**基于样本熵SVM的运动想象研究**

本文中，EEG数据通过28个头皮点击采集。由于EEG是非线性非平稳信号，所以采用了非线性动力学方法中的样本熵算法来提取EEG特征。然后利用SVM进行模式识别最终结果显示样本熵是一种提取不同大脑状态的特征的有效算法。

受试者想象不同的肢体运动，以提高或阻塞特定的大脑节律。本文中，ERD和ERS是研究的关键领域，特征通过样本熵算法提取，然后受试者的意图通过SVM确定。将样本熵应用于提取不同大脑状态特征，不同的参数设定将会影响样本熵的值和最终的分类精度。本文的最后一节将会讨论这些问题。

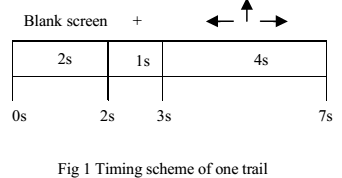
2.实验和数据

A．受试者

三名健康受试者（男性，右利手，20-32岁，平均年龄27.6岁）参与实验。所有受试者视力正常，所有受试者均完成了BCI训练。

B．过程

受试者按照箭头指示，分别想象左手、右手和脚运动。每个trial开始时，屏幕空白持续2s，t=2s开始，一个十字出现在屏幕中间，t=3s时，十字消失，同时一个箭头出现在屏幕中央，受试者开始想象。4s后箭头消失，受试者停止想象，休息2-3s。



一个session中，每个动作重复10次。每个受试者执行8-20个session，这取决于受试者的状态和承受能力。

C．数据记录和预处理

电极等距分布，从鼻根到枕骨隆突，从左耳到右耳，对称的覆盖头皮。参考电极为左右乳突。记录水平眼电和垂直眼电。采样频率250Hz。原始EEG信号使用Neuroscan4软件进行预处理。在这个过程中要做两件事。1、去除眼电干扰；2、带通滤波0.1-40Hz。

选择分布在主要运动区域的28个EEG通道作为研究对象。电极分别是F1, Fz, F2, FC5, FC3, FC1, FCz, FC2,FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz,CP2, CP4, CP6, P1, Pz, P2 and PO2。使用3-7s数据进行研究分析。

3.方法

A．样本熵

样本熵是一种统计学方法，用来量化确定性信号和随机信号的波动的不可预知性，可以用于短时间序列分析，适合随机的、噪声确定的复合过程。

对一个时间序列X={x(1),x(2),…,x(N)}，N是时间序列的长度，需要先确定两个参数，m是向量长度，r是相似容限，样本熵算法如下：

Step1. 形成一系列向量X(1)~X(N-m+1),他们的长度均为m：



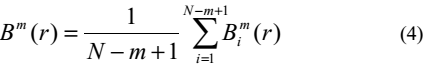
Step2. 计算两个向量之间的距离d[X(i),X(j)]，距离d定义为两个向量间对应元素的差的绝对值的最大值。



Step3. 定义为满足d[X(i),X(j)]<r的X(j)的数量。定义为：



Step4. 然后对所有求平均，得到



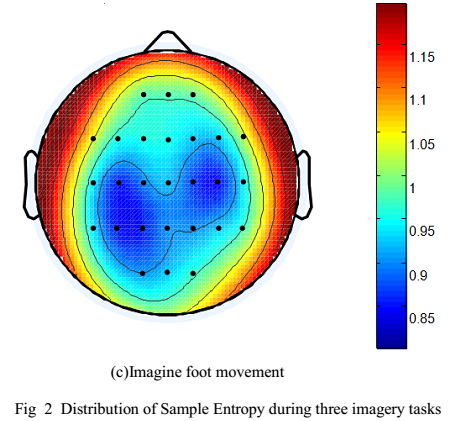
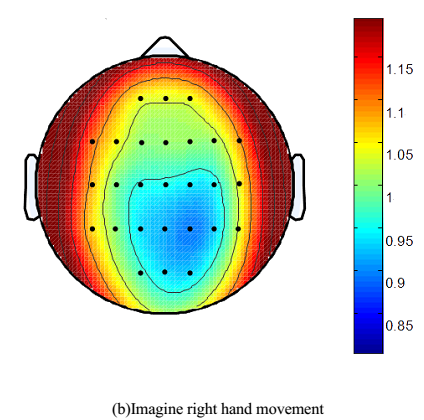
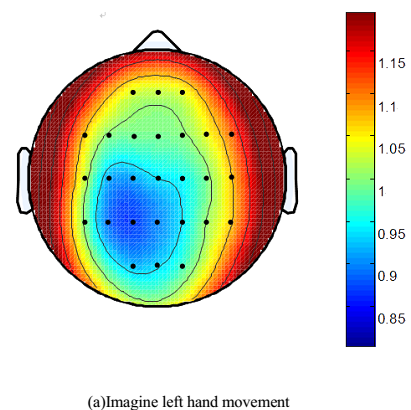
Step5. m增加一维变成m+1，重复step1-step4，计算得到

Step6. 理论上样本熵公式为：





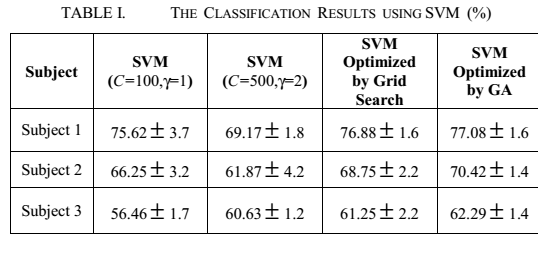
参考其他文献，m设为2，相似容限r设在0.1~2SD之间，SD为时间序列x(i)的标准差。R的值影响到最终精度，不同受试者，参数设置不同。计算28个EEG通道的样本熵的值作为特征。图2为受试者1在进行三种运动想象任务时平均样本熵的分布，r值设为0.1SD。

从图2可以看出，不同的想象状态具有不同的分布图样。

B．本文构建了一个基于RBF核函数的SVM分类器。SVM要设置两个参数C和r。对比了4种参数组合。其中两个为：C=100，γ=1和C=500，γ=2。第三种是通过网格寻优获得；最后一种是利用最优化方法GA（遗传算法）来确定合适的参数值。

4.结果

每个受试者的数据包含480个trial，每类想象任务各有160次。



5.讨论

相似容限r会影响到样本熵的值，如果r值设置的太大，那么满足step3的向量数就会减少，样本熵表示的是信号的趋势而不是细节；如果r设置的太小，会变大，样本熵表现的是信号的细节。图3和图2的区别是：图3的r=1.5SD，可以看出图2和图3的区别，图3中，三类任务不能明显区别出来。

图4曲线显示了相似容限和分类精度的关系，可以看出，每个受试者的最高精度有很大的差异，精度变化随r值变化。

