

基于样本熵的注意力相关脑电特征信息提取与分类

燕楠, 王珏, 魏娜, 宗良

(西安交通大学生物医学信息工程教育部重点实验室, 710049, 西安)

摘要: 提出一种用样本熵作为特征进行注意力相关脑电信号的分析与分类处理, 并采用支持向量机(SVM)算法实现分类器的方法. 7位年龄在20~30岁之间的男性受试者接受了执行3种不同注意任务状态下的测试. 数据分析结果显示: 样本熵分类法对注意任务相关脑电信号分类的正确率可达85.5%, 优于传统频段能量法获得的分类精度(77.9%). 这个结果暗示了样本熵能有效地识别出自发脑电中注意力相关信息, 因而它可在脑电生物信息反馈治疗系统设计中获得广泛的应用.

关键词: 脑电; 生物反馈; 样本熵; 支持向量机

中图分类号: R318.4 **文献标识码:** A **文章编号:** 0253-987X(2007)10-1237-05

Feature Exaction and Classification of Attention Related Electroencephalographic Signals Based on Sample Entropy

Yan Nan, Wang Jue, Wei Na, Zong Liang

(Key Laboratory of Biomedical Information Engineering of Education Ministry, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A method regarding the sample entropy (SampEn) as features is proposed to carry out the analysis and classification of attention related electroencephalographic(EEG) signals, and the support vector machine (SVM) algorithm is used as classifiers for classification. seven males (aged from 20 to 30) are recruited to perform three attention-related tasks, including attention, inattention, and relaxation states. The processing results demonstrate that the classification accuracy of the SampEn gets up to 85.5% for classifying the relation between attention and inattention, obviously much higher than that with frequency band power (77.9%). It indicates that the SampEn is more effective to extract the information attention-related in EEG to show the clinical application prospects in EEG biofeedback systems.

Keywords: electroencephalography; biofeedback; sample entropy; support vector machine

脑电信号(EEG)包含着大脑皮层活动的丰富信息, 通过对 EEG 的分析, 进一步理解人的思维机理、实现神经系统损伤患者与周围环境的信息交换, 已成为当前脑-机接口(BCI)技术、神经功能康复等领域的重要研究手段^[1-2]. 脑电生物反馈治疗是 BCI 的一个典型应用^[2-5], 其应用操作性条件反射的原理, 以神经反馈仪为手段, 通过训练, 选择性地强化或抑制某一特征频率的脑电波, 从而实现改善和治疗疾病的目的. 图 1 是一个典型的脑电生物反馈训练系统框图, 主要由 EEG 信号采集、信号处理(特征

提取和分类)、反馈显示 3 部分组成, 其中信号特征的有效提取是影响分类正确率的一个重要因素. 由于这些疾病均与某个频段的脑电相关, 所以当前的脑电生物反馈系统大多直接采取某个波段(如 δ 波 1~4 Hz、 θ 波 4~8 Hz、 α 波 8~13 Hz、 β 波 14~30 Hz)的能量或者功率谱作为信号特征. 这种方法原理简单, 便于实现, 但也存在一定的问题, 由于不同状态下大脑头皮的脑电能量水平较为接近, 单纯以不同波段脑电的能量(功率)来分类, 存在误判率较高等缺点.

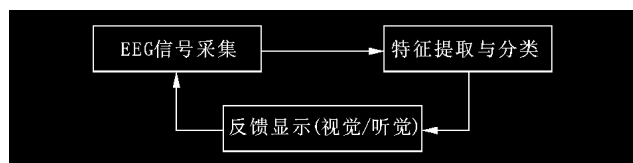


图1 脑电生物反馈系统框图

本文致力于提出一种更加有效的注意力相关脑电特征的特征提取与分类方案,利用非线性理论中的样本熵(SampEn)分析,通过计算 EEG 时间序列在模式上自相似性的得到注意力相关脑电的特征值,进而采用支持向量机(SVM)进行分类,并将结果与常用的不同波段脑电的能量特征分类方法进行比较,取得了满意的结果。

1 方法

1.1 样本熵的定义

样本熵是 Richman 等研究发展的一种有别于近似熵的不计数自身匹配的统计量,是对于近似熵算法的改进^[6]。序列越复杂,其样本熵值就越大,若序列的自我相似性越高,则样本熵值越小。在实际应用中,由于样本熵只需比较短的数据就能得出稳健的估计值,有较好的抗噪和抗干扰能力,又可以用于由随机成分和确定性成分组成的混合信号,并且具有分析效果优于简单统计参数(如均值、方差、标准差等)和不需要对原始信号进行粗粒化等特点,使之尤为适合对生物信号进行分析^[7-8]。

