第 31 卷第 2 期 电 子 与 信 息 学 报 Vol.31No.2
2009 年 2 月 Journal of Electronics & Information Technology Feb. 2009

脑-机接口中基于 ERS/ERD 的自适应空间滤波算法

吕 俊 谢胜利 章晋龙
(华南理工大学电信学院 广州 510641)

摘 要：在与运动相关的脑-机接口(Brain-Computer Interface, BCI)研究中，如果样本规模小，共同空间模式(Common Spatial Patterns, CSP)滤波算法对离群点(可能为噪声)敏感，鲁棒性不好。为此该文提出自适应空间滤波(Adaptive Spatial Filter, ASF)算法，抽取滤波后脑电信号的方差作为特征，并寻找最优滤波器使两类特征中心的比值最大。与 CSP 不同，ASF 是迭代算法，具有软判决机制，能够依据历代更新后的滤波器，自适应地降低离群点对各类特征中心计算带来的影响。采用 BCI competition 2003 和 2005 中两套数据集进行实验，结果表明：尤其是在训练样本少的情况下，相对于 CSP，ASF 所提取的特征分类效果更好。

关键词：脑-机接口(BCI)；特征提取；共同空间模式(CSP)滤波法
中图分类号：TN911.7; R318 **文献标识码**：A **文章编号**：1009-5896(2009)02-0314-05

Adaptive Spatial Filter Based on ERD/ERS for Brain-Computer Interfaces

Lü Jun Xie Sheng-li Zhang Jin-long
(*School of Electronic and Information Engineering*, *South China University of Technology*, *Guangzhou* 510641, *China*)

Abstract: For motor related Brain-Computer Interface (BCI), if the sample size is small, Common Spatial Patterns (CSP) algorithm is sensitive to outlier data and lacks of robustness. In this paper, an Adaptive Spatial Filter (ASF) algorithm is proposed to take filtered samples' variances as the features and seek the spatial filter to maximize the ratio of two classes' means. Unlike CSP, ASF is an iterative algorithm and have soft determination. ASF can adaptively decrease outliers' effects according to the updated filters. Using two datasets from BCI competition 2003 and 2005, the experimental results show that ASF outperforms CSP，especially when training samples are few.

Key words: Brain-Computer Interface (BCI); Feature extraction; Common Spatial Patterns (CSP) algorithm

1 引言

近年来，与运动相关的脑-机接口(Brain-Computer Interface, BCI)研究取得了长足的进展，受到广泛的关注[1,2]。其主要生理背景之一即：事件相关同步/去同步现象(Event-Related Synchronization/De-Synchronization，ERS/ERD)[3]。针对 ERS/ERD，学者们提出了各种特征抽取算法[4]，其中著名的共同空间模式(Common Spatial Patterns，CSP)滤波法已在许多 BCI 实验中获得了成功[5,6]。然而 CSP 是基于大样本的，忽略了离群点(可能为噪声)对特征中心统计的影响。在实际应用中为了减少使用者的精神负荷，训练样本的数量往往不多，而且与运动相关的脑电信号容易受到肌电、眼电等各种伪迹的污染，这些因素使得 CSP 难以准确地估计滤波后各类特征的中心，造成分类效果欠佳。为此，以往的实验不得不依靠有经验的研究人员预先剔除受噪声污染严重的样本[7]。但是手工操作效率低，成本高，不能满足 BCI 作为实时通信系统的需要。

于是本文提出一种自适应空间滤波(Adaptive Spatial Filter, ASF)算法，抽取滤波后脑电信号的方差作为特征，并寻找最佳空间滤波器使得两类特征中心的比值最大。ASF 算法的思想是: 依据特征值之间的相邻关系，采用软判决的方式估计样本的噪声概率以及离群程度，并自适应地降低它们对各类特征中心计算的权重，从而提高算法的鲁棒性。考虑到特征值的邻域关系和滤波器的优化是密切相关的，ASF 算法采用迭代的形式，即先设置初始滤波器，接着计算滤波后特征的邻域关系，然后估计离群点，调整样本权重，最后更新滤波器使得投影后两类特征中心的比值最大，如此反复直至滤波器收敛。与 CSP 比较，ASF 的抗噪性更好，对离群点不敏感，能够更准确地估计各类特征的中心，更适合 BCI 的实际应用。在实验部分我们选择 Adaboost(LDA)算法作为分类器[8]，随机抽取 BCI 2003 dataset IV 和 BCI 2005 dataset I[9, 10] 不同数量的训练样本，比较了 CSP 和 ASF 所提取特征的分类性能。

