

多尺度多变量模糊熵分析*

李鹏¹⁾ 刘澄玉¹⁾ 李丽萍²⁾ 纪丽珍¹⁾ 于守元¹⁾ 刘常春^{1)†}

1) (山东大学控制科学与工程学院, 济南 250061)

2) (山东中医药大学理工学院, 济南 250355)

(2013年2月1日收到; 2013年2月24日收到修改稿)

多尺度多变量样本熵评价同步多通道数据的多变量复杂度, 是非线性动态相互关系的一种反映, 但其统计稳定性差, 且不适用于非线性非平稳信号. 研究利用模糊隶属度函数代替模式相似判断的硬阈值准则, 并分析模糊隶属度函数形式的影响; 研究利用多变量经验模态分解算法进行多尺度化, 并对比其处理效果. 仿真试验表明, 模糊隶属度函数的引入可以有效提高算法的统计稳定性, 所构造的物理模糊隶属度函数的性能最为显著; 基于多变量经验模态分解算法的多尺度化过程可更有效地捕获信号的不同尺度成分, 从而更敏感地区分具有不同复杂度的信号. 对临床试验数据的分析支持以上结论, 且结果提示随着年龄增加或心脏疾病的发生, 心率变异性 and 心脏舒张间期变异性的多变量复杂度以不同的方式降低: 年龄增加会使低尺度熵值降低, 表示近程相关性的丢失; 而心脏疾病会同时影响各个尺度的熵值, 即同时丢失了近程和长时相关性. 该结论可用于指导心血管疾病的无创预警研究.

关键词: 多变量复杂度, 多尺度多变量模糊熵, 物理模糊隶属度函数, 多变量经验模态分解

PACS: 05.45.Tp, 05.45.-a, 89.75.Da

DOI: 10.7498/aps.62.120512

1 引言

随着多传感测量技术的发展, 对同步观测的多通道数据序列内或序列间动态相互关系的评价日益受到研究者的重视. 但目前学界对这种动态相互关系并没有严格意义上的定义, 面对不同的应用背景, 研究者提出了多种评价方法. 2011年 Ahmed 等^[1] 基于传统的单变量复杂度评估思想^[2,3] 和多维嵌入重构理论^[4], 提出了多变量样本熵 (multivariate sample entropy, MSampEn) 算法. MSampEn 可认为是传统单变量样本熵 (sample entropy, SampEn)^[3] 的一般形式, 除了对多通道数据序列中 (within channel) 每一个序列自身的复杂度, 即序列内模式的多样性或不可预知性作出评估之外, MSampEn 同时考虑了多个通道之间互预测性, 因此由 MSampEn 所评价的动态相互关系又被称为多变量复杂度 (multivariate complexity). 与 SampEn 类似, 多变量随机序列具有大于多变量混沌序列的

MSampEn 值, 而学界普遍认为后者应具有较高的复杂度 (因此具有较高的 MSampEn 值). 为克服这种矛盾, Ahmed 等借鉴了 Costa 等^[5] 提出的多尺度熵 (multiscale entropy, MSE) 思想, 对 MSampEn 进行了多尺度扩展, 即多尺度多变量样本熵 (multiscale multivariate sample entropy, MMSE). 多变量随机序列的 MMSE 值在高尺度上急剧下降, 而多变量混沌序列在高尺度上仍具有很高的 MMSE 结果. 作为时间序列分析的一种手段, MMSE 从复杂度、互预测性以及长时相关性角度评价了多通道时间序列的动态相互关系的一个方面, 从而为研究者展现出其内在的非线性耦合特征. 自其提出以来, MMSE 已在物理、生理等学科多种领域中获得应用, 比如三维风速数据分析^[1]、多导联脑电数据分析^[6]、中心血压序列分析^[7] 以及心跳和呼吸序列分析^[1,8,9] 等, 显示了其潜在的理论和应用价值.

作为 SampEn 的多变量扩展形式, MSampEn 仍然存在其固有缺陷, 主要包括短序列分析时的统计稳定性较差以及对阈值参数 r 的依赖性较

* 国家自然科学基金青年科学基金 (批准号: 61201049)、山东大学研究生自主创新基金 (批准号: yzc12082)、山东省优秀中青年科学家科研奖励基金 (批准号: BS2012DX019) 和中国博士后科学基金 (批准号: 2013M530323) 资助的课题.

† 通讯作者. E-mail: changchunliu@sdu.edu.cn, lskyp@mail.sdu.edu.cn

强等. Chen 等^[10]对 SampEn 的研究表明,造成这种缺陷的本质原因在于其评估两个模式是否相似时采用了硬阈值判据,即 Heaviside 函数 (Heaviside function, HF),当两个模式的距离在阈值参数 r 周围出现微弱变动时,其判断结果可能完全相反,从而引起较差的统计稳定性.举例来说,对相空间中的三个模式: $\mathbf{X}(i)$, $\mathbf{X}(m)$ 和 $\mathbf{X}(n)$, 以及一个任意小的正数 δ ($\delta \ll r$), 不失一般性,假设 $d\{\mathbf{X}(i), \mathbf{X}(m)\} = r - \delta$, $d\{\mathbf{X}(i), \mathbf{X}(n)\} = r + \delta$, 其中 d 表示两个模式之间的距离,那么按照 MSampEn 或者 SampEn 的判断准则, $\mathbf{X}(m)$ 会被认为是与 $\mathbf{X}(i)$ 相似的模式,而 $\mathbf{X}(n)$ 与 $\mathbf{X}(i)$ 不相似,但实际上 $d\{\mathbf{X}(m), \mathbf{X}(n)\} = 2\delta \ll r$, 即相比 $\mathbf{X}(i)$ 而言, $\mathbf{X}(n)$ 在距离上与 $\mathbf{X}(m)$ 更接近,这就是上述硬阈值判据带来的不稳定现象.为此,Chen 等^[10]利用模糊隶属度函数代替硬阈值判据,提出了模糊熵 (fuzzy entropy, FuzzyEn) 算法,该思想已被后续很多研究者所承袭^[11-14]. MSampEn 的缺陷也有望通过引入模糊隶属度函数而得到改善,但这些研究全部使用高斯分布函数作为模糊隶属度函数,当距离大于 0 时,其隶属度按指数规律平滑下降,而实际的物理记录由于不可避免地受到系统扰动或者采样误差的影响,理论上完全等同 (距离为 0) 的两个模式,可能会出现微小的距离增大,从而造成隶属度值小于 1,因此以高斯分布函数为代表的光滑函数族存在不能完全反映实际物理现象相互隶属关系的可能,隶属度函数的形式仍需进一步研究.

