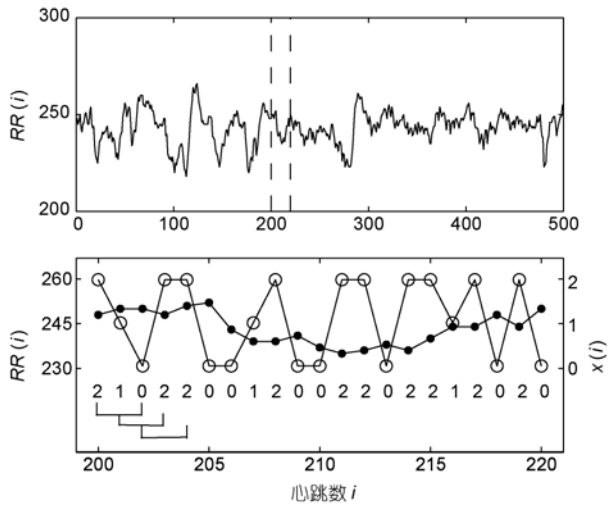

图 1 心跳间期符号化及使用滑动窗方法构造字(m=3)示意图

对于符号化的方向信号，为揭示其时序和结构的规律，使用滑动窗的方法构造字长宽度为 m 的向量序列:

$$X(i)=[x(i),x(i+1),\cdots x(i+(m-1))],\quad i=1,2,\cdots N-m. \tag{2}$$

图 1 给出了构造向量 $X(i)$的实例，向量 $X(i)$代表方向的连续变化情况. 当字长宽度为 m 时，连续变化共有 M=3^m 种可能的模式. 统计每种模式出现的概率(如图 2):

$$p_j=\frac{N_j}{N-m},\quad j=1,2,\cdots M, \tag{3}$$

其中N_j为第j种模式出现的次数. 计算信息熵[16]:

$$SSE(m)=-\sum_{j=1}^{M}p_j\log_2 p_j. \tag{4}$$

熵值大小代表了变化模式出现的平均不规则程度: 如果只出现一种模式，此时熵最小，值为 0; 如果出现的模式无特异性分布，即各种模式出现的概率相等，此时熵最大.

首先使用国际标准数据库测试SSEA分析方法. 测试使用的HRV信号取自MIT-BIH Fantasia数据库[17,18]，其中包括 40 个健康样本: 20 位年轻人(21~34 岁)和 20 位老年人(68~85 岁). 所有人在进行ECG信号采集时保持仰卧休息状态，窦性心率，观看Fantasia影片(Disney, 1940 出品)帮助保持清醒. ECG 信号采样频率为 250 Hz，采集时间为 120 min. 数据库中给出了从连续ECG信号自动检测得到的R波波峰的时间标注，从而获得心跳间期时间序列.

图 2 为字长宽度 m=2 时所有样本数据 9 种变化模式的统计分布. 可以看出老年组在增大和减小交替的两种模式上与年轻组接近(Mode = “02”, P = 0.80; Mode = “20”, P=0.82)；在连续减小(00)和连续增大

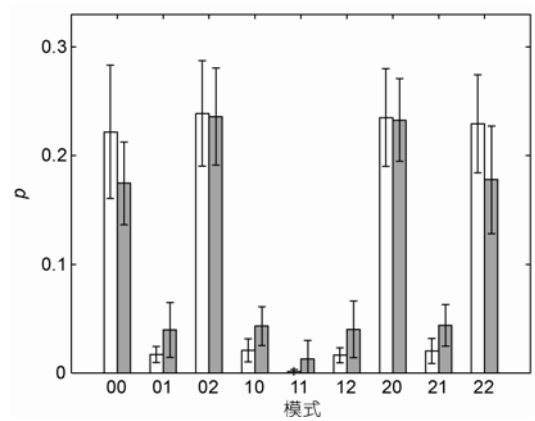

图 2 所有样本变化模式的统计分布

白色为年轻人，灰色为老年人；柱状图表示相应模式的概率均值，误差棒表示标准差

(22)两种模式上显著低于年轻人(P<0.05); 而在含有持平的几种模式(01, 10, 11, 12, 21)上则显著高于年轻人(P<0.01). 因此, 相对于老年人来说, 年轻人的符号变化模式有更明显的特异性分布, 在某些模式上相对更加集中; 而老年人则相对均匀.