2007-09-13 收到，2008-04-08 改回
国家自然科学基金重点项目(U0635001)和国家自然科学基金项目(60505005，60774094)资助课题

2　方法

设训练集为 $\mathrm{Tr}=\{(\boldsymbol{X}_i,y_i)\}_{i=1}^{n}$，测试集为 $\mathrm{Ts}=\{\boldsymbol{X}_p\}_{p=1}^{l}$。$\boldsymbol{X}_i,\boldsymbol{X}_p\in\mathbb{R}^{N\times T}$，其中 N 为电极数，T 为时间采样点数，y_i 表示 $\boldsymbol{X}_i$ 对应的运动类别，$y_i\in\{c_1,\cdots,c_j\}$，训练集 c_j 类运动的实验次数为 n_j。本文研究 BCI 二分类问题，$j=2$。

2.1 共同空间模式(CSP)滤波法

CSP 基于 ERS/ERD 现象，寻求最佳空间滤波器，使得滤波后两类样本的方差中心的比值最大。设空间滤波器为 $\boldsymbol{w}$，($\boldsymbol{w}\in\mathbb{R}^{N\times 1}$)，$\widetilde{\boldsymbol{X}}_i$ 为逐行零均值化后的 $\boldsymbol{X}_i$，$\boldsymbol{R}_i$ 为 $\widetilde{\boldsymbol{X}}_i$ 的自相关矩阵，则 CSP 的目标函数表达如下：

$$\max_{\boldsymbol{w}} J_1=\frac{\dfrac{1}{n_1}\sum_{i\in c_1}\left\|\boldsymbol{w}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_i\right\|^2}{\dfrac{1}{n_2}\sum_{i\in c_2}\left\|\boldsymbol{w}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_i\right\|^2}=\frac{\dfrac{1}{n_1}\sum_{i\in c_1}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_i\boldsymbol{w}}{\dfrac{1}{n_2}\sum_{i\in c_2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_i\boldsymbol{w}}=\frac{\boldsymbol{w}^{\mathrm{T}}\overline{\boldsymbol{R}}_1\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\overline{\boldsymbol{R}}_2\boldsymbol{w}} \tag{1}$$

应用奇异值分解可以求得 $\boldsymbol{w}$：

$$(\overline{\boldsymbol{R}}_2^{-1}\overline{\boldsymbol{R}}_1)\boldsymbol{w}_s=\beta_s\boldsymbol{w}_s \tag{2}$$

其中 β_s 为特征值，按降序排列，$s=1,2,\cdots,\xi$。取特征值 β_1 和 β_ξ 对应的特征向量作为最优滤波器，于是滤波后的时间序列可表示为

$$\boldsymbol{z}_{1h}=\boldsymbol{w}_1^{\mathrm{T}}\widetilde{\boldsymbol{X}}_h,\quad \boldsymbol{z}_{2h}=\boldsymbol{w}_\xi^{\mathrm{T}}\widetilde{\boldsymbol{X}}_h \tag{3}$$

其中 $\boldsymbol{X}_h\in\mathrm{Tr}\cup\mathrm{Ts}$，$h=1,2,\cdots,n+l$。最后提取 $\boldsymbol{z}_{1h}$ 和 $\boldsymbol{z}_{2h}$ 的方差作为 ERS/ERD 特征。设训练样本数为 N，电极数为 C，CSP 算法的计算复杂度为 $O((N+3)C^3+NC^2)$。