与 MSE 类似, MMSE 的多尺度化是一种低通滤波和降采样操作^[15],低通滤波采用滑动平均器,即通过矩形窗对原序列进行平滑操作,而有研究证实通过这种方式滤波后的序列在进行降采样时无法满足采样定理的要求,降采样操作会引入频率折叠,另外矩形窗具有很大的边瓣,造成高尺度序列存在很多虚假振荡成分 (详细分析可以参考文献 [16]),从而影响熵的计算.为此,Valencia 等^[16]通过 Butterworth 滤波器组来实现多尺度化从而避免上述问题,但不论是滑动平均还是 Butterworth 滤波,都是基于平稳信号处理的傅里叶变换思想,而实际物理或生理信号多数是非平稳非线性的,上述尺度化方法并不能有效地处理此类问题.在多尺度化操作上,Huang 等^[17]提出的经验模态分解 (empirical mode decomposition, EMD) 是一种完全数据自适应的非线性非平稳信号处理算法,通过 EMD 分解,可获得一个信号序列的多个振荡模式,

被称为本征模函数 (intrinsic mode function, IMF). Amoud 等^[18]基于 EMD 多尺度化法提出了本征模多尺度熵,提高了 MSE 刻画非平稳信号不同尺度的能力.但 EMD 算法处理多变量问题时,视每一个序列为一项独立元素,单独对每一个序列进行分解,往往出现同一尺度上的模式不匹配问题.2010 年,Rehman 等^[19]提出的多变量经验模态分解 (multivariate empirical mode decomposition, MEMD) 算法对此进行了更正,此后, Hu 等^[20]和 Ahmed 等^[21]先后利用 MEMD 对多变量数据进行多尺度化,从而获得每个序列的 MSE 以及多变量序列的 MMSE,证实了基于 MEMD 的多尺度化方法在刻画非平稳信号尺度上表现出良好的性能.因此基于 MEMD 多尺度化的多变量熵分析有可能提高其分析非平稳信号的能力,解决上述传统多尺度化过程存在的问题.这是本文研究的第二个目的.

为此,本文进行如下研究:类比传感器的灵敏阈静态特性,提出物理模糊隶属度函数 (physical fuzzy membership function, PFMF) 的概念,并以此为基础定义多变量模糊熵 (multivariate fuzzy entropy, MFuzzyEn).同时,基于 MEMD 多尺度框架理论,定义 MFuzzyEn 的多尺度扩展形式,即多尺度多变量模糊熵 (multiscale multivariate fuzzy entropy, MMFE),并通过仿真序列验证其有效性和优势.最后,通过多通道生理信号变异性序列的多变量复杂度分析,验证其应用价值.

2 多尺度多变量模糊熵算法

2.1 物理模糊隶属度函数

下式所描述的降半高斯分布函数可见诸 FuzzyEn 的相关研究^[12,13]:

$$A(x) = e^{-\ln(2)(x/r)^2} \quad (x \geq 0). \quad (1)$$

相比 HF 判据函数而言,当距离参数 $x > 0$ 时,降半高斯分布函数按对数规律平滑下降.实际上,由于序列不可避免地受到扰动、采样误差等的影响,本身完全一致的两个模式 (距离为 0) 可能会出现微小的距离增大,若直接采用 (1) 式,其隶属度值会小于 1,这种误差会对熵的计算带来干扰.因此,类比传感器的灵敏阈静态特性,本文认为模糊隶属度函数应具有以下特性.

1) $x \leq r$ 为“不应”范围,该距离完全是由于系统扰动或采样误差而造成的干扰,隶属度函数不作

出反应, 其值为固定值 1.

2) $x > r$ 为“反应”范围, 该距离包含了系统扰动和采样误差, 同时也可能包含了本身就存在的距离, 因此, 隶属度值随 x 的增加按对数规律平滑下降.

由此, 本文提出 PFMF 的概念, 并称由 (1) 式描述的函数形式为理想模糊隶属度函数 (ideal fuzzy membership function, IFMF).

定义 1 物理模糊隶属度函数

令域 X 表示一个距离集合, 对 $\forall x \in X$ 且 $x \geq 0$, $\exists A(x) \in [0, 1]$, 称 $A(x)$ 为域 X 上的隶属度函数, 并称由 (2) 式描述的函数形式为物理模糊隶属度函数:

$$A(x) = \begin{cases} 1 & (0 \leq x \leq r) \\ e^{-\ln(2)(\frac{x-r}{r})^2} & (x > r) \end{cases} \quad (2)$$

当 $r = 0.2$ 时, HF 判据函数、IFMF 函数和 PFMF 函数如图 1 所示.

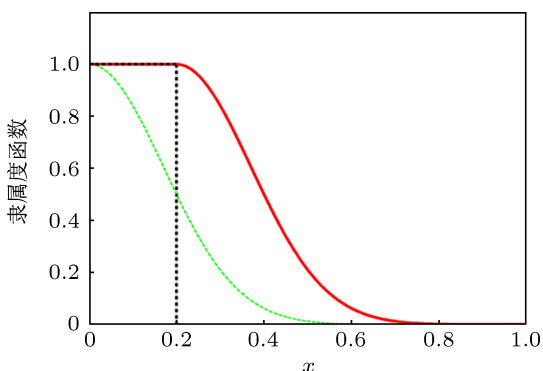


图 1 相似性判断的 HF 判据函数 (点划线)、IFMF 函数 (虚线) 和 PFMF 函数 (实线)

2.2 多变量模糊熵算法

本文利用 (1) 式或 (2) 式所描述的模糊隶属度函数代替 MSampEn 算法中的 HF 函数, 从而定义 MFuzzyEn 算法如下.