对于给定的 N 点时间序列 $u(i)$,其样本熵通过如下步骤计算得到,其中 m 是预先选定的模式维数, r 是预先选定的相似容限。

(1)设原始数据为 $u(1), u(2), \dots, u(N)$,共 N 个点。

(2)按序号连续顺序组成一组 m 维矢量,从 $X_m(1)$ 到 $X_m(N-m+1)$,其中

$$X_m(i) = [u(i), u(i+1), \dots, u(i+m-1)] \quad (1)$$

$$1 \leq i \leq N-m+1$$

这些量代表从第 i 个点开始连续的 $m+1$ 个 u 的值。

(3)定义

$$d[X_m(i), X_m(j)] = \max(u(i+k) - u(j+k)) \quad (2)$$

式中: $0 \leq k \leq m-1; 1 \leq i, j \leq N-m+1$ 。

(4)给定阈值 r ,对每个 $1 \leq i \leq N-m$ 的值,统计 $d[X(i), X(j)]$ 小于 r 的数目(称为模板匹配数)及此数目与距离总数 $N-m-1$ 的比值,记作

$$B_i^m(r) = \frac{1}{N-m-1} \{d[X_m(i), X_m(j)] < r \text{ 的数目}\}$$

其中 $1 \leq j \leq N-m, j \neq i$ 。求其对所有 i 的平均值

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \quad (3)$$

对于 $m+1$ 点矢量,同样有

$$A_i^m(r) = \frac{1}{N-m-1} \{d[X_{m+1}(i), X_{m+1}(j)] < r \text{ 的数目}\}$$

其中 $1 \leq j \leq N-m, j \neq i$,对所有 i 的平均值

$$A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r) \quad (4)$$

(5) $B^m(r)$ 和 $A^m(r)$ 分别是 m 点和 $m+1$ 点两序列相似的概率。在这里我们记样本熵为 H ,则理论上序列的样本熵为

$$H(m, r) = \lim_{N \rightarrow \infty} \{-\ln[A^m(r)/B^m(r)]\} \quad (5)$$

当 N 为有限值时,可得序列长度为 N 时样本熵估计值,记作

$$H(m, r) = -\ln[A^m(r)/B^m(r)] \quad (6)$$

因此 H 被认为是条件概率 $A^m(r)/B^m(r)$ 的精确值。

1.2 支持向量机分类方法

支持向量机是基于统计学习理论和结构风险最小化原则,根据线性可分情况下的最优分类面提出的^[9]。SVM 算法使用分类间隔控制学习机器的容量,从而使结构风险最小,使其在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势。

2 注意力相关脑电数据采集

针对不同注意力水平,特别是注意力的稳定性水平,本文设计了 3 种实验方案由受试者完成。① 注意任务。受试者坐在屏幕前,视角与屏幕保持平齐,屏幕上自动播放视频游戏(如图 2 所示),要求受试者集中精力,并假想游戏是由自己控制,控制一条小鱼不停地游动并吞食其他生物。实验进行之前,让受试者了解这个游戏的玩法,领会实验要求完成的任务。② 非注意任务。受试者坐在屏幕前,屏幕显示与任务 1 相同的游戏画面,但屏幕保持静止,受试者注视屏幕,不去想游戏相关内容。③ 静息任务。受试者处于尽可能放松状态,睁眼。在整个实验过程中,保持实验环境安静,让受试者端坐于一个舒适的靠背椅子上,全身肌肉处于一种放松状态,不产生任何的肌肉紧张与运动,保证受试者不受外界干扰。

7 位健康男性受试者,年龄在 20~30 岁,均在头脑清醒的情况下接受测试。脑电电极按照国际标准 10-20 系统放置,实验中共采用 12 个导联 FP1、

FP2、F3、F4、C3、C4、CZ、PZ、P3、P4、O1、O2 以及四路眼电 (EOG) 和参考电极. EEG 采集选用 NeuroScan 公司的 SCAN4.3 系统,该仪器在采集过程中已自动将 ECG、EOG 和 50 Hz 工频干扰滤除. 采样频率为 1 000 Hz,精度为 32 b,并采用 0.1~75 Hz 的带通滤波器. 每个任务持续 30 s,重复 10 次,这样每次实验每个任务可以得到 10 组数据.