使用 SSEA 计算信息熵以定量分析. 图 3 为 5 名年轻人和 5 名老年人连续 8 段数据长度为 500 点的计算结果, 横轴为样本序号, 即 1~5 名年轻人和 1~5 名老年人, 纵轴为各数据段的 SSEA 结果. 可以看出, 同一个对象相邻数据段的结果很接近, 说明该方法可以从很短的数据中获得有效的序列特征参数, 受信号非平稳性的影响很小.

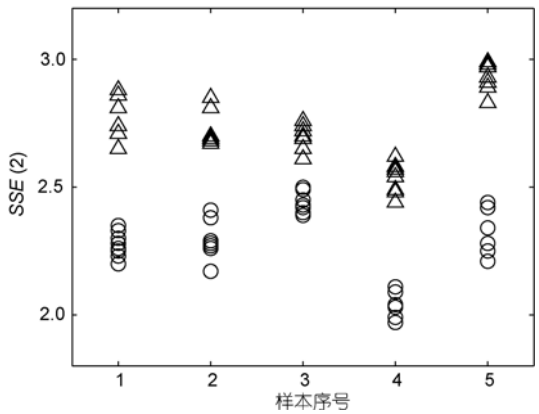

图 3 相邻连续短时数据段的计算结果比较

○为年轻人, △为老年人

将全部对象的结果用均值±标准差形式表示于图 4(a), 结果显示数据长度为 500 点时, SSEA 分析在实际应用表现出很好的敏感性, t 检验显示年轻人和老年人组间存在明显差异(P<0.001); 并且在增加数据长度为 1000, 2000, 4000 点进行长时分析时, 其结果和短时数据结果一致. 进一步说明 SSEA 方法可以有效地应用于短时数据分析, 具有很好的临床实用性.

图 4(b)显示了字长宽度 m 对于分析结果的影响, 数据长度 N=500. 可以看出字长宽度的增加会使 SSE 值增大, 但在 m=2~6 时, SSEA 方法均能有效地区分年轻组和老年组的差异(P<0.001).

为研究噪声以及 R 波波峰检测误差对于 SSEA 分析结果的影响, 我们在 RR 间期上增加随机白噪声, 幅度为 2~10 个采样间隔(对应于 250 Hz 的 ECG 采样频率, 为 8~40 ms, 一般误差不会超出这一范围). 对应于每个噪声水平进行了 100 次模拟实验. 图 5 为一个年轻人样本和老年人样本模拟实验的统计结果.