定义 2 多变量模糊熵

1) 对归一化 p 变量时间序列 $\{x_{k,i}\}_{i=1}^N$, $k = 1, 2, \dots, p$, 进行多维时延嵌入重构, 从而获得复合延迟向量:

$$\begin{aligned} \mathbf{X}_m(i) &= [x_{1,i}, x_{1,i+\tau_1}, \dots, x_{1,i+(m_1-1)\tau_1}, x_{2,i}, \\ &\quad x_{2,i+\tau_2}, \dots, x_{2,i+(m_2-1)\tau_2}, \dots, \\ &\quad x_{p,i}, x_{p,i+\tau_p}, \dots, x_{p,i+(m_p-1)\tau_p}] \\ &= [z_i, z_{i+1}, \dots, z_{i+m-1}] \\ &\quad (i = 1, 2, \dots, N-n), \end{aligned} \quad (3)$$

其中, $\mathbf{M} = [m_1, m_2, \dots, m_p] \in \mathbb{R}^p$ 为嵌入维数向量, $\boldsymbol{\tau} = [\tau_1, \tau_2, \dots, \tau_p]$ 为时间间隔向量, $m = \sum_{k=1}^p m_k$, $n = \max\{\mathbf{M}\} \times \max\{\boldsymbol{\tau}\}$.

2) 定义复合延迟向量中两个向量 $\mathbf{X}_m(i)$ 和 $\mathbf{X}_m(j)$ 之间的距离为

$$d[\mathbf{X}_m(i), \mathbf{X}_m(j)] = \max_{l=1}^m \{|z_{i+l-1} - z_{j+l-1}|\}. \quad (4)$$

3) 记 $d_{ij}^m = d[\mathbf{X}_m(i), \mathbf{X}_m(j)]$, 定义 $B_i^m(r) = \frac{1}{N-n-1} \sum_{j=1, j \neq i}^{N-n} A(d_{ij}^m)$, 其中 $A(x)$ 为由 (1) 或 (2) 式定义的模糊隶属度函数, 定义 $B^m(r) = \frac{1}{N-n} \sum_{i=1}^{N-n} B_i^m(r)$.

4) 将步骤 1) 中嵌入重构维度 m 扩展为 $m+1$, 由于包含有 p 个序列, 这种扩展可分别通过 $(m_k+1)_{k=1,2,\dots,p}$ 来完成, 因此可获得 $p \times (N-n)$ 个复合延迟向量 $\mathbf{X}_{m+1}(i)$ (对扩展方式的具体描述可参考文献 [1]).

5) 记 $d_{ij}^{m+1} = d[\mathbf{X}_{m+1}(i), \mathbf{X}_{m+1}(j)]$, 定义 $B_i^{m+1}(r) = \frac{1}{p(N-n)-1} \sum_{j=1, j \neq i}^{p(N-n)} A(d_{ij}^{m+1})$, 其中 $A(x)$ 为由 (1) 或 (2) 式定义的模糊隶属度函数, 定义 $B^{m+1}(r) = \frac{1}{p(N-n)} \sum_{i=1}^{p(N-n)} B_i^{m+1}(r)$.

6) 定义 MFuzzyEn 为

$$\text{MFuzzyEn}(\mathbf{M}, \boldsymbol{\tau}, r, N) = -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right]. \quad (5)$$

2.3 多尺度多变量模糊熵算法

为克服传统多尺度化方法不适用于处理非平稳信号的缺陷, 本文首先通过 MEMD 分解获得多变量序列的多个 IMF 函数, 由于 IMF 函数本身为窄带信号 [17], 类比传统多尺度化方法逐步去除高频成分的过程可知, 从低频到高频 IMF 函数的逐步累加, 或从原始序列中逐步去除高频 IMF 函数, 则可实现多变量序列的多尺度化. 由此, 本文定义 MMFE 算法如下.

定义 3 多尺度多变量模糊熵

1) 对归一化 p 变量时间序列 $\{x_{k,i}\}_{i=1}^N$ ($k = 1, 2, \dots, p$) 进行 MEMD 分解, 获得 p 变量 IMF 函数 (趋势项视为最后一个 IMF), 记每一序列的 IMF 数目为 Φ , 则 p 变量 IMF 函数可表示为 $\{c_{k,i}^\varphi\}_{i=1}^N$ ($k = 1, 2, \dots, p, \varphi = 1, 2, \dots, \Phi$).

2) 多变量多尺度化, 取每个序列的前 Ψ ($\Psi \leq \Phi$) 个 IMF 函数进行, 则其中第 ε 尺度上的序列可表示为 (若 $\Psi = \Phi$, 表示完全尺度化, 即第一尺度的序列为原始序列, 若 $\Psi < \Phi$, 则舍弃部分低频 IMF 分量, 这些分量一般被认为是非平稳趋势):

$$y_{k,i}^{\varepsilon} = \sum_{\varphi=\varepsilon}^{\Psi} c_{k,i}^{\varphi}. \quad (6)$$

3) 对 p 变量第 ε 尺度上的序列 $y_{k,i}^{\varepsilon}$, 按定义 2 计算 MFuzzyEn.

4) 定义 $1-\Psi$ 尺度上的 MMFE 为

$$\text{MMFE}(\varepsilon) = \text{MFuzzyEn} \{y_{k,i}^{\varepsilon}\}. \quad (7)$$

3 仿真结果与性能分析

单变量 MSE 在应用到白噪声序列时, 其熵值随尺度的增大单调递减, 而应用于 $1/f$ 噪声序列时, 高尺度上的熵值逐渐趋于稳定, 并且大于白噪声序列的结果 [5]. 这与研究者所普遍接受的 $1/f$ 噪声序列相比白噪声而言具有更为复杂的结构一致. 同样, 对于多通道数据而言, 当含有 $1/f$ 噪声序列的通道数增多时, 其多变量复杂度增大, 使用 MMSE 的仿真结果支持该结论 [1].