2.2 自适应空间滤波(ASF)算法

由式(1)可见：在 CSP 算法中，同类样本的特征值 $\left\|\boldsymbol{w}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_i\right\|^2$ ($\boldsymbol{X}_i\in c_j$)拥有一样的权重 $1/n_j$。在小样本情况下，一些离群点(可能是噪声)会影响到各类特征中心的计算。因而，ASF 算法将估计各样本的噪声概率和离群程度，并相应地调整它们的权重。ASF 算法的设计依据以下两条假设：

假设 1　若某样本附近存在较多的异类样本，那么该样本属于噪声的可能性较大。

假设 2　若某样本到同类样本的距离较近，那么该样本离群的程度较小，对类别中心估计的权重应该较大。

定义与样本 $\boldsymbol{X}_i$ 异类的样本集合为 $D_i=\{\boldsymbol{X}_k:1\le k\le n,\ y_k\ne y_i\}$，因为 $\boldsymbol{X}_i\in c_j$，故训练集 $\mathrm{Tr}=c_j\cup D_i$。用 $P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w})$ 表示样本 $\boldsymbol{X}_i$ 属于噪声的概率，用 $F_{\mathrm{cong}}(c_j\mid\boldsymbol{X}_i,\boldsymbol{w})$ 表示样本 $\boldsymbol{X}_i$ 对 c_j 类中心估计的权重。以特征值之间的欧式距离度量相应样本 $\boldsymbol{X}_a$ 和 $\boldsymbol{X}_b$ 之间的距离：

$$d_{ab}=\left\|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_a\boldsymbol{w}-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_b\boldsymbol{w}\right\|_F^2 \tag{4}$$

根据以上 2 条假设得

$$P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w})=\left[\sum\nolimits_{\boldsymbol{X}_k\in D_i}f(d_{ik})\right]\Big/\left[\sum\nolimits_{\boldsymbol{X}_u\in\mathrm{Tr}\setminus\boldsymbol{X}_i}f(d_{iu})\right] \tag{5}$$

$$F_{\mathrm{cong}}(c_j\mid\boldsymbol{X}_i,\boldsymbol{w})=\left[\sum\nolimits_{\boldsymbol{X}_g\in c_j\setminus\boldsymbol{X}_i}f(d_{ig})\right]\Big/\left[\sum\nolimits_{\boldsymbol{X}_u,\boldsymbol{X}_v\in c_j}f(d_{uv})\right] \tag{6}$$

其中 $f(d)$ 是核函数，各种核函数的特性可以参见文献[11]。为了减轻离群点对 $P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w})$ 和 $F_{\mathrm{cong}}(c_j\mid\boldsymbol{X}_i,\boldsymbol{w})$ 估计的影响，本文选择高斯核函数，令 $f(d)=\exp(-d/\sigma)$，σ 为待定参数。然后对 CSP 进行改造，ASF 的目标函数形式如下：

$$\max_{\boldsymbol{w}} J_2=\frac{\sum_{i\in\{y_i=c_1\}}(1-P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w}))F_{\mathrm{cong}}(c_1\mid\boldsymbol{X}_i,\boldsymbol{w})\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_i\boldsymbol{w}}{\sum_{i\in\{y_i=c_2\}}(1-P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w}))F_{\mathrm{cong}}(c_2\mid\boldsymbol{X}_i,\boldsymbol{w})\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_i\boldsymbol{w}} \tag{7}$$

令

$$\boldsymbol{A}=\sum_{i\in\{y_i=c_1\}}(1-P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w}))F_{\mathrm{cong}}(c_1\mid\boldsymbol{X}_i,\boldsymbol{w})\boldsymbol{R}_i \tag{8}$$

$$\boldsymbol{B}=\sum_{i\in c_2}(1-P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w}))F_{\mathrm{cong}}(c_2\mid\boldsymbol{X}_i,\boldsymbol{w})\boldsymbol{R}_i \tag{9}$$

则式(7)可改写为

$$\max_{\boldsymbol{w}} J_2=(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{w})/(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{w}) \tag{10}$$