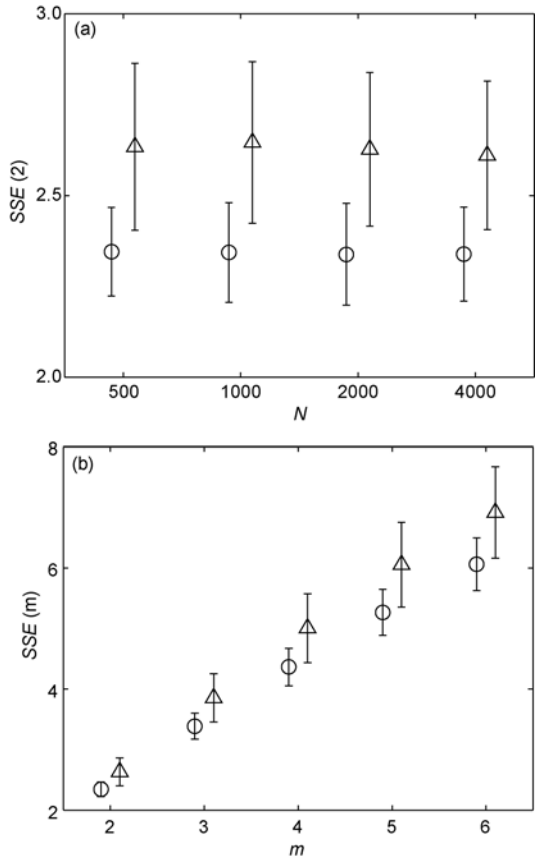

图 4 数据长度 N 与字长 m 对 SSEA 分析的影响

(a) m=2 时, 不同数据长度的 SSE 统计结果(均值±标准差); (b) m=2~6 时, SSE 统计结果比较. ○为年轻人, △为老年人

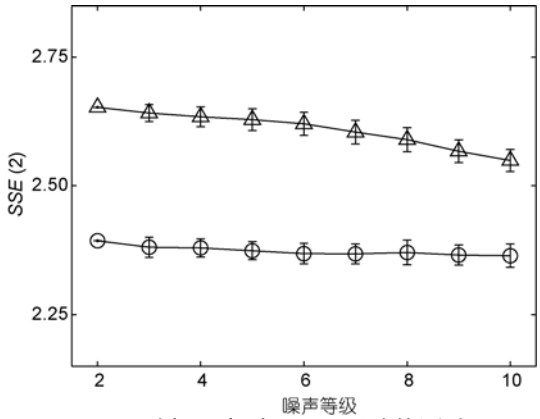

图 5 随机噪声对 SSEA 影响的测试

○为年轻人, △为老年人

可以看出, 白噪声的增加对于结果没有显著的影响.

2 临床试验

在测试了方法的有效性后, 我们进行了临床试验. 实验对象包括健康人 36 例, 无心脏病史, 血压正

常; 瓣膜性心脏病房颤患者 10 例. 实验装置采用自行研制的多功能心电工作站. 所有实验对象在数据采集时保持平躺休息状态. ECG 采样频率与 Fantasia 库一致, 为 250 Hz, 采集 500 个心跳周期, 时间约 6~7 min. HRV 信号从采集的 ECG 信号自动提取.

使用本文提出的 SSEA 方法对临床采集的健康人和房颤病人的 HRV 信号进行分析, 并与 Fantasia 库中的年轻与老年组比较, 结果如图 6 所示(字长 m = 2, 序列长度 N=500). 从图中可以看出, 健康人的 SSE 随着老年化呈明显的上升趋势, 并且和 Fantasia 数据库结果一致. 将健康人(包括 Fantasia 数据库)的 SSE 随年龄的变化进行线性回归分析, 模型采用 $SSE(2) = p1\times$年龄$+p2$. 分析结果为: $p1$=0.006578, $p2$ = 2.225; 相关系数(adjusted R-square)为 0.4202.

此外, 房颤病人的 SSE 在各个年龄段都显著低于健康人(P<0.001). 该结果显示, SSEA 方法不仅可以用于评价老年化对于 HRV 的影响, 也可以用于临床疾病的诊断.

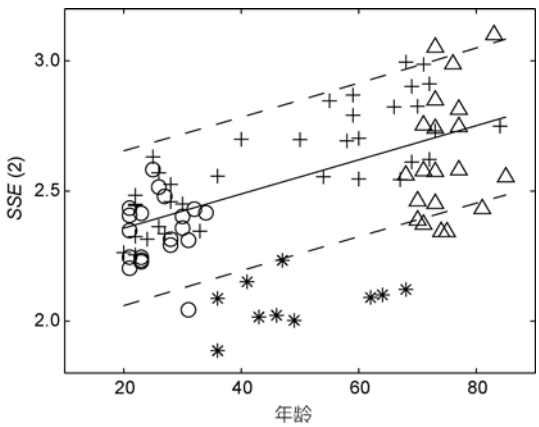

图 6 健康人与房颤病人 SSE 的年龄分布

○为 Fantasia 数据库年轻人, △为 Fantasia 数据库老年人, +为自行采集的健康人, *为自行采集的房颤病人; 实线为健康人(包括 Fantasia 数据库样本)SSE 参数随年龄变化的线性拟合结果, 虚线为其 95%可信区间