为对比分析 MMSE 和 MMFE 的不同特性, 本文使用三变量数据序列进行仿真验证. 根据三通道数据中含有的白噪声和 $1/f$ 噪声组合情况分为四组: (A) 所有通道为 $1/f$ 噪声; (B) 两个通道为 $1/f$ 噪声, 另外一个通道为白噪声; (C) 一个通道为 $1/f$ 噪声, 另外两个通道为白噪声; (D) 所有通道为白噪声. 其中, $1/f$ 噪声序列的产生可参考文献 [5]. 在频域上看, 其频谱密度与频率 f 成反比, 介于白噪声 (与频率无关) 和布朗噪声 (与 f^2 成反比) 之间, 局部无序而宏观上具有一定的相关性, 或称之为长时相关性 (long-range correlations), 而长时相关性被认为是诸如人类安静状态下脑电 α 波以及心率变异性 (heart rate variability, HRV) 序列等复杂序列所具有的特质. 另外, 以上所有序列长度均为 4000 点.

为同步对比 IFMF 与 PFMF 的不同效果, 以及基于 MEMD 增强非平稳处理能力的多尺度化过程与传统多尺度化过程的分析效果, 本文对以上数据进行分析时, 分别计算如下六种多变量多尺度熵: (A) MMSE; (B) 基于 IFMF 函数的 MMFE (i-MMFE); (C) 基于 PFMF 函数的 MMFE (p-MMFE);

(D) 基于 MEMD 多尺度化的 MMSE (m-MMSE); (E) 基于 MEMD 多尺度化和 IFMF 函数的 MMFE (mi-MMFE); (F) 基于 MEMD 多尺度化和基于 PFMF 函数的 MMFE (mp-MMFE). 因为是三通道信号, 所以嵌入维数向量设置为 $M = [2, 2, 2]$, 时间间隔向量设置为 $\tau = [1, 1, 1]$, 取阈值参数 $r = 0.45$, 传统多尺度化过程中取尺度为 1—20, 而基于 MEMD 过程的多尺度化尺度数取决于 MEMD 自身, 仿真试验中依次去除前 8 个 IMF 分量, 即 $\Psi = 8$. 试验结果如图 2 所示.

由图 2 可知, 随含有 $1/f$ 噪声通道数的增多, 六种类型的多变量多尺度熵在高尺度上增大, 预示着系统多变量复杂度的增大, 这与设想一致. 但六种多尺度多变量熵的性能又具有明显的差别, 具体分析如下.

由图 2(a) 可见, MMSE 处理白噪声比重较大的三变量序列时统计稳定性可以接受 (序列组 (C) 和 (D) 的误差线较短), 但处理 $1/f$ 噪声比重较大时具有最差的统计稳定性 (序列组 (A) 和 (B) 的误差线很长且已经重叠), 表明 MMSE 并不适于处理具有长时相关性的复杂序列. 另外, 图 2(a) 与文献 [1] 中图 2 的区别仅在于后者在仿真实验中使用的序列长度为 10000 点, 所以在文献 [1] 中序列组 (A) 和 (B) 的误差线也相对较短且无重叠现象, 说明只有数据序列长度达到一定程度之后, MMSE 在处理具有长时相关性的复杂序列时, 其统计稳定性才可被接受, 否则就会出现类似于图 2(a) 的重叠现象, 单次试验则极有可能获得完全矛盾的分析结果.

图 2(b)—(f) 的统计性能相比图 2(a) 均有所改善, 表明 MMSE 统计稳定性差的原因的确来自于多尺度化过程和评价模式相似性时使用的硬阈值 HF 函数. 因此无论是从改进传统多尺度化过程出发, 还是从使用模糊隶属度函数代替 HF 函数的思路出发, 其结果均会有明显改进: 对比图 2(b) 和 (c), 本文提出的 PFMF 函数相比文献中常用的 IFMF 函数显示了更优的处理结果, 可能正是由于 PFMF 函数考虑到实际物理过程时不可避免的扰动或采样误差, 引入了“敏感阈”的概念, 才会明显增强 MMFE 在处理复杂序列时的统计性能; 对比同一列上下两个子图可知, MEMD 在处理复杂序列时显示了较好的捕获尺度的能力, 第二行的子图其统计性能明显优于第一行. 因此, 图 2(c) 和 (f) 所表示的 MMFE 在统计性能上是可以接受的, 且以图 2(f) 为最佳, 即 mp-MMFE 统计稳定性最优.