图 2 注意状态任务界面图

3 结果分析与讨论

3.1 样本熵参数选择

参数 m 、 r 的选择是样本熵估计的关键,而样本熵与 m 、 r 、 N 取值有关. Pincus 建议取 $m=2$, $r=(0.1\sim0.25)\delta$,其中 δ 是原始数据 $u(i)$, $i=1\sim N$ 的标准差^[10]. $H(m=2,r,N)$ 反映当维数 m 由 2 增至 3 时产生新模式可能性的 大小,样本熵越大,说明产生新模式的机会越大,曲线越复杂. 本文对原始信号采样至 250 Hz 以后,分析计算大量 EEG 信号后发现,当 $N=500$ (即 2 s 的数据分析) 时,样本熵随 r 的变化关系如图 3 所示. 这里仅显示了 FP2 导联数据计算的结果,其他有差异的导联也符合这一变化趋势. 从图 3 中可看出,当 r 为 0.1~0.3 时,两注意状态样本熵的差值相对稳定. 事实上,样本熵的绝对值随参数不同变化很大,而我们仅关心其差异大小,因此本文在分析注意相关脑电数据时,取 $r=0.2\delta$. 在实验中,经过对大量 EEG 信号的计算发现,当数据长度 N 大于等于 500 个采样点后,样本熵达到相对稳定,其他导联也符合这一变化趋势,所以在本文中 N 取 500 个采样点,即滑动窗口为 2 s 的数据.

3.2 特征提取结果与讨论

由于脑电频率主要分布在 0.5~32 Hz 这个频段内,因此我们对原始脑电进行滤波,只取 0.5~32 Hz 内的脑电信号进行处理. 同时,为了方便处理,我们对原始信号进行 250 Hz 的重新采样,这样当 N

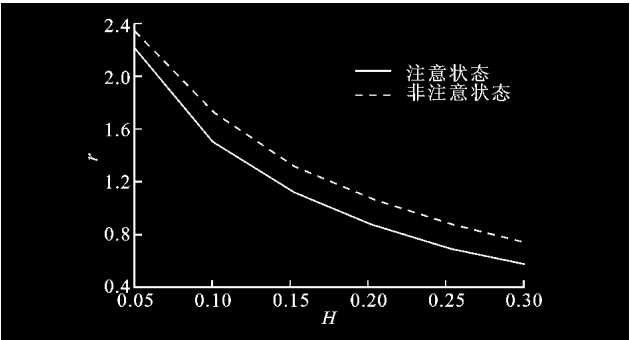


图 3 $N=500$ 时, r 与样本熵的关系

$=500$ 时,相当于采用 2 s 的滑动窗口进行样本熵估计运算.

对每一个受试者的每组数据以 2 s 为滑动窗口,50%重叠,计算所有导联 3 种不同任务状态下的样本熵,其平均值如表 1 所示. 依照统计学原理 t 检验,注意、非注意状态分别与放松状态比较,所有导联样本熵值都有显著性差异 ($P<0.01$),均呈现出放松状态的样本熵值明显高于注意状态. 注意与非注意状态间仅 FP1、FP2、F3、F4 导联的样本熵值存在显著性差异 ($P<0.01$),非注意状态的样本熵值高于注意状态,可将此作为注意状态的特征值.

表 1 任务状态 2 s 滑动窗口样本熵值 ($n=6$)

导联	H 的平均值		
	注意状态	非注意状态	放松状态
FP1	0.65±0.30	0.92±0.28	1.44±0.22
FP2	0.55±0.38	0.92±0.33	1.42±0.22
CZ	0.88±0.14	0.89±0.18	1.45±0.12
PZ	0.87±0.16	0.90±0.14	1.49±0.16
F3	0.73±0.27	0.89±0.21	1.49±0.17
F4	0.70±0.35	0.90±0.27	1.45±0.20
P3	0.94±0.16	0.95±0.15	1.61±0.17
P4	0.91±0.17	0.95±0.15	1.51±0.17
C3	0.91±0.15	0.93±0.17	1.58±0.12
C4	0.90±0.18	0.96±0.21	1.51±0.12
O1	1.03±0.11	1.10±0.15	1.81±0.09
O2	0.99±0.17	1.05±0.15	1.73±0.12

随机选取某一受试者,以 FP1 导联为例,对 30 s 数据进行动态分析,以 2 s 为滑动窗口,50%重叠,从图 4 得出该导联非注意状态的样本熵值明显高于注意状态,对于 F3、F4 导联以及其他受试者,也可得到相同分析结果.