给定 $\boldsymbol{w}$，则 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为常数，可依据式(10)用奇异值分解更新 $\boldsymbol{w}$：$(\boldsymbol{B}^{-1}\boldsymbol{A})\boldsymbol{w}^{+}=\lambda\boldsymbol{w}^{+}$，由于记录到的脑电信号是高维矩阵，且样本数量不大，所以 $\boldsymbol{B}$ 可能不可逆。这里应用 PCA [12]：令 $\boldsymbol{P}=(\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_d)$，$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_d$ 为 $(\boldsymbol{A}+\boldsymbol{B})$ 前 d 个最大的特征值对应的特征向量，且 $d=\mathrm{rank}(\boldsymbol{A}+\boldsymbol{B})$，则式(10)可改写为

$$\max_{\boldsymbol{w}} J_2=\max_{\boldsymbol{v}} J_2=\frac{\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P})\boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P})\boldsymbol{v}}=\frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}''\boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{B}''\boldsymbol{v}} \tag{11}$$

有

$$(\boldsymbol{B}''^{-1}\boldsymbol{A}'')\boldsymbol{v}_q=\gamma_q\boldsymbol{v}_q \tag{12}$$

式(12)中 γ_q 为特征值，$\boldsymbol{v}_q$ 是与 γ_q 对应的特征向量，$q=1,2,\cdots,d$，$\gamma_1>\gamma_2>\ldots>\gamma_d$，那么更新后的空间滤波器为

$$\boldsymbol{w}^{+}=\boldsymbol{P}\boldsymbol{v}_1/\left\|\boldsymbol{P}\boldsymbol{v}_1\right\| \tag{13}$$

考虑到 $P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w})$，$F_{\mathrm{cong}}(c_j\mid\boldsymbol{X}_i,\boldsymbol{w})$ 是关于滤波器 $\boldsymbol{w}$ 的函数，ASF 以迭代的方式寻找最优的 $\boldsymbol{w}$。ASF 算法的步骤如下：

(1)给定 $\mathrm{Tr}=\{(\boldsymbol{X}_i,y_i)\}_{i=1}^{n}$，计算 $\boldsymbol{R}_i$，采用 CSP 得到 $\boldsymbol{w}^{(0)}$，设置高斯半径 σ，截止门限 θ；

(2)将 $\boldsymbol{w}^{(t-1)}$ 代入式(5)，式(6)计算 $P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w}^{(t-1)})$ 和 $F_{\mathrm{cong}}(c_j\mid\boldsymbol{X}_i,\boldsymbol{w}^{(t-1)})$；

(3)据式(8)-式(13)计算 $\boldsymbol{w}^{(t)}$；

(4)如果 $||\boldsymbol{w}^{(t)}-\boldsymbol{w}^{(t-1)}||<\theta$ 那么：输出 $\boldsymbol{w}^{(t)}$，停机；否则：t=t+1，转到步骤(2)。

因为使 c_1 类特征中心与 c_2 类特征中心的比值最大或者最小都是利于分类的，所以式(7)的分子分母可以互换，ASF 算法需要运行 2 次，求得最优滤波器 $\boldsymbol{w}_1^*$，$\boldsymbol{w}_2^*$，提取特征：

$$\mathrm{feature1}_h=\boldsymbol{w}_1^{*\mathrm{T}}\boldsymbol{R}_h\boldsymbol{w}_1^*,\ \mathrm{feature2}_h=\boldsymbol{w}_2^{*\mathrm{T}}\boldsymbol{R}_h\boldsymbol{w}_2^*,\quad h=1,2,\cdots,n+l \tag{14}$$

设 ASF 收敛时迭代次数为 I，则 ASF 的计算复杂度为 $O((N+3)C^3+NC^2+2I(N^2/2+NC^2+3C^3))$；相对于 CSP，ASF 的计算复杂度增加量为 $\Delta O(2I(N^2/2+NC^2+3C^3))$，其中迭代次数 I 受到高斯半径 σ，截至门限 θ 以及训练样本数 N 等因素的影响。当 N 和电极数 C 较小时，调整 σ 和 θ 的值限制 I，能够有效地控制 ASF 的计算复杂度。

