

doi:10.11835/j.issn.1000-582X.2014.06.013

改进的 FCM 算法及其在脑电信号处理中的应用

余 炜^{1,2}, 万代立¹, 杨喜敬¹, 周 娅¹

(1. 昆明理工大学 信息工程与自动化学院, 云南 昆明 650500; 2. 成都工业学院 电气与电子工程系, 成都 611730)

摘 要: 针对脑电信号的常用识别方法都基于监督型分类算法, 需要一定数量的训练数据对分类器进行训练, 无法满足实时应用的要求。提出基于数据点密度大小和马氏距离的改进模糊 C-均值(FCM)非监督分类算法, 对 2003 年第二届 BCI 大赛脑电信号分类。首先采用经验模式分解(EMD)算法对脑电信号进行分解, 提取相应特征值, 再经改进的 FCM 算法对输入的特征值进行分类。实验结果证明了改进算法在脑电信号分类应用中的可行性和有效性。

关键词: 脑-机接口; 经验模式分解; 模糊聚类; 马氏距离

中图分类号: TN911.6

文献标志码: A

文章编号: 1000-582X(2014)06-083-07

An improved FCM algorithm and its application to EEG signal processing

YU Wei^{1,2}, WAN Daili¹, YANG Xijing¹, ZHOU Ya¹

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 2. Department of Electrical & Electronics Engineering, Chengdu Technological University, Chengdu 611730, China)

Abstract: Most of the popular EEG classifiers need to be supervised and their parameters have to be trained by a number of train data in advance. That's the reason why they cannot be used in the real-time circumstances. In this paper, a new FCM unsupervised classification algorithm is proposed which is based on the density size of data dot and mahalanobis distance. Then, the algorithm is used to classify the EEG signals from the database of the second session of 2003 BCI competition. The EMD algorithm is used to decompose the EEG and extract the characteristic values, and then these values are classified by the proposed FCM algorithm. The experimental results show the algorithm's feasibility and validity in the EEG classification field.

Key words: brain-computer interface (BCI); empirical mode decomposition (EMD); fuzzy clustering; Mahalanobis distance

脑-机接口(brain computer interface, BCI)技术形成于 20 世纪 70 年代(1973, Vidal)^[1], 作为一种新型的人机接口方式, 引起人们广泛的关注并得到迅速的发展。BCI 技术绕开了外周神经和肌肉组织, 直接建立了人脑同计算机和其他电子设备的信息交流和控制通路, 从而实现了无需语言和动作, 直接通过人脑想法操纵和控制外部设备, 增强肢体残疾患者与外界交流和沟通的能力^[2]。BCI 分为入侵和非入侵两种, 与前者相比后者脑电信号(electroencephalography, EEG)具有信息量大、时间分辨率高、设备便于携带及无创性等特点, 为脑-机接口提供了实际可行性^[3]。但由于 EEG 信号具有非平稳、信噪比低和统计特性不断变化等特

收稿日期: 2013-05-12

基金项目: 四川省教育厅重点项目(2013SZA0153); 四川省应用基础研究计划项目(2013SZZ026)。

作者简介: 余炜(1974-), 男, 重庆大学博士, 副教授, 主要从事通信与信号处理, 智能信息处理等方向研究。(Tel) 18608710817; (E-mail) cdtuyw@qq.com。

点^[4-5],其特征提取和信号判别、分类成为 BCI 的关键环节和技术难点之一。

BCI 领域主要的特征提取方法有^[2-3,6]:短时傅里叶变换法、功率谱密度估计法、模型参数法和小波分析法。以上方法各有其特点,但对于非平稳、数据长度较短的脑电信号,都存在一定的不足,如短时傅里叶变换特征信息单一,实际应用中分类准确率不高;功率谱密度估计法当信号数据较短时统计特性不明显,识别率低;模型参数法更适用于平稳信号分析;小波变换和小波包变换分解信号需要预设分层数和小波函数,不具备对信号自适应的分解能力等。针对上述情况,使用经验模式分解(empirical mode decomposition, EMD)提取信号的特征值,它是一种有效的,非线性和非平稳数据分析方法^[3],EMD 通过信号的时间尺度来获得本征波动模式,然后以此为依据用基于信号本身的基函数对信号进行分解。目的是将信号分解为一系列的表征信号特征时间尺度的固有模态函数(intrinsic mode function, IMF)。

