

基于多变量符号转移熵的癫痫脑电分析

刘倩倩¹ 戴加飞² 李锦³ 王俊¹ 侯凤贞⁴

(1. 南京邮电大学图像处理与图像通信江苏省重点实验室, 南京, 210003; 2. 南京军区南京总医院, 南京, 210002; 3. 陕西师范大学物理学与信息技术学院, 西安, 710062; 4. 中国药科大学理学院, 南京, 210009)

摘要: 大脑神经元细胞群的异常同步放电是癫痫的病因, 这种异常放电是目前诊断癫痫的重要依据。利用复杂度理论来分析癫痫信号已经成为研究热点, 而符号转移熵是反应系统混乱程度的一种非线性指标, 在研究癫痫脑电信号特征的提取中有重要的作用。符号转移熵一般都是用来衡量两个变量之间的动力学特征及方向性信息, 忽略了多个变量之间相互作用。本文基于多变量符号转移熵研究分析了癫痫脑电信号, 实验中将原始信号符号化后通过数值分析, 对导联信号及信号长度的选取以及稳健性分析, 表明该方法能够对正常人与癫痫病人的脑电信号进行显著区分, 且该算法稳健可靠, 该研究结果对临床辅助诊断有帮助。

关键词: 脑电信号; 多变量符号转移熵; 复杂度理论

中图分类号: TP391 **文献标志码:** A

Epileptic EEG Based on Improved Multivariate Symbolic Transfer Entropy

Liu Qianqian¹, Dai Jiafei², Li Jin³, Wang Jun¹, Hou Fengzhen⁴

(1. Image Processing and Image Communications Key Lab, Nanjing University of Posts & Telecommunications, Nanjing, 210003, China; 2. Nanjing General Hospital of Nanjing Military Command, Nanjing, 210002, China; 3. College of Physics and Information Technology, Shanxi Normal University, Xi'an, 710062, China; 4. School of Science, China Pharmaceutical University, Nanjing, 210009, China)

Abstract: Epilepsy is caused by abnormal synchronous discharge of neurons in the brain, which constructs the main basis of its diagnosis. The use of complexity theory to study the epileptic signal has become a hot spot. The symbolic transfer entropy as a reflection of the degree of chaos of nonlinear system of indicators can be used as a characteristic of epilepsy. It plays an increasingly important role in the study of epilepsy in electroencephalogram signals (EEG) feature extraction. But symbolic transfer entropy is generally used to measure the dynamic characteristics and directional information between two variables and ignores the interaction between multivariate. Epileptic EEG signals are analyzed based on multivariate symbolic transfer entropy. By choosing the lead signal and the signal length to analyze the robustness, the method can be used to distinguish normal person and patients with epilepsy. It is proved that the algorithm is robust and reliable for clinical diagnosis.

Key words: EEG signal; multivariate symbolic transfer entropy; complexity theory

基金项目: 国家自然科学基金(61271082, 61201029, 61102094, 61401518)资助项目; 江苏省重点研发计划(BE2015700)资助项目; 江苏省自然科学基金(BK20141432)资助项目; 南京军区南京总医院基金(2014019)资助项目; 中国药科大学中央高校基本科研业务费专项资金(FY2014LX0039)资助项目。