2.3 关于参数 σ 的讨论

当 $\sigma \to +\infty$ 时，用 $|D_i|$ 表示与 $\boldsymbol{X}_i$ 异类的样本的个数，据式(5)，式(6)可得

$$\left.\begin{aligned}\lim_{\sigma\to+\infty} P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w}) &= |D_i|/(n-1)\\ \lim_{\sigma\to+\infty} F_{\mathrm{cong}}(c_j \mid \boldsymbol{X}_i,\boldsymbol{w}) &= 1/n_j\end{aligned}\right\} \tag{15}$$

由式(15)可见：此时 $P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w})$，$F_{\mathrm{cong}}(c_j \mid \boldsymbol{X}_i,\boldsymbol{w})$ 为常数与 $\boldsymbol{w}$ 无关，特征的邻域关系不影响中心的估计，ASF 算法一步收敛，。比较式(1)和式(7)可见：CSP 没有自适应估计噪声的机制，$P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w})=0$，而且 CSP 没有依据同类运动特征之间的相邻关系调整样本权重，$F_{\mathrm{cong}}(c_j \mid \boldsymbol{X}_i,\boldsymbol{w})=1/n_j$，所以 CSP 是 ASF 在 $P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w})=0$，$\sigma \to +\infty$ 时的特例。

当 $\sigma \to 0$ 时：

$$\begin{aligned}&\lim_{\sigma\to 0} P_{\mathrm{nois}}(\boldsymbol{X}_i,\boldsymbol{w})\\ &\quad=\lim_{\sigma\to 0}\left\{1+\left[\sum\nolimits_{\boldsymbol{X}_u\in \mathrm{Tr}\setminus(\boldsymbol{X}_i\cup D_i)} f(d_{iu})\right]\Big/\left[\sum\nolimits_{\boldsymbol{X}_k\in D_i} f(d_{ik})\right]\right\}^{-1}\\ &\quad=\begin{cases}1,\ d_{ik}=\min\limits_{\boldsymbol{X}_j\in \mathrm{Tr}/\boldsymbol{X}_i} d_{ij}, & \exists \boldsymbol{X}_k\in D_i\\ 0,\ d_{ik}>\min\limits_{\boldsymbol{X}_j\in \mathrm{Tr}/\boldsymbol{X}_i} d_{ij}, & \forall \boldsymbol{X}_k\in D_i\end{cases}\end{aligned} \tag{16}$$

$$\begin{aligned}&\lim_{\sigma\to 0} F_{\mathrm{cong}}(c_j \mid \boldsymbol{X}_i,\boldsymbol{w})\\ &\quad=\left\{1+\left[\sum\nolimits_{\boldsymbol{X}_u,\boldsymbol{X}_v\in c_j\setminus \boldsymbol{X}_i} f(d_{uv})\right]\Big/\left[\sum\nolimits_{\boldsymbol{X}_g\in c_j} f(d_{ig})\right]\right\}^{-1}\\ &\quad=\begin{cases}1,\ d_{ig}=\min\limits_{\boldsymbol{X}_u,\boldsymbol{X}_v\in c_j} d_{uv}, & \exists \boldsymbol{X}_g\in c_j\\ 0,\ d_{ig}>\min\limits_{\boldsymbol{X}_u,\boldsymbol{X}_v\in c_j} d_{uv}, & \forall \boldsymbol{X}_g\in c_j\end{cases}\end{aligned} \tag{17}$$

式(16)，式(17)的物理意义是：当 $\sigma \to 0$ 时，若距离样本 $\boldsymbol{X}_i$ 最近的样本 $\boldsymbol{X}_j$ 与 $\boldsymbol{X}_i$ 异类，那么 $\boldsymbol{X}_i$ 为噪声点，否则 $\boldsymbol{X}_i$ 不是噪声点；若存在样本 $\boldsymbol{X}_g$ 与 $\boldsymbol{X}_i$ 同属于 c_j 类，并且 $\boldsymbol{X}_g$ 与 $\boldsymbol{X}_i$ 是 c_j 类样本中两两距离最近的一对，那么 $\boldsymbol{X}_i$ 对 c_j 类中心估计的权重为 1，否则，$\boldsymbol{X}_i$ 对 c_j 类中心估计的权重为 0。易知：当 $\sigma \to 0$ 时，ASF 算法不收敛。