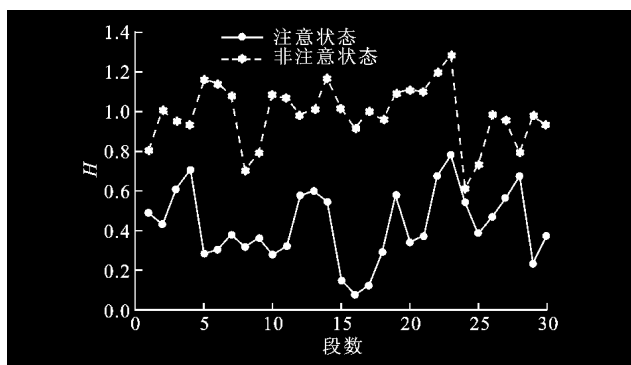


图4 FP2导联两注意状态样本熵值

随着脑负荷的增加,复杂性测度减小^[11]. 本文研究也发现在注意状态下样本熵值低于非注意状态,说明注意状态复杂性低. 就生理角度而言,受试者处于不同状态时大脑皮层的兴奋区域和兴奋程度不同,而样本熵通过衡量脑电数据序列中产生新模式概率的大小以表征受试者的生理状态的变化. 注意状态时大脑处于高负荷状态,有意识的调控作用增强,脑电活动更富有自主独立性,从而减弱了大脑神经细胞间的非线性程度,使脑电信号呈现“有序”状态. 非注意状态时大脑处于低负荷状态,即处于惰性状态,信息加工能力弱,大脑处于“无序”状态,复杂性程度高.

由前面的分析可以得出,FP1、FP2、F3、F4导联非注意状态的样本熵明显高于注意状态,我们用非注意状态样本熵减去注意状态样本熵的差(ΔH)作为纵坐标,得到对称导联样本熵差,见图5,发现除FP1、FP2、F3、F4导联外,其他导联两注意状态的样本熵差值无显著性差异($P < 0.01$),提示注意状态改变时,额部脑区变化显著,这与文献[12]提出的注意与大脑前额叶的控制区相关的结论相符合. 因此,我们选取这4个导联的样本熵值作为特征值进行后续分类处理.

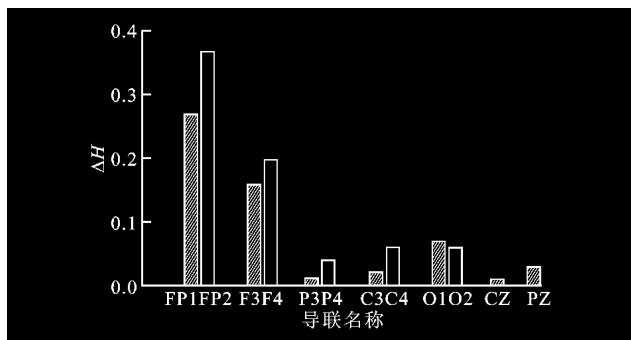


图5 对称导联两注意状态样本熵差

3.3 分类结果与讨论

采用SVM分类器对注意与非注意状态(放松

状态作为参考,不列在内)进行分类. 取FP1、FP2、F3、F4导联样本熵为特征值,输入SVM分类器进行分类,每次选取120组数据作为训练样本,30组数据作为测试集. SVM采用不同核函数将形成不同的算法,进行分类比较,表2给出了分类结果. 可以看出,采用不同核函数,分类准确率相差并不是很大,多项式核函数的性能较好一些.