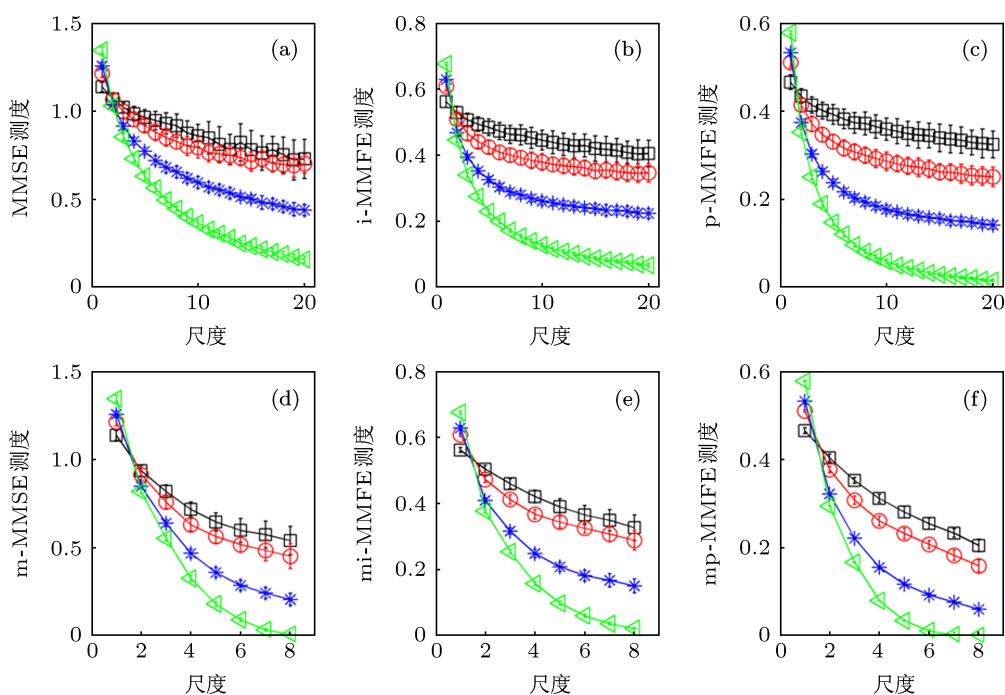


图2 六种多变量多尺度熵处理四种三变量序列的结果对比, 方框示意三通道都为 $1/f$ 噪声 (序列组 (A)), 圆圈示意两通道 $1/f$ 噪声和一通道白噪声 (序列组 (B)), 星号示意一通道 $1/f$ 噪声和两通道白噪声 (序列组 (C)), 三角示意三通道都为白噪声 (序列组 (D)), 曲线和误差线分别表示 20 次仿真实验的均值和标准差 (a) MMSE; (b) i-MMFE; (c) p-MMFE; (d) m-MMSE; (e) mi-MMFE; (f) mp-MMFE

需要重点说明的是, 上述仿真试验选取了固定的 r 值 ($r = 0.45$), 正如本文引言中所阐述的一样, r 值的选取对上述各测度的统计稳定性都具有潜在的影响. 这种影响与其对单变量熵的影响机制类似, 读者可以参考 Lu 等^[22] 以及 Liu 等^[23] 的研究. 针对 MMSE, 也有研究系统阐述了 r 值选取的影响^[9]. 在 MMSE 测度中, r 值是向量相似与不相似的分界值, 因此 r 本身正是为了消除序列可能存在的系统干扰或采样误差, 但这种误差的绝对值有可能超出 r 这个界限, 这正是本文所引入的“敏感阈”关键所在. 敏感阈以内的向量被认为完全相同 (隶属度为 1), 敏感阈之外的向量根据它们之间距离的大小赋予不同的隶属度值. 上面的仿真结果证实了敏感阈的引入的确能够提高统计稳定性, 也就是说 PFMF 函数消除了序列随机性或系统干扰的影响. 另外, 已有文献研究证实单变量 FuzzyEn 的 r 值依赖性远低于 SampEn^[10], 因此有理由相信 i-MMFE 以及 im-MMFE 的 r 值依赖性会有较高的改善 (因其使用与 FuzzyEn 同样的隶属度函数 IFMF), 即选取不同 r 值对试验结果的影响较小. 那么基于 PFMF 的 p-MMFE 以及 ip-MMFE 的 r 值依赖性是否与 i-MMFE 以及 im-MMFE 一样甚至更低呢? 本文相信会得到肯定的结果, 但这需要进一步的严格测试,

由于篇幅所限, 本文将该问题留待后续工作中进行.

4 多通道生理信号序列的多变量复杂度分析

心脏逐搏电、机械活动间期序列的复杂度会因心血管疾病或自然衰老等的发生而降低, 比如最常见诸于各研究文献的 HRV 分析. 该理论已被很多研究文献所证实, 并广泛应用于心血管疾病的早期预警领域^[9,14,24,25]. HRV 是心脏窦房结在交感神经和副交感神经协调交互调控下自律除极化的结果, 心脏的机械收缩和舒张正是在这种自律除极化 (电) 活动介导下完成的, HRV 中包含了这种机械收缩和舒张的变异性, 一般称为心脏收缩间期变异性 (systolic time interval variability, SIV) 和心脏舒张间期变异性 (diastolic time interval variability, DIV), 生理研究表明 HRV 主要由 DIV 来体现, 而 SIV 相对变异性较小, 从而保证稳定的心脏射血^[26]. 我们在上文中已经提到, HRV 序列本身具有 $1/f$ 分布, 因此研究者普遍认为这是一种相对复杂的序列, 这与生理系统本身的自组织性不谋而合, 生理系统的自组织性使得生理信号本身表现出自相似和长时相关性, 这正是上文所使用的 $1/f$ 噪声的独特性质.

既然 HRV 主要由 DIV 来体现, 那么 DIV 序列本身也应展现出 $1/f$ 分布, HRV 和 DIV 序列之间也许存在着某种非线性交互关系. 这正是本文使用这两类序列进行分析的出发点.

本文对 HRV 和 DIV 序列进行多变量复杂度分析 (以下简记为 HRV-DIV 复杂度), 探究 HRV 和 DIV 的非线性交互关系对无创评价心脏功能的潜在价值. 为此, 本文从 5 min 心电数据 (ECG) 和同步记录的 5 min 脉搏波数据 (PW) 中构建 HRV 和 DIV 序列, 具体构建方法描述如下: 首先, 对 ECG 和 PW 数据进行预处理, 使用模板匹配法寻找 ECG 中的 R 波位置 [27], 连续两个 R 波之间对应的 PW 最小值和最大值分别为心脏收缩期起点 (标记为 a) 和快速射血期终止时刻 (标记为 b), 本周期的 b 点与下周期 a 点之间 PW 二阶差分最大值为心脏收缩期终止时刻 (也即舒张期开始时刻, 标记为 c 点), 则 HRV 由逐搏 RR 间期构成, DIV 序列由下一搏动 a- 本搏动 c 构成, 如图 3 所示. 本试验所用数据包括 30 例心衰患者 (纽约心脏分级 II—IV 级, 年龄 40—75 岁, 平均 63.4 岁) 以及 30 例健康志愿者 (年龄范围 23—72 岁, 平均 47.4 岁), 来源于课题组 2009 年在山东大学齐鲁医院进行的临床试验 (由山东大学齐鲁医院伦理委员会批准), 数据采集仪器为济南汇医融工科技有限公司生产的 CV FD-I 型心血管状态监测仪, 采样率 1000 Hz.