由以上分析可见：参数 σ 度量了特征邻域关系对 ASF 算法的影响程度。决定着 ASF 算法的收敛性。设 $\Psi=\{\psi:\psi=(P_{\mathrm{nois}},F_{\mathrm{cong}})\}$，$W=\{\boldsymbol{w}^{(t)} \mid \|\boldsymbol{w}^{(t)}\|=1, t=1,2,\cdots,+\infty\}$，ASF 的计算过程为 $A1(\boldsymbol{w}^{(t-1)})=\psi$，$A2(\psi)=\boldsymbol{w}^{(t)}$，即 $\boldsymbol{w}^{(t)}=A2(A1(\boldsymbol{w}^{(t-1)}))\triangleq E(\boldsymbol{w}^{(t-1)})$，$E$ 是范数空间完备子集 W 到其自身的映射。ASF 算法的收敛性问题实质是 E 的不动点问题。该问题的讨论可以参见文献[13]。本文在实验中选择较大的 σ 值使得 ASF 算法收敛。

3 实验结果及分析

3.1 BCI 2003 dataset IV(self-paced 1s)实验

BCI 2003 dataset IV 的实验目的是区分左右手指运动前产生的脑电图信号(EEG)。该实验数据取自 28 导电极，共有训练样本 316 个，从中随机抽出 N 个样本(N=100, 200, 300；两类运动样本数大致相等)，然后进行 8-30Hz 的带通滤波。在不同规模的训练集和参数设置下，CSP 和 ASF 的测试误分率以及 ASF 的迭代次数 I(各 200 次实验均值)如表 1 所示。

由表 1 可见：(1)给定高斯半径 σ 和截止门限 θ：随着训练样本数量 N 的增加 CSP 和 ASF 的误分率均呈下降趋势，而且 ASF 比 CSP 的误分率低；当训练样本数少 N=100 时，ASF 的误分率明显低于 CSP；当训练样本数 N=300 时，收敛迭代次数 I 较小，其原因是 ASF 的初始滤波器 $\boldsymbol{w}^{(0)}$ 源自 CSP，样本多时 CSP 对 $\boldsymbol{w}^{(0)}$ 的估计较准确；(2) 给定高斯半径 σ 和训练样本数量，如果截至门限 θ 过大(取 $1.0\mathrm{e}-3$)，ASF 的分类性能优势较小，$\theta=1.0\mathrm{e}-6$ 和 $\theta=1.0\mathrm{e}-9$ 时，ASF 的分类性能优势较大；但是 θ 越小，收敛迭代次数 I 越大；(3)给定截止门限 θ 和训练样本数量，σ 过大(取 2000)时，ASF 和 CSP 的分类性能较接近，$\sigma=300$ 和 $\sigma=500$ 时，ASF 的分类性能优势较大；但是 σ 越小，收敛迭代次数 I 越大。

3.2 BCI 2005 dataset I(motor imagery in ECoG recordings)实验

BCI 2005 dataset I 的实验目的是区分舌头和手指运动想象产生的皮层脑电图信号(ECoG)。相对于 EEG，ECoG 避开了颅骨和头皮噪声较小。该实验数据取自 64 导电极，共有训练样本 278 个。本文预先挑选了部分电极(No.12, 21, 22, 29, 30, 31, 37, 38, 39, 40, 46)[10]进行 8-30Hz 的带通滤波，然后从中随机抽出 N 个样本(N=30, 90, 270；两类运动样本数大致相等)。在不同规模的训练集和参数设置下，CSP 和 ASF 的测试误分率以及 ASF 的迭代次数 I(各 200 次实验均值)如表 2 所示。