收稿日期: 2015-11-07; **修订日期:** 2016-01-23

引言

癫痫是一种由多种原因引起的高发病率的中枢神经系统疾病^[1,2]。癫痫主要是由脑部中枢神经系统功能失常引起的,脑部中枢神经系统功能失常是因为大脑神经元细胞群异常同步放电。这种放电有着突发性、暂时性及反复性等特点,这些特点也是目前癫痫诊断的主要依据。因为病人在癫痫未发作前以及癫痫发作后的体检中均未发现异常,所以怎样区分病人脑部是否存在异常放电具有重要意义。利用复杂度方法来研究癫痫信号已经成为研究热点。熵作为一种非线性指标,因它能够反应系统混杂程度,故在癫痫脑电信号特征提取的研究中作为一种癫痫特征有着非常重要的作用。目前有许多方法可以对信号复杂度进行测量,如近似熵^[3]、符号熵^[4]、部分互信息熵^[5]、相对熵^[6]及相对转移熵^[7]等。这些测量信号复杂度的方法在技术已越来越成熟,如近似熵是一种适用于段数据的复杂性度量,它主要应用于生物电信号中的脑电信号和心电信号。生物信号的复杂度在某种疾病下可能与人体正常的情况下不一样,故许多文献都把近似熵当作脑电信号特征参数来诊断神经疾病或研究大脑生命迹象。但是这些方法都是从信号中提取单个变量的特征或两个变量的特征,然而只研究两个变量可能会得出一种伪关系,比如,如果想要研究 $X \rightarrow Y \rightarrow Z$ 之间的关系,利用两个变量分析就只能给出 X 和 Z 之间的关系而忽略了 Y 与它们之间的关系,而用多变量分析能研究 X, Y 和 Z 之间的关系,所以对信号多变量特性的研究具有极大的意义。

本文研究的是多变量符号转移熵在癫痫脑电信号中的分析,它通过应用条件互信息理论^[8]来实现,希望用该算法来区分正常人和癫痫病人。本文使用的正常人和癫痫病人原始脑电时间序列的数据来自于某军区总医院。本文通过对正常人和癫痫病人的脑电信号的原始时间序列进行符号化划分,选取区分度最好的导联以及最佳的序列长度,并对脑电信号原始时间序列的多变量符号转移熵进行求值。对所得到的多变量符号转移熵值分析后发现该算法能显著区分正常人和癫痫病人。最后,对该算法的抗噪性进行分析,发现该算法叠加上高斯白噪声后无明显变化并且趋势完全相同,表明该算法有较高的鲁棒性。

1 基本原理

1.1 多变量转移熵

转移熵^[9]的定义涉及到无穷向量,因此它很难可靠地估计高维变量。为了克服这一问题,提出了分解转移熵^[10]。基于分解转移熵,本文引入以下的结论:给定一个平稳多元离散时间随机过程 X ,其子过程表示为 x, y, z, w, \dots ,在 t 时刻的值为 X_t, x_t, \dots 。它们的过去时刻过程定义为 $X_t^- = (X_{t-1}, X_{t-2}, \dots)$ 和 $x_t^- = (x_{t-1}, x_{t-2}, \dots)$,并把 x_t^- 作为 X_t^- 的一个子集。转移熵 $I_{x \rightarrow y}^{\text{TE}} = I(x_t^-; y_t | X_t^- \setminus x_t^-)$ 是指在已知条件 $X_t^- \setminus x_t^-$ 下, x_t^- 到 y_t 的不确定度, x_t^- 和 $X_t^- \setminus x_t^-$ 为两个无限维部分。首先利用链式法则解决个别 x 滞后的问题。即

$$I(x_t^-; y_t | X_t^- \setminus x_t^-) = \sum_{\tau=1}^{\infty} I(x_{t-\tau}; y_t | X_t^- \setminus x_t^-, x_{t-\tau}^-) \quad (1)$$

然后利用理论图形模型^[11]来解决仍然存在的无限维问题,它利用的是独立过程中的马尔可夫性质,如图 1 所示。其表示为

$$I(x_{t-\tau}; y_t | X_t^- \setminus x_t^-, x_{t-\tau}^-) = I(x_{t-\tau}; y_t | S_{y, x_{t-\tau}}) \quad (2)$$

式中: $S_{y, x_{t-\tau}} \subset X_t^- \setminus x_t^- \cup x_{t-\tau}^-$ 为确定的有限子集。

如果过程的时间序列图已知,那么有限子集 $S_{y, x_{t-\tau}}$ 就可以确定。剩余的无穷和也可以在一些确定的 τ^* 截断,因此通常以指数 τ 衰减,即

$$I_{x \rightarrow y}^{\text{TE}} \approx I_{x \rightarrow y}^{\text{DTE}} = \sum_{\tau=1}^{\tau^*} I(x_{t-\tau}; y_t | S_{y, x_{t-\tau}}) \quad (3)$$