3 总结与讨论

由于传统时域分析方法不考虑时序和结构信息, 所以复杂性和非线性方法在心率变异信号的研究中受到了普遍的重视. 而传统非线性参数如关联维和李氏指数计算所需数据量大[19], 当应用到HRV信号分析时信号采集需要一小时甚至更长时间, 对于临床使用很不方便, 因此短时分析方法的研究受到了高度重视. 本文提出了适用于短时HRV信号分析的符号序列熵分析方法SSEA. 使用三个符号表示心跳间期的变化(减小、持平或增大), 并构造一定长度的字以反映其时序与结构, 通过统计各种模式的概率分布并计算香农熵量化各种模式出现的特异性, 进而表征生理与病理状态的改变.

我们将 SSEA 应用于实际 HRV 信号的分析, 包括长时的国际标准 MIT-BIH Fantasia 数据库和临床采集的短时 HRV 数据. 研究发现, 老年人心率变异连续减小(00)和连续增大(22)两种模式都显著低于年轻人(如图 2). 从生理上来说, 人体各种反射机制产生的信息都是通过交感神经和副交感神经这两条通路传输到心脏的, 而人体神经自律能力随年龄老化而逐渐下降, 从而导致其对心脏的控制能力下降. 值得注意的是, 瓣膜性房颤患者使用本方法也可以从同年龄段正常人中正确分出, 其隐含的生理意义有进一步深入研究的价值.

在文献[20]中提出老年化使得HRV信号的分形维数(fractal dimension, FD)下降, 并计算出FD 与年龄的相关系数为 0.34, 低于SSE与年龄的相关系数(0.4202), 说明SSEA在评价老年化时比FD更 敏感.

近似熵(approximate entropy, ApEn)是近些年信号复杂度分析使用较为广泛的方法. ApEn反映当维数增加时产生新模式可能性的大小, 愈大, 说明产生新模式的机会愈大, 因此该信号的不规则性越强[7]. 理论上近似熵值的计算与数据长度无关, 但是由于HRV信号的高度不平稳性和干扰噪声的存在, 导致实际应用HRV序列分析中往往不能获得可靠的结果. 图 7

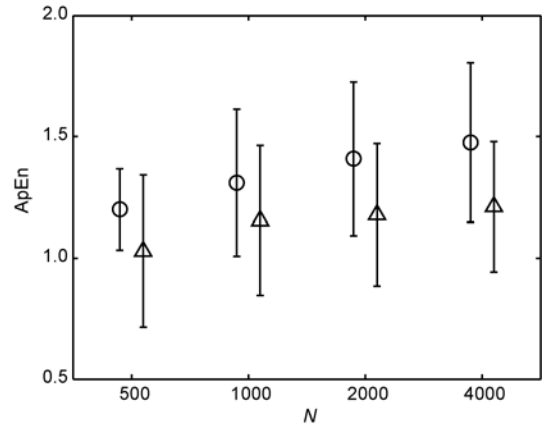

图 7 HRV 信号近似熵分析

t 检验的结果为: N=500, P=0.056; N=1000, P=0.087; N=2000, P=0.093; N=4000, P<0.01

为使用近似熵对 Fantasia 数据库中年轻人和老年人 HRV 序列分析的结果. t 检验结果显示, 当数据长度为 N=500, 1000, 2000 时近似熵均不能区分出年轻人和老年人组间的差异; 增加数据长度为 4000 点时才可以有效区分出组间差异. 而 SSEA 并不受数据长度的影响, 短时数据即可给出准确的结果(见图 4(a)).