由表 2 可见：(1)给定高斯半径 σ 和截止门限 θ：随着训练样本数量 N 的增加 CSP 和 ASF 的误分率均呈下降趋势，当训练样本数少 N=30 时, ASF 的误分率明显低于 CSP；当训练样本数较多 N=90 和 N=270 时，ASF 与 CSP 的误分率几乎没有差别，收敛迭代次数 I 也较小，说明该数据集受噪声污染小；(2) 给定高斯半径 σ 和训练样本数量，如果截至门限 θ 过大(取 $1.0\mathrm{e}-3$)，ASF 的分类性能优势较小，$\theta=1.0\mathrm{e}-6$ 和 $\theta=1.0\mathrm{e}-9$ 时，ASF 的分类性能优势较大；但是 θ 越小，收敛迭代次数 I 越大；(3)给定截止门限 θ 和训练样本数量，σ 过大取 2000 时，ASF 和 CSP 的误分率十分接近，$\sigma=300$ 和 $\sigma=500$ 时，ASF 的分类性能优势较大；但是 σ 越大，收敛迭代次数 I 越小。

4 结论

本文基于ERS/ERD现象提出了一种新的自适应空间滤波(ASF)算法，用于二分类 BCI 的特征提取。ASF 是迭代算法，根据历代更新后的滤波器，计算特征值之间的邻近关系，然后通过软判决自适应地降低离群点对各类特征中心估计带来的影响。实质上，若去除软判决机制，不调整样本权重，直接求解最佳滤波器，ASF 即为经典的 CSP 算法。实验表明：相对与 CSP，ASF 算法的鲁棒性更好，更加适合小样

表 1　BCI 2003 dataset IV CSP 与 ASF 分类性能的比较

		$\theta=1.0\,\mathrm{e}-3$			$\theta=1.0\,\mathrm{e}-6$			$\theta=1.0\,\mathrm{e}-9$		
		100	200	300	100	200	300	100	200	300
$\sigma=300$	CSP	0.359	0.262	0.223	0.354	0.264	0.223	0.357	0.259	0.228
	ASF	0.337	0.243	0.213	0.327	0.247	0.211	0.328	0.244	0.213
	I	43.830	31.435	4.240	44.805	35.265	7.850	46.220	38.110	11.525
$\sigma=500$	CSP	0.365	0.261	0.228	0.356	0.264	0.231	0.356	0.262	0.227
	ASF	0.343	0.239	0.218	0.328	0.238	0.220	0.325	0.238	0.215
	I	21.850	17.415	3.240	32.440	23.145	6.020	34.610	25.065	8.640
$\sigma=1000$	CSP	0.354	0.262	0.229	0.354	0.263	0.226	0.357	0.263	0.227
	ASF	0.337	0.245	0.222	0.335	0.245	0.219	0.336	0.243	0.218
	I	11.845	12.015	2.275	18.010	14.465	4.605	22.280	16.080	6.960
$\sigma=2000$	CSP	0.343	0.262	0.226	0.357	0.265	0.225	0.359	0.268	0.225
	ASF	0.335	0.249	0.224	0.343	0.249	0.222	0.345	0.247	0.223
	I	9.770	5.655	2.060	9.280	8.320	4.055	10.735	9.840	5.970