式中: τ^* 为估计值小于某些给定的绝对公差下的最小的 τ 值。

图 1 是 $x_{t-\tau}$ ($\tau=2$) 和 y_t 的条件互信息。其中无限区域代表无限集合 $X_t^- \setminus x_t^- \cup x_{t-\tau}^-$, 有限区域代表有限集合 $S_{y_t, x_{t-\tau}}$, 这两个集合能够满足式(3)。

1.2 原始序列符号化

符号动力学分析是指将原始时间序列转化成几个符合组合而成的序列然后再进行分析的方法, 它是在符号动力学、混沌序列分析理论的基础上发展起来的^[12,13]。原始序列的符号化方法定义为

$$s_i(x_i) = \begin{cases} 0 & u_1 < x_i \leq (1+a)u_1 \text{ 或者 } (1+a)u_2 \leq x_i < u_2 \\ 1 & (1+a)u_1 < x_i < \infty \text{ 或者 } -\infty < x_i < (1+a)u_2 \\ 2 & (1-a)u_1 < x_i \leq u_1 \text{ 或者 } u_2 \leq x_i < (1-a)u_2 \\ 3 & (1-a)u_2 \leq x_i \leq (1-a)u_1 \end{cases} \quad (4)$$

式中: $i=1, 2, \dots, N$ 表示序列位置的索引值; μ_1, μ_2 分别为原始

序列中小于零以及大于等于零的取样信号的均值; a 为取值范围为 $[0.03, 0.07]$ 的特殊常数, 因为 a 的值过大会使细节丢失, a 的值过小会使噪声的影响变大, 从而无法有效地获取信号中的动态信息, 为了使处理后的信号不会丢失时间序列的动力学特性, 通常取 $a=0.05$ ^[14]。

1.3 多变量符号转移熵

因为转移熵对参数的协调性要求较高而且对噪声也比较敏感, 因此符号转移熵(Symbolic transfer entropy, STE)^[15]是在符号化技术的基础上发展出来的新方法。

由多变量转移熵^[16]的定义 $I_{x \rightarrow y}^{TE} \approx I_{x \rightarrow y}^{DTE} = \sum_{\tau=1}^{\tau^*} I(x_{t-\tau}; y_t | S_{y_t, x_{t-\tau}})$ 可推导出多变量符号转移熵的定义, 把原始时间序列 x 转化成符号序列 $S = \{s_1, s_2, \dots, s_N\}$, $s_i \in A (A=0 \sim 3)$; 原始时间序列 y 转化成符号序列 $J = \{j_1, j_2, \dots, j_N\}$, $j_i \in A (A=0 \sim 3)$, 原始时间序列 z 转化成符号序列 $K = \{k_1, k_2, \dots, k_N\}$, $k_i \in A (A=0 \sim 3)$ 。多变量符号转移熵定义为

$$I_{S \rightarrow J}^{TE} \approx I_{S \rightarrow J}^{DTE} = \log \frac{p(s_{t-\tau}, j_t, s_{t-\tau-1}, j_{t-\tau+1}, j_{t-\tau}, j_{t-\tau-1}, k_{t-\tau}, k_{t-\tau-1})}{p(j_t, s_{t-\tau-1}, j_{t-\tau+1}, j_{t-\tau}, j_{t-\tau-1}, k_{t-\tau}, k_{t-\tau-1})} \frac{p(s_{t-\tau-1}, j_{t-\tau+1}, j_{t-\tau}, j_{t-\tau-1}, k_{t-\tau}, k_{t-\tau-1})}{p(s_{t-\tau}, s_{t-\tau-1}, j_{t-\tau+1}, j_{t-\tau}, j_{t-\tau-1}, k_{t-\tau}, k_{t-\tau-1})} \quad (5)$$