BCI 脑电信号判别研究中目前使用较为广泛的方法主要包含有^[7-9]:线性判别分析(LDA)、人工神经网络(ANN)、支持向量机(SVM)。以上均为监督型方法,虽然在 BCI 信号判别领域取得了不错的分类效果,但上述判别方法在分类判别之前均需要一定数量的真实数据标签(训练集)来训练分类器,缺乏在线实时性,对于实际应用存在一定困难。聚类分析是多元统计分析的方法之一,作为模式识别中非监督识别的重要分支,其方法不需要样本先验知识,也无需训练样本。模糊 C-均值聚类(fuzzy C-means clustering, FCM)作为一种聚类算法,基于划分聚类,依据最小二乘原理,采用迭代法优化目标函数,最终得到每个样本的隶属度。FCM 分类无需训练,对于非平稳的生物医学信号具有较强的鲁棒性,因此其作为脑电信号的分类器具有一定优势^[10]。但经典的 FCM 只对球形结构聚类有一定优势,存在样本数据等划分趋势,并假定样本矢量的各维特征对聚类贡献大小相同,这对于真实非均匀、非对称的脑电数据样本,其聚类效果并不是很理想。由于数据点密度大小可以反映数据内部结构^[11-12],同时在一定程度上说明不同样本对聚类的形象不同,而且马氏距离计算仅与样本数目有关且能自适应调节数据几何分布,因此研究将样本数据的点密度大小归一化结果作为权值,用马氏距离代替 FCM 中的欧式距离,改进了 FCM 算法的聚类目标函数,提出了加权的马氏距离模糊 C-均值算法(WFCM-M),并应用到脑电信号分类中,提高了分类正确率。

1 经验模式分解及特征提取

1.1 经验模式分解(EMD)

EMD 是 Huang 等人提出的用于非线性、非平稳时间序列信号的分解算法,完全由数据驱动,具有自适应和多尺度特性。EMD 的本质是基于信号的时间尺度来获得本征波动模式,以此分解信号^[13]。目的是将复杂信号分解成有限个具有物理意义的固有模态函数(IMF)。IMF 满足以下 2 个约束条件

- 1) 数据段内的极值点和零点数目相等或最多相差一个;
- 2) 任意一点处,分别由局部极大值和局部极小值形成包络线的平均值为零。

给出一个信号 $x(t)$,有效的 EMD 算法总结如下^[13]

- 1) 求得信号 $x(t)$ 的所有极值点。在极值点间用三次样条函数进行插值,得到上下包络 e_{\max} 和 $e_{\min}(t)$;
- 2) 计算 2 个包络线的均值 $m_1(t) = [e_{\max}(t) + e_{\min}(t)]/2$,从信号中提取该均值,形成第一个 IMF 原型

$h_1(t) = x(t) - m_1(t)$;

3) 以 $h_1(t)$ 代替 $x(t)$,重复以上步骤,直到 $h_{1k}(t)$ 为一个 IMF,记作 $c_1(t) = h_{1k}(t)$ 把 $c_1(t)$ 从 $x(t)$ 中提取出来, $r_1(t) = x(t) - c_1(t)$ 同时令 $x(t) = r_1(t)$;

- 4) 重复上述步骤,直到 $R_N(t)$ 或 $C_1(t)$ 满足预设条件, $x(t)$ 的分解结束。

1.2 特征提取

对于脑电信号正确的分类不仅取决于分类器的性能,更与输入分类器的数据相关。有效的脑电信号特征提取对于分类效果有重要的影响,其中, Khorshidtalab 已提取脑电信号的绝对均值(mean absolute value, MAV)作为特征值应用于脑电信号分类^[15],并达到良好的实验效果。相关电位(event-related potential, ERP)是指人体感觉系统受某种刺激时在大脑皮层相应区域产生的电位变化。研究表明人脑想象左右手运动时产生的 ERP 信号在 C_3 和 C_4 两电极比较明显^[14],本实验对 C_3 、 C_4 记录脑电信号分别做 EMD 分解。如图 1~图 2 所示是分别对 C_3 、 C_4 同一时刻想象运动脑电记录数据的 EMD 分解,提取主要 IMF 分量的绝对均值作为特征值进行分类。