首先, 研究 HRV-DIV 复杂度在心衰患者和健康志愿者之间的区别. 为避免年龄的潜在影响, 进行分析时两组人群需满足年龄匹配, 最终选取 21 例心衰患者 (年龄范围 40—71 岁, 平均 59.5 岁) 和 17 例健康志愿者 (年龄范围 46—72 岁, 平均 57.3 岁) 的数据进行分析. 计算过程与本文第 3 部分中的仿真试验一致: 分别计算上述 6 种多变量多尺度熵, 由于是两通道数据, 嵌入维数向量设置为 $M = [2, 2]$, 时间间隔向量设置为 $\tau = [1, 1]$, 取阈值参数 $r = 0.2$, 传统多尺度化过程中取尺度为 1—6, 而基于 MEMD 过程的多尺度化尺度数取决于 MEMD 自身, 本文计算前 3 个尺度 ($\Psi = 3$). 试验结果如图 4 所示.

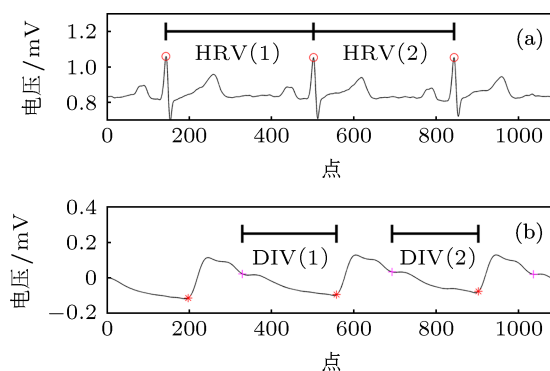


图 3 示意 HRV 序列和 DIV 序列的构建过程 (a) ECG 信号, 圆圈标记 R 波位置; (b) 与 ECG 同步记录的 PW 信号, 星号和加号分别表示收缩期起始位置 a 和终止位置 c

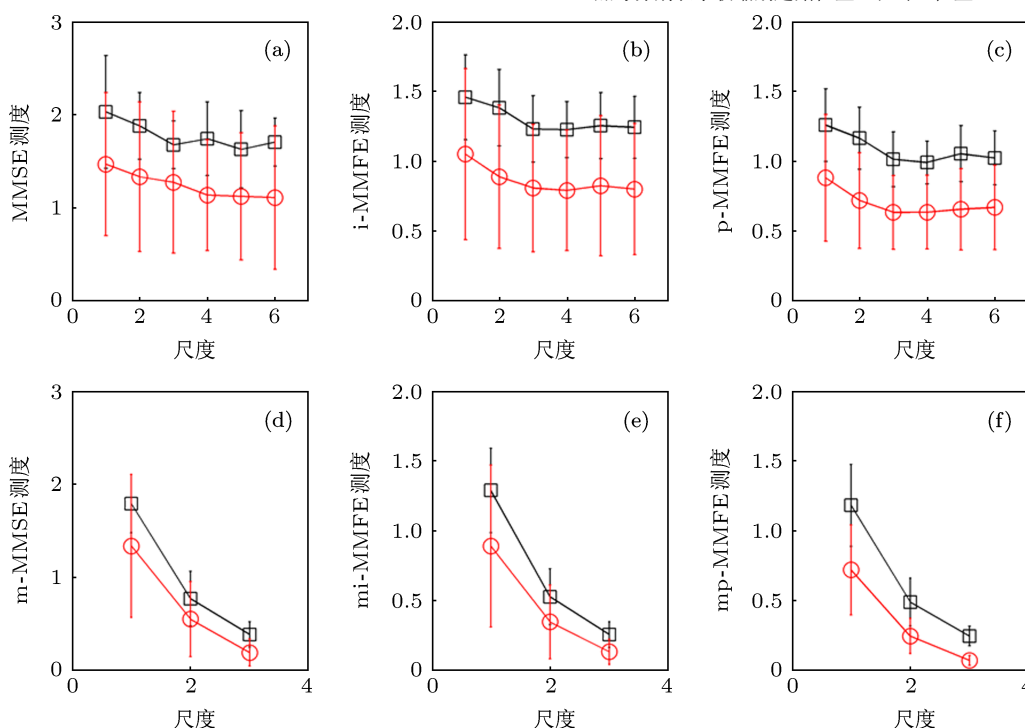


图 4 心衰患者和健康志愿者的多变量复杂度, 方框示意健康对照组, 圆圈示意心衰患者组, 曲线和误差线分别表示 17 次 (健康对照组) 或 21 次 (心衰患者组) 计算的均值和标准差 (a) MMSE; (b) i-MMFE; (c) p-MMFE; (d) m-MMSE; (e) mi-MMFE; (f) mp-MMFE

健康志愿者 HRV-DIV 复杂度均大于心衰患者, 提示心衰的出现会引起心血管复杂度的降低. 对比图 4(b), (c) 与 (a) 可知, 引入模糊隶属度函数可提高多变量熵的统计稳定性 (误差线变小), 这与本文的仿真实验结果一致. 另外, 对比对应上下两个子图, 基于 MEMD 的多尺度化过程可以更有效地刻画生理信号的不同尺度成分, 仅采用 MEMD 而不利用模糊隶属度函数时, 高尺度上的熵值稳定性已经比较高 (图 4(d)). 图 4(f) 性能最佳, 这也与仿真实验结

果一致. 另外, 心衰患者与健康对照组 HRV-DIV 多变量复杂度在高尺度上的区别更为明显, 这提示心衰的发生可能首先影响到 HRV-DIV 的长时相关性.

本文继续通过 HRV-DIV 复杂度研究年龄的影响, 从健康志愿者中取 10 例健康年青志愿者 (年龄范围 23—34 岁, 平均 27.5 岁) 和 10 例健康老年志愿者 (年龄范围 56—72 岁, 平均 63.4 岁). 进行如上的分析, 结果如图 5 所示.