表2 对样本熵特征值采用不同SVM算法的分类结果

受试者	分类准确率/%			
	线性SVM	多项式核函数	径向基核函数	Sigmoid核函数
1	86.2	86.5	88.0	85.5
2	84.7	89.0	86.5	83.6
3	87.6	87.4	86.1	86.7
4	82.3	83.9	82.5	80.9
5	85.5	86.2	84.9	83.9
6	80.9	80.2	79.9	78.7
平均值	84.5	85.5	84.6	83.2

作为对比,我们采用Daubechies6小波对C3、C4、CZ、PZ、FP1、FP2、F3、F4导联原始脑电信号进行8极小波包分解,得到 δ 、 θ 、 α 、 β 不同频段的重构信号,定义小波包重构信号的能量为重构信号的平方和^[13],由此可得到不同导联相应频段的能量特征. 以C3、C4、CZ、PZ、FP1、FP2、F3、F4导联 δ 、 θ 、 α 、 β 频段的能量作为分类器的输入,同以样本熵为特征值的分类效果进行对比分析,表3给出了分类结果. 对比两种特征提取方法的分类效果,发现样本熵特征的分类效果优于能量特征,说明在本文实验方案的注意状态和非注意状态的特征提取过程中,样本熵特征提取效果较好,对注意状态的变化差异更加敏感. 由分类结果推断,样本熵作为注意相关脑电的特征提取方法效果更佳.

表3 对能量特征采用不同SVM算法的分类结果

受试者	分类准确率/%			
	线性SVM	多项式核函数	径向基核函数	Sigmoid核函数
1	78.3	80.4	81.0	77.5
2	77.6	79.3	80.4	78.3
3	77.6	78.8	79.2	78.6
4	74.6	72.6	74.6	73.8
5	75.7	76.1	77.6	75.2
6	73.8	73.2	74.3	73.8
平均值	76.3	76.7	77.9	76.2

4 结 论

本文提出了一种采用样本熵分析的方法来提取注意力相关脑电特征,并采用 SVM 分类器进行分类.样本熵分析方法只需要较短数据就可得出稳健的估计值,是一种具有较好的抗噪和抗干扰能力的非线性分析方法.通过我们设计的实验表明,此方法对注意相关脑电特征提取非常有效,分类结果也优于传统的以频段能量作为特征的方法.

参考文献:

- [1] Ebrahimi T, Vesin J M. Brain-computer interface in multimedia communication [J]. IEEE Signal Processing Magazine, 2003, 20(1):14-24.
- [2] Phillips H. Controlling pain by watching your brain [J]. New Scientist, 2004, 182(2445):9.
- [3] McFarland D J, McCane L M. EEG-based communication and control: short-term role of feedback [J]. IEEE Transactions on Rehabilitation Engineering, 1998, 6(3):7-11.
- [4] Green C S, Bavelier D. Action video game modifies visual selective attention [J]. Nature, 2003, 423(6939):534-537.
- [5] 王珏,魏娜,张彤,等.注意力缺陷障碍伴多动症诊断技术的现状与展望 [J]. 中国康复医学杂志, 2004, 19(5):392-394.
- [6] Richman J S, Moorman J R. Physiological time-series analysis using approximate entropy and sample entropy [J]. Am J Physiol: Heart Circ Physiol, 2000, 278(6): 2039-2049.
- [7] Abasolo I D, Hornero R, Espino P, et al. Entropy analysis of the EEG background activity in Alzheimer's disease patients [J]. Physiological Measurement, 2006, 27(3):241-253.
- [8] 刘慧,和卫星,陈晓平.生物时间序列的近似熵和样本熵方法比较 [J]. 仪器仪表学报, 2004, 25(4):806-807.
Liu Hui, He Weixing, Chen Xiaoping. Approximate entropy and sample entropy for physiological time-series [J]. Chinese Journal of Scientific Instrument, 2004, 25(4):806-807.
- [9] Vapnik V N. Statistical learning theory [M]. New York: John Wiley and Sons Inc, 1998.
- [10] Pincus S M, Goldberger A L. Physiological time-series analysis: what does regularity quantify? [J]. Am J Physiol: Heart Circ Physiol, 1994, 266(4):643-656.
- [11] 刘建平,贺太纲,郑崇勋,等. EEG 复杂性测度用于大脑负荷状态的研究 [J]. 生物医学工程学, 1997, 14(1):33-37.
Liu Jianping, He Taigang, Zheng Chongxun, et al. Measuring EEG complexity for studying the state of mental load [J]. Journal of Biomedical Engineering, 1997, 14(1):33-37.
- [12] LaBerge D. Attention, awareness, and triangular circuit [J]. Consciousness and Cognition, 1997, 6(3): 149-181.
- [13] 薛建中. 基于自发脑电的脑-计算机接口研究[D]. 西安:西安交通大学生命科学与技术学院, 2003.

(编辑 杜秀杰)