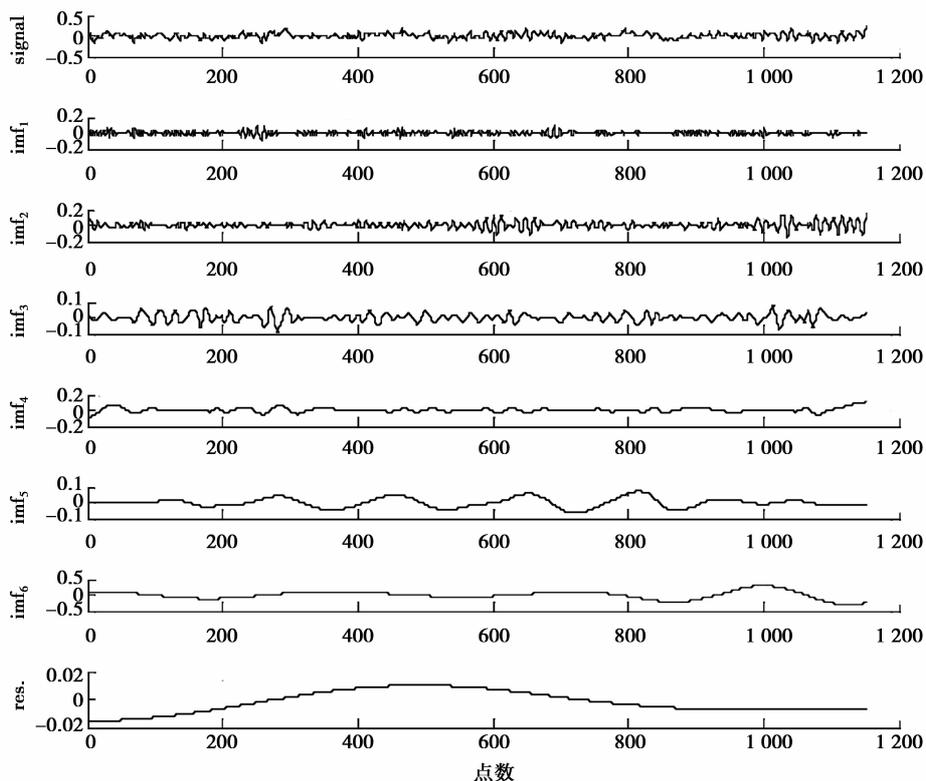


图 1 EMD 分解的 C₃ 脑电信号

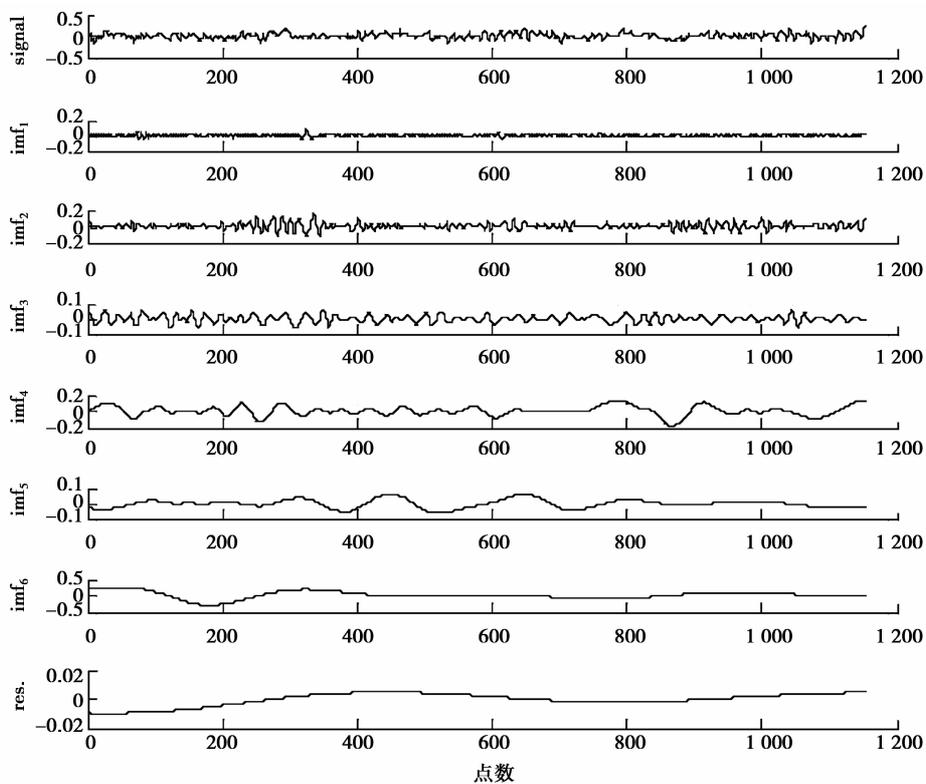


图 2 EMD 分解的 C₄ 脑电信号

绝对均值是指每段数据样本点绝对值的平均,如式(1),其中 $0 \leq i \leq N$, x_i 是第 i 段数据中的样本点, $0 \leq k \leq N$ 。

$$\text{MAV}_k = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (1)$$

2 加权的马氏距离模糊 C 均值聚类(WFCM-M)