2 多变量符号转移熵的癫痫脑电分析

2.1 实验数据

本文使用的正常人与癫痫病人的原始脑电信号的时间序列从临床诊断中采集, 该数据来源于某军区总医院。该库中记录了包括 16 个导联信号的多参数脑电数据, 该信号的记录长度均大于 1 min, 采样周期为 512 Hz, 序列长度大于 40 000。本文使用的数据分别从正常脑电数据与癫痫脑电数据中随机各选取 10 个人的脑电数据。在脑电数据中取其 O_1, O_2 , ECG 导联信号, 即计算在已知 ECG 的条件下, $O_1 \rightarrow O_2 | ECG$ 的多变量符号转移熵(O_1, O_2 导联位于左右半脑对称的位置; 实验证明: O_1, O_2 导联在导联 ECG 的条件下的多变量符号转移熵有良好的区分度, 故选 O_1, O_2 导联作为研究对象)。本文将这两组数据分别记为“正常”及“癫痫”。

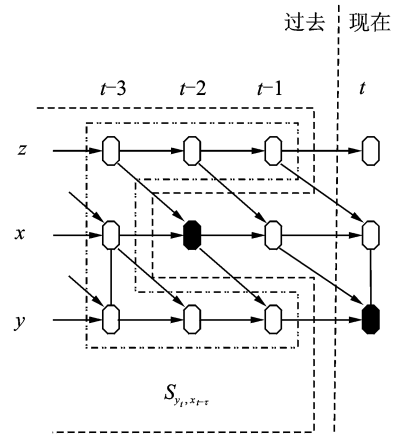


图 1 时间序列图

Fig. 1 Time series graph

2.2 分析方法

先使用式(4)符号化处理各个样本脑电信号的有效原始时间序列,再使用式(5)计算已经符号化处理过的有效原始脑电时间序列的多变量符号转移熵值。最后为保证该算法的有效性,对计算结果使用统计分析软件 SPSS 进行显著性检验。

2.3 数值计算结果与分析

2.3.1 导联信号

对于不同的导联组,从样本“正常”和“癫痫”中对每个个体的原始脑电时间序列取其序列长度 $L=80$ 。针对每一个体,取时间起点为 $t=100, 300, 600, \dots, 2400, 2700$ 的 10 个不同时间点,对每个时间序列进行符号化处理。把计算的每个个体的 10 组多变量符号转移熵值取均值,这个均值作为个体的多变量符号转移熵值。最后对这 10 个正常人和 10 个癫痫病人的多变量符号转移熵值进行平均,得到样本“正常”和“癫痫”的多变量符号转移熵值,结果如图 2 所示。

由图 2 可知,正常人和癫痫病人在不同导联组相同的序列长度下脑电信号的多变量符号转移熵值比较(横坐标表示:1: F_{P_1}, F_{P_2} ;2: F_3, F_4 ;3: C_3, C_4 ;4: P_3, P_4 ;5: O_1, O_2 ;6: F_7, F_8 ;7: T_3, T_4 ;8: T_5, T_6 ;STE:符号转移熵)。在计算 8 组导联的样本多变量符号转移熵中,第 9 导联 O_1 (位于左半球)和第 10 导联 O_2 (位于右半球)在 ECG 条件下多变量符号转移熵值的差值最大,由此可见信号区分度最好,所以取 O_1 和 O_2 导联组数据做分析对象。

2.3.2 多变量转移熵与数据长度的关系

从样本“正常”和“癫痫”中对每个个体的原始脑电时间序列分别取序列长度为 $L=40, 40 \times 2, 40 \times 3, 40 \times 4, 40 \times 5$ 。对每个个体在不同 L 下计算每个个体的多变量符号转移熵值。分别取得样本“正常”和“癫痫”中所有个体的多变量符号转移熵值的均值,该均值就是样本“正常”和“癫痫”在序列长度 L 下的多变量符号转移熵值。由不同 L 得到多变量符号转移熵与 L 的关系如图 3 所示。由图 3 可以看出,在同样 L 下,正常人原始脑电时间序列的多变量符号转移熵值明显比癫痫病人原始脑电时间序列的多变量符号转移熵值高。为兼顾算法效率与实验精度,取 $L=40 \times 4$ 。