SSEA 方法在实际应用中表现出很好的敏感性和可靠性, 能从短时的(500 点, 6~7 min)HRV 信号中敏感地检测出生理和病理状态的变化, 并且受信号非平稳性和噪声的影响较小. 这些特点决定了其具有一定的临床实用价值, 并且也可以将其应用于其他生理信号, 或其他领域复杂信号的分析.

致谢 衷心感谢南京大学医学院附属鼓楼医院胸心外科邵俊医生及南京大学范爱华同学在临床数据采集中所做的工作和给予的有益讨论!

参考文献

1 Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability, standards of measurement, physiological interpretation, and clinical use. Circulation, 1996, 93(5): 1043—1065

2 Bogaert C, Beckers F, Ramaekers, et al. Analysis of heart rate variability with correlation dimension method in a normal population and in heart transplant patients. Auton Neurosci, 2001, 90: 142—147[doi]

3 Bian C H, Ning X B. Evaluating age-related loss of nonlinearity degree in short-term heartbeat series by optimum modeling dimension. Physica A, 2004, 337(1-2): 149—153[doi]

4 卞春华, 宁新宝. 短时心率变异信号非线性程度的研究. 科学通报, 2004, 49(2): 130—134

5 Ivanov P Ch, Amaral L A N, Goldberger A L, et al. Multifractality in human heartbeat dynamics. Nature, 1999, 399: 461—465[doi]

6 李锦, 宁新宝. 短时心率变异性信号的基本尺度熵分析. 科学通报, 2005, 50(14): 1438—1441

7 Pincus S M. Approximate entropy as a measure of system complexity. Proc Natl Acad Sci USA, 1991, 88: 2297—2301[doi]

8 Staniek M, Lehnertz K. Symbolic transfer entropy. Phys Rev Lett, 2008, 100: 158101[doi]

9 Costa M, Goldberger A L, Peng C K. Multiscale entropy analysis of biological signals. Phys Rev E, 2005, 71: 021906[doi]

10 Martins J F, Santos P J, Pires A J, et al. Entropy-based choice of a neural network drive model. IEEE Trans Ind Electr, 2007, 54(1): 110—116[doi]

11 Pincus S M. Approximate entropy as an irregularity measure for financial data. Econom Rev, 2008, 27(4): 329—362[doi]

12 郝柏林. 复杂性的刻画与复杂性科学. 科学, 1999, 51(3): 3—8

13 Daw C S, Finney C E A, Tracy E R. A review of symbolic analysis of experimental data. Rev Sci Instrum, 2003, 74: 915—930[doi]

14 Wessel N, Ziehmann C, Kurths J, et al. Short-term forecasting of life-threaten[doi]

15 Guzzetti S, Borroni E, Garbelli P E. Symbolic dynamics of heart rate variability: A probe to investigate cardiac autonomic modulation. Circulation, 2005, 112: 465—470[doi]

16 Shannon C E. A mathematical theory of communication. Bell Syst Tech, 1948, 27(3): 379—423

17 Goldberger A L, Amaral L A N, Glass L, et al. PhysioBank, physioToolkit, and physioNet: Components of a new research resource for complex physiologic signals. Circulation, 2000, 101(23): e215—e220

18 Iyengar N, Peng C K, Morin R, et al. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. Am J Physiol 1996, 271: 1078—1084

19 Eckman J P, Ruelle D. Fundamental limitations for estimation dimensions and Lyapunov exponents in dynamical systems. Physica D, 1992, 56: 185—187[doi]

20 Beckers F, Verheyden B, Aubert A E. Aging and nonlinear heart rate control in a healthy population. Am J Physiol Heart Circ Physiol, 2006, 290: 2560—2570[doi]