经典模糊 C-均值算法(fuzzy c-means clustering, FCM)是通过不断的迭代计算,更新隶属度矩阵和聚类中心从而对目标样本自动分类。FCM 作为一种非监督聚类算法近年来广泛的应用于生物医学信号的分类和脑-机接口方面^[15]。然而,FCM 也存在初始聚类中心选择对于算法性能影响大,易陷入局部最优解,只对球形结构聚类有一定优势,对数据集等划分趋势,假设数据对象所有属性对于聚类结果贡献同等缺陷。

针对 FCM 的不足,提出改进加权的马氏距离的模糊 C 均值聚类方法(WFCM-M),通过计算样本数据的点密度大小^[17-18],归一化后作为权系数,使得不同样本对于数据分类的贡献有所差异。并且 S. Xiang 已经论述了马氏距离在分类中的优势^[19],利用马氏距离可以自适应的调整数据的几何分布,使得相似数据距离较小,从而解决了欧式距离对于数据属性相关且分布类似于高斯分布时分类正确率低的特点。因此,使用仅与样本数目相关的马氏距离替代 FCM 的欧式距离,改进了 FCM 聚类目标函数。设 \mathbf{X} 为 $l \times n$ 输入矩阵,对于每个样本点 x_i ,点密度定义为^[11-12]

$$D_i = \sum_{j=1, j \neq i}^l \frac{1}{d_{ij}}, 1 \leq i \leq l, \quad (2)$$

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|, 1 \leq i, j \leq l, \quad (3)$$

对 D_i 进行归一化

$$t_i = \frac{D_i}{\sum_{i=1}^l D_i}, 1 \leq i \leq l. \quad (4)$$

使用 t_i 表示第 i 个样本对于分类的影响程度。

样本 x_i 到总体 \mathbf{X} 的马氏距离定义为^[17-18]

$$d_m(\mathbf{x}_i, \mathbf{X}) = (\mathbf{x}_i - \mathbf{u})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \mathbf{u}), \quad (5)$$

其中 \mathbf{u} 为样本的均值向量, $\boldsymbol{\Sigma}$ 是协方差矩阵,定义为

$$\boldsymbol{\Sigma} = \frac{1}{l} \sum_{j=1}^l (\mathbf{x}_j - \mathbf{u})(\mathbf{x}_j - \mathbf{u})^T, \quad (6)$$

采用马氏距离替代模糊 C-均值中的欧式距离,并加入权系数,对应的目标函数表示为

$$\min J_{fcm}(U, C, \boldsymbol{\Sigma}) = \sum_{i=1}^c \sum_{j=1}^l t_j \mathbf{u}_{ij}^m (\mathbf{x}_j - c_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - c_i). \quad (7)$$

运用拉格朗日乘子法,得到如下公式不断更新协方差矩阵、聚类中心和隶属度矩阵,

$$\mathbf{v}_i = \frac{\sum_{j=1}^l t_j \mathbf{u}_{ij}^m \mathbf{x}_j}{\sum_{j=1}^l t_j \mathbf{u}_{ij}^m}, 1 \leq i \leq c. \quad (8)$$

$$\mathbf{u}_{ij} = \left[\sum_{k=1}^c (D_{ij}/D_{kj})^{2/m-1} \right]^{-1}, 1 \leq i \leq c, 1 \leq j \leq l. \quad (9)$$

其中,马氏距离的平方

$$D_{ij}^2 = (\mathbf{x}_j - c_i)^E \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - c_i), \quad (10)$$

$$\boldsymbol{\Sigma}_i = \frac{\sum_{j=1}^l \mathbf{u}_{ij}^m (\mathbf{x}_j - c_i)(\mathbf{x}_j - c_i)^T}{\sum_{j=1}^l \mathbf{u}_{ij}^m}, 1 \leq i \leq c, \quad (11)