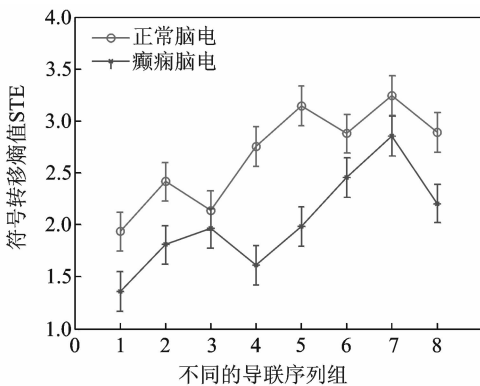


图 2 不同的导联组下符号转移熵的比较

Fig. 2 Comparison of the symbolic transfer entropy in different lead groups

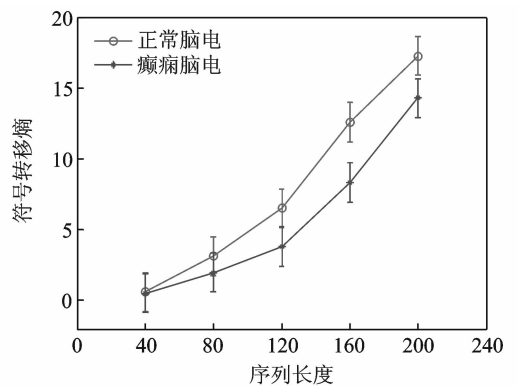


图 3 转移熵与数据长度 L 的关系

Fig. 3 Relationship between transfer entropy and data length L

2.3.3 序列长度 $L=160$ 时样本熵值差异

从样本“正常”及样本“癫痫”中对每个个体的脑电信号取固定序列长度为 $L=40 \times 4$ 的原始脑电时间序列。与上一实验相同,计算每个个体在序列长度为 40×4 的多变量符号转移熵值,得到的结果如表

1 所示。

基本研究正常人和癫痫病人原始脑电时间序列的多变量符号转移熵后,将样本“正常”和样本“癫痫”依照表 1 中的均值和方差绘成图,如图 4 所示。

表 1 正常人与癫痫患者的符号转移熵值

Tab. 1 Symbolic transfer entropy of normal and epileptic

个体	1	2	3	4	5	6	7	8	9	10	均值
正常	10.943 6	12.407 8	13.932 7	15.798 4	10.167	16.816 7	12.281 2	9.697 7	8.875 6	15.003	12.592 37
癫痫	9.683 4	9.525 8	11.468 4	9.580 5	8.570 9	11.723 9	4.646 3	9.368 7	1.331 6	7.431 8	8.333 13

2.3.4 统计分析 with 假设检验

为检验本文提出的算法的确切性和有效性,应用统计分析软件 SPSS 对计算结果进行独立 T 检验分析,结果为 $T=2.666, P=0.015 < 0.05$ 的差异性显著。这说明正常人和癫痫病人原始脑电时间序列的多变量符号转移熵的差异性十分明显,该算法能有效区分正常人和癫痫病人。

2.3.5 算法稳健性

根据前期的实验,从样本“正常”和“癫痫”中分别取序列长度 $L=40 \times i (i=1 \sim 5)$ 的两类脑电信号原始时间序列,对其进行叠加高斯白噪声的计算(假设脑电信号原始时间序列为 Data,本试验中采用的叠加高斯噪声的计算式为 $awgn(data, 15)$,即叠加 15 dB 的信噪比的高斯白噪声),结果如图 5 所示。由图 5 可知,原始脑电时间序列叠加高斯白噪声后的多变量符号转移熵与未叠加高斯白噪声的多变量符号转移熵没有显著的区别。因此说明本文中提出的算法鲁棒性较高且可靠有效。