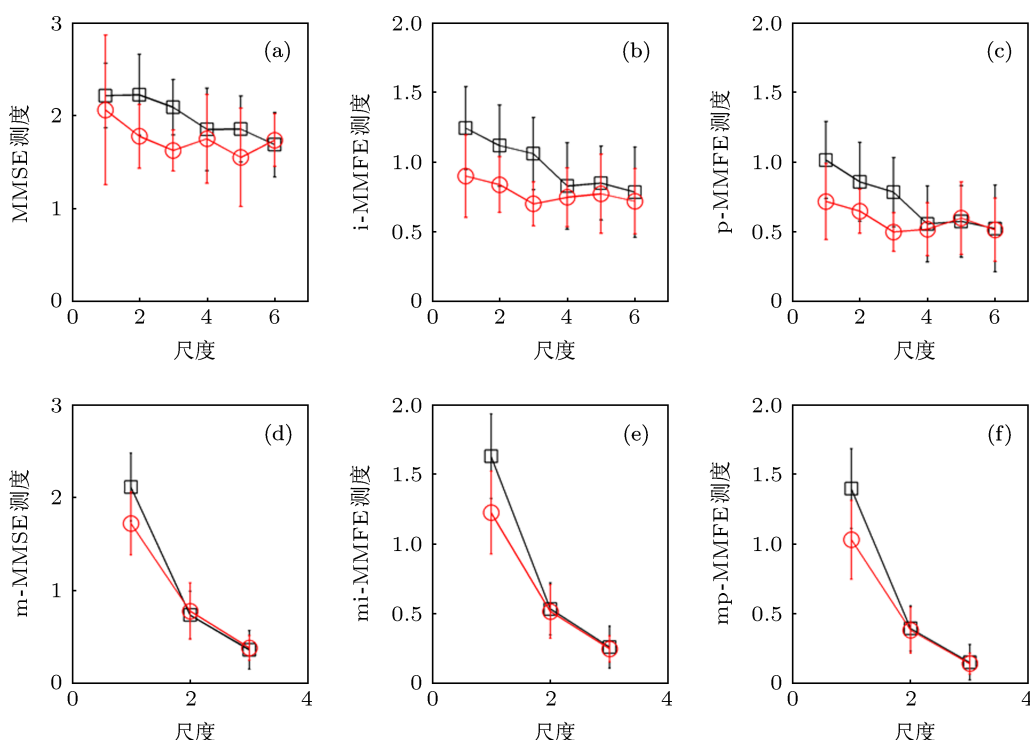


图 5 健康年青志愿者与健康老年志愿者的多变量复杂度. 方框示意健康年青组, 圆圈示意健康老年组, 曲线和误差线分别表示 10 次计算的均值和标准差 (a) MMSE; (b) i-MMFE; (c) p-MMFE; (d) m-MMSE; (e) mi-MMFE; (f) mp-MMFE

分析图 5 可得到与图 4 类似的结论, MMSE 并不能很好地区分两组之间的差异, 且统计稳定性较差. 不同的是, 在高尺度上健康年青志愿者与健康老年志愿者的 HRV-DIV 复杂度并无显著差异 (图 5(b)—(f)), 两组之间的差异表现在低尺度中. 这提示我们衰老与心脏疾病引发的复杂度丢失可能具有不同的变化进程: 衰老首先引起近程 HRV-DIV 复杂度的降低 (比如相近搏动之间的相互关系变弱, 而正常生理系统应该是高度耦合和自组织的), 但从长期来看, 这种自组织性尚未有明显改变, 然而, 心衰会明显同时改变近程和长时 HRV-DIV 复杂度. 该结论可能有助于心血管疾病的无创预警研究, 但需要增加入组人数进行进一步的临床试验来证实.

5 结论

针对多尺度多变量样本熵存在的两个主要问题: 1) 统计稳定性差; 2) 尺度化过程不适用于处理非线性、非平稳信号, 本文从改变模式相似判断准则出发, 研究了模糊隶属度函数的形式对统计稳定性的影响, 提出了统计稳定性能较优的多尺度多变量模糊熵. 并通过改进尺度化过程, 引入多变量经验模态分解算法, 使之适用于处理非线性、非平稳信号. 仿真试验和临床数据分析结果表明:

1) 模糊隶属度函数的形式对 MMFE 的统计稳定性具有显著影响, 本文所构造的物理模糊隶属度函数考虑到系统扰动或采样误差的影响, 更加符合实际物理或生理现象, 统计稳定性最佳;

2) MEMD 多尺度化过程避免了滑动平均造成的虚假振荡, 并且对数据的平稳性没有要求, 在处理生理信号时显示了较优的能力;

3) 本文提出的 mp-MMFE 可以敏感且鲁棒地分析非线性、非平稳信号的多变量复杂度; 临床应用表明, 基于 mp-MMFE 分析, 心衰患者的 HRV-

DIV 复杂度在各尺度上都明显降低, 而健康老年志愿者的 HRV-DIV 复杂度仅在低尺度上具有明显下降, 提示心脏疾病和自然衰老过程对 HRV-DIV 复杂度影响的进程不同, 该结论对心血管疾病的无创预警具有显著参考意义.

- [1] Ahmed M U, Mandic D P 2011 *Phys. Rev. E* **84** 061918
- [2] Pincus S M 1991 *Proc. Nat. Acad. Sci. U.S.A.* **88** 2297
- [3] Richman J S, Moorman J R 2000 *Am. J. Physiol. Heart. Circ. Physiol.* **278** H2039
- [4] Cao L, Mees A, Judd K 1998 *Physica D* **121** 75
- [5] Costa M, Goldberger A L, Peng C K 2002 *Phys. Rev. Lett.* **89** 068102
- [6] Ahmed M U, Li L, Cao J, Mandic D P 2011 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society Boston, Massachusetts, USA, Aug. 30–Sep. 3, 2011* p810
- [7] Ahmed M U, Mandic D P 2012 *IEEE Signal Process Lett.* **19** 91
- [8] Looney D, Ahmed M U, Mandic D P 2012 *Natural Intelligence: the INNS Magazine* **1** 40
- [9] Li P, Liu C Y, Wang X P, Li L P, Yang L, Chen Y C, Liu C C 2013 *Med. Biol. Eng. Comput.* **58** 581
- [10] Chen W T, Zhuang J, Yu W X, Wang Z Z 2009 *Med. Eng. Phys.* **31** 61
- [11] Xiong G L, Zhang L, Liu H S, Zou H J, Guo W Z 2010 *J. Zhejiang University-Science A* **11** 270
- [12] Chen X J, Li Z, Bai B M, Pan W, Chen Q H 2011 *J. Electron. Inform. Tech.* **33** 1198 (in Chinese) [陈小军, 李赞, 白宝明, 潘玮, 陈清华 2011 电子与信息学报 **33** 1198]
- [13] Sun K H, He S B, Yin L Z, A D L-Duo L K 2012 *Acta Phys. Sin.* **61** 130507 (in Chinese) [孙克辉, 贺少波, 尹林子, 阿地力·多力坤 2012 物理学报 **61** 130507]
- [14] Liu C Y, Li K, Zhao L N, Liu F, Zheng D C, Liu C C, Liu S T 2013 *Comput. Biol. Med.* **43** 100
- [15] Nikulin V V, Brismar T 2004 *Phys. Rev. Lett.* **92** 089803
- [16] Valencia J F, Porta A, Vallverdu M, Claria F, Baranowski R, Orłowska-Baranowska E, Caminal P 2009 *IEEE Trans. Biomed. Eng.* **56** 2202
- [17] Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N C, Tung C C, Liu H H 1998 *Proc. R. Soc. London A* **454** 903
- [18] Amoud H, Snuoussi H, Hewson D, Doussot M, Duchene J 2007 *IEEE Signal Process Lett.* **14** 297
- [19] Rehman N, Mandic D P 2010 *Proc. R. Soc. London, Ser. A* **466** 1291
- [20] Hu M, Liang H 2012 *IEEE Trans. Biomed. Eng.* **59** 12
- [21] Ahmed M U, Rehman N, Looney D, Rutkowski T M, Mandic D P 2012 *Bull. Pol. Ac.: Tech.* **60** 433
- [22] Lu S, Chen X, Kanters J K, Solomon I C, Chon K H 2008 *IEEE Trans. Biomed. Eng.* **55** 1966
- [23] Liu C Y, Liu C C, Shao P, Li L P, Sun X, Wang X P, Liu F 2011 *Physiol. Meas.* **32** 167
- [24] Kaplan D T, Furman M I, Pincus S M, Ryan S M, Lipsitz L A, Goldberger A L 1991 *Biophys. J.* **59** 945
- [25] Iyengar N, Peng C K, Morin R, Goldberger A L, Lipsitz L A 1996 *Am. J. Physiol. Regulatory Integrative Comp. Physiol.* **271** R1078
- [26] Liu C Y, Liu C C, Li L P, Zhang Q G, Li B 2009 *3rd International Conference on Bioinformatics and Biomedical Engineering Beijing, China, June 11–13, 2009* p2609
- [27] Li P, Liu C C, Zhang M, Che W B, Li J 2011 *Acta Biophys. Sin.* **27** 222 (in Chinese) [李鹏, 刘常春, 张明, 车文彪, 李键 2011 生物物理学报 **27** 222]

Multiscale multivariate fuzzy entropy analysis*

Li Peng¹⁾ Liu Cheng-Yu¹⁾ Li Li-Ping²⁾ Ji Li-Zhen¹⁾
Yu Shou-Yuan¹⁾ Liu Chang-Chun^{1)†}

1) (School of Control Science and Engineering, Shandong University, Jinan 250061, China)

2) (College of Science and Technology, Shandong University of Traditional Chinese Medicine, Jinan 250355, China)

(Received 1 February 2013; revised manuscript received 24 February 2013)

Abstract

Multiscale multivariate sample entropy can test the multivariate complexity, which is accepted as a kind of reflection of nonlinear dynamical interactions in multichannel data. It is however relatively unstable due to the rigid ranking scheme used in comparison among different patterns. It is not applicable to the nonlinear and non-stationary data because the multiscale framework used is in fact handled by moving average succeeded by down-sampling, which actually has a premise of stationary data. We substitute a fuzzy membership function for the original rigid one and compare the performances of different kinds of fuzzy membership functions. In addition, we employ the multivariate empirical mode decomposition (MEMD) to capture different scales. Results show that the substitution of fuzzy membership function brings in significant stability. It is much more obvious by using the introduced physical fuzzy membership function (PFMF). Also MEMD could capture scales more robustly. In conclusion, the introduced PFMF- and MEMD-based MMFE perform best. Final analysis on the interactions between heart rate variability (HRV) and heart diastolic time interval variability (DIV) validates it. In addition, the results show that the multivariate complexity between HRV and DIV decreases in aging or heart failure group but in a distinctly different decreasing manner—it decreased at low scales with aging, indicating a loss of short-range correlation but both at low and high scales with heart failure, which shows the losses of both short- and long-range correlations. Studies in noninvasive detection of cardiovascular diseases should benefit from the above conclusions.

Keywords: multivariate complexity, multiscale multivariate fuzzy entropy, physical fuzzy membership function, multivariate empirical mode decomposition

PACS: 05.45.Tp, 05.45.-a, 89.75.Da

DOI: 10.7498/aps.62.120512

* Project supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 61201049), the Graduate Independent Innovation Foundation of Shandong University, China (Grant No. yzc12082), the Excellent Young Scientist Awarded Foundation of Shandong Province, China (Grant No. BS2012DX019), and the China Postdoctoral Science Foundation (Grant No. 2013M530323).

† Corresponding author. E-mail: changchunliu@sdu.edu.cn, lskyp@mail.sdu.edu.cn