表 2　BCI 2005 dataset I CSP 与 ASF 分类性能的比较

		$\theta=1.0\,\mathrm{e}-3$			$\theta=1.0\,\mathrm{e}-6$			$\theta=1.0\,\mathrm{e}-9$		
		30	90	270	30	90	270	30	90	270
$\sigma=300$	CSP	0.187	0.103	0.079	0.187	0.101	0.082	0.182	0.102	0.081
	ASF	0.162	0.102	0.080	0.159	0.101	0.080	0.155	0.100	0.080
	I	141.815	38.610	4.175	152.505	42.100	9.620	168.320	65.320	15.525
$\sigma=500$	CSP	0.189	0.099	0.080	0.188	0.098	0.081	0.184	0.105	0.080
	ASF	0.161	0.100	0.079	0.155	0.096	0.080	0.160	0.105	0.077
	I	66.505	8.830	3.350	88.275	17.945	7.840	100.610	21.765	12.620
$\sigma=1000$	CSP	0.186	0.102	0.080	0.185	0.102	0.079	0.190	0.105	0.079
	ASF	0.173	0.102	0.080	0.158	0.101	0.077	0.165	0.104	0.078
	I	16.485	4.270	3.050	26.470	9.050	7.605	33.240	14.920	11.380
$\sigma=2000$	CSP	0.183	0.099	0.810	0.180	0.098	0.082	0.185	0.099	0.080
	ASF	0.169	0.100	0.800	0.168	0.096	0.079	0.163	0.098	0.079
	I	8.615	3.520	3.010	11.280	7.920	6.025	17.485	13.340	9.745

本情况下 BCI 的实际应用。下一步，我们将针对多分类的 BCI 拓展 ASF，并研究相应的在线算法。

参 考 文 献

[1] Lemm S, Schafer C, and Curio G. BCI competition 2003-data set III: Probabilistic modeling of sensorimotor μ rhythms for classification of imaginary hand movements [J]. *IEEE Trans. on Biomedical Engineering*, 2004, 51(6): 1077–1080.

[2] Li Y Q and Guan C T. A semi-supervised SVM learning algorithm for joint feature extraction and classification in brain computer interfaces [C]. The 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, New York, USA, Aug.30–Sep.3, 2006: 2570–2573.

[3] Lemm S, Blankertz B, and Curio G, *et al*.. Spatio-spectral filters for improving the classification of single trial EEG [J]. *IEEE Trans. on Biomedical Engineering*, 2005, 52(9): 1541–1548.

[4] McFarland D J, Anderson C W, and Müller K R, *et al*.. BCI meeting 2005-workshop on BCI signal processing: Feature extraction and translation [J]. *IEEE Trans. on Neural and Rehabilitation Systems Engineering*, 2006, 14(2): 135–138.

[5] Hammon P S and deSa V R. Preprocessing and meta-

classification for brain- computer interfaces [J]. *IEEE Trans. on Biomedical Engineering*, 2007, 54(3): 518–525.

[6] Wang Y J, Zhang Z G, and Li Y, *et al.*. BCI competition 2003-data set IV: An algorithm based in CSSD and FDA for classifying single-trial EEG [J]. *IEEE Trans. on Biomedical Engineering*, 2004, 51(6): 1081–1086.

[7] Liao X, Yao D Z, and Wu D, *et al.*. Combining spatial filters for the classification of single-trial EEG in a finger movement task [J]. *IEEE Trans. on Biomedical Engineering*, 2007, 54(5): 821–831.

[8] Friedman J, Hastie T, and Tibshirani R. Additive logistic regression: A statistical view of boosting [J]. *The Annals of Statistics*, 2000, 28(2): 337–407.

[9] Blankertz B, Müller K R, and Curio G, *et al.*. The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials [J]. *IEEE Trans. on Biomedical Engineering,* 2004, 51(6): 1044–1051.

[10] Wei Q G, Gao X G, and Gao S K. Feature extraction and subset selection for classifying single-trial ECoG during motor imagery [C]. The 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, New York, USA, Aug.30–Sep.3, 2006: 1589–1592.

[11] Atkeson C G, Moore A W, and Schaal S. Locally weighted learning [J]. *Artificial Intelligence Review*, 1997, 11(15): 11–73.

[12] Yang J and Yang J Y. Why can LDA be performed in PCA transformed space [J]. *Pattern Recognition*, 2003, 36(12): 563–566.

[13] Sun Y J. Iterative RELIEF for feature weighting: algorithms, theories, and applications [J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007, 29(6): 1035–1051.

吕 俊： 男，1979 年生，博士生，研究方向为脑电信号处理、脑机接口.

谢胜利： 男，1958 年生，教授，博士生导师，研究方向为自动控制、盲信号处理.