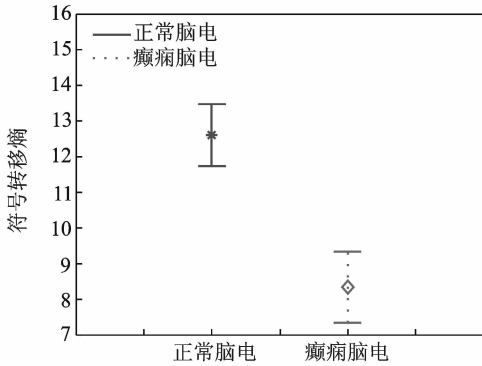


图 4 正常人和癫痫患者符号转移熵比较
Fig. 4 Comparison of symbolic transfer entropy between normal and epilepsy

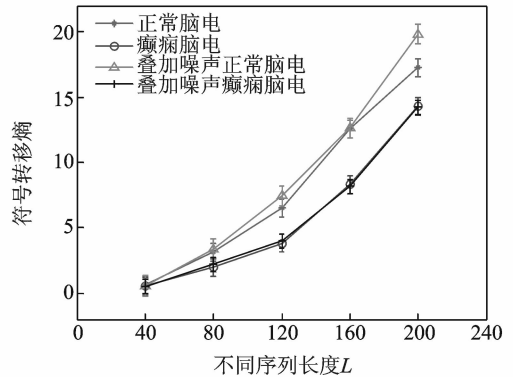


图 5 多变量符号转移熵算法鲁棒性分析
Fig. 5 Robustness analysis of multi-variable symbolic transfer entropy algorithm

3 结束语

本文主要是基于多变量转移熵的癫痫脑电信号分析,它是利用条件互信息理论来研究多个变量之间的关系,数值实验中通过确定导联信号的选取、信号序列长度的选取最终确定多变量符号转移熵值,最后对该算法加上高斯白噪声对其稳健性进行分析。在 8 组导联信号中,计算正常人与癫痫病人的多变量符号转移熵中其中第 9 导联 O_1 (位于左半脑) 和第 10 导联 O_2 (位于右半脑) 在 ECG 条件下多变量符号转移熵值得差值最大,由此可见信号区分度最好,所以取 O_1 和 O_2 导联组数据做分析对象。对于同一序列长度 L ,正常人的多变量符号转移熵值明显高于癫痫病人。本文研究左右脑之间的信息连通性,通过研究表明,正常人的左右脑连通性高于癫痫病人。这意味着如果左右脑连通性下降预示着脑功能发生了衰退或病变,这对于研究脑功能连接有十分重要的意义。分析实验结果还发现,该算法仅用较短的序列长度就

能区分正常人和癫痫病人,对原始脑电时间序列叠加上高斯白噪声后多变量符号转移熵值并没有太大的变化,且受噪声的影响较小,可见该算法鲁棒性较高,这一研究结果对临床辅助诊断有益。

参考文献:

- [1] Iasemidis L D, Shiau D S, Chaovalitwongse W, et al. Adaptive epileptic seizure prediction system[J]. *IEEE Transactions on Biomedical Engineering*, 2003,50(5):616-627.
- [2] Gadhomi K, Lina J M, Gotman J. Discriminating preictal and interictal states in patients with temporal lobe epilepsy using wavelet analysis of intracerebral EEG[J]. *Clinical Neurophysiology Official Journal of the International Federation of Clinical Neurophysiology*, 2012,123(10):1906-1916.
- [3] 郭兴明,李传鹏,卢德林. 定量递归分析和近似熵在心音分类识别中的应用[J]. *数据采集与处理*, 2013,28(5):559-564.
- [4] Guo Xingming, Li Chuanpeng, Lu Delin. Application of recurrence quantification analysis and approximate entropy in classification and recognition of heart sound[J]. *Journal of Data Acquisition and Processing*, 2013,28(5):559-564.
- [5] 成谢锋,于淼,姬汉贵,等. 基于概率分布的符号熵在心音分析中的应用[J]. *数据采集与处理*, 2015,30(5):948-955.
- [6] Cheng Xiefeng, Yu Miao, Ji Hangui, et al. Application of symbol entropy based on probability distribution to heart sound analysis [J]. *Journal of Data Acquisition and Processing*, 2015,30(5):948-955.
- [7] 张梅,崔超,马千里,等. 基于符号化部分互信息熵的多参数生物电信号的耦合分析[J]. *物理学报*, 2013,62(6):068704.
- [8] Zhang Mei, Cui Chao, Ma Qianli, et al. Coupling analysis of multivariate bioelectricity signal based symbolic partial mutual information[J]. *Acta Physica Sinica*, 2013,62(6):068704.
- [9] 张梅,王俊. 基于改进的符号相对熵的脑电信号时间不可逆性研究[J]. *物理学报*, 2013,62(3):038701.
- [10] Zhang Mei, Wang Jun. Modified symbolic relative entropy based electroencephalogram time irreversibility analysis[J]. *Acta Physica Sinica*, 2013,62(3):038701.
- [11] 王莹,侯凤贞,戴加飞,等. 改进的相对转移熵的癫痫脑电分析[J]. *物理学报*, 2014,63(21):21807.
- [12] Wang Ying, Hou Feizhen, Dai Jiawei, et al. Analysis on relative transfer of entropy based on improved epileptic EEG[J]. *Acta Physica Sinica*, 2014,63(21):21807.
- [13] Vejmelka M, Palus M. Inferring the directionality of coupling with conditional mutual information[J]. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2008, 77(2):026214.
- [14] Schreiber T. Measuring information transfer[J]. *Physical Review Letters*, 2000,85(2):461-464.
- [15] Faes L, Nollo G. Decomposing the transfer entropy to quantify lag-specific Granger causality in cardiovascular variability[C]//International Conference of the IEEE Engineering in Medicine and Biology Society. [S. l.]: IEEE, 2013:5049-5052.
- [16] Eichler M. Graphical modelling of multivariate time series [J]. *Probability Theory and Related Fields*, 2012,153(1/2):233-268.
- [17] 沈韡,王俊. 基于符号相对熵的心电信号时间不可逆性分析[J]. *物理学报*, 2011,60(11):744-747.
- [18] Shen Wei, Wang Jun. Time irreversibility analysis of ECG based on symbolic relative entropy [J]. *Acta Physica Sinica*, 2011,60(11):744-747.
- [19] Kostelich E J. The analysis of chaotic time-series data[J]. *Systems & Control Letters*, 1997,31(5):313-319.
- [20] Wessel N, Ziehmann C, Kurths J, et al. Short-term forecasting of life-threatening cardiac arrhythmias based on symbolic dynamics and finite-time growth rates[J]. *Physical Review E Statistical Physics Plasmas Fluids & Related Interdisciplinary Topics*, 2000,61(1):733-739.
- [21] Staniek M, Lehnertz K. Symbolic transfer entropy[J]. *Physical Review Letters*, 2008,100(15):3136-3140.
- [22] Runge J, Heitzig J, Petoukhov V, et al. Escaping the curse of dimensionality in estimating multivariate transfer entropy [J]. *Physical Review Letters*, 2012,108(25):1-5.

作者简介:



刘倩倩(1990-),女,硕士研究生,研究方向:医学信号处理, E-mail: daoyingliqianxiao@163.com.



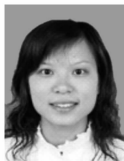
戴加飞(1981-),男,医师,研究方向:脑电图,癫痫鉴别诊断。



李锦(1973-),女,博士,副教授,研究方向:信号与信息处理、超声医学电子。



王俊(1973-),男,博士,教授,博士生导师,研究方向:生物医学信息处理。



侯凤贞(1990-),女,博士,副教授,研究方向:生物医学信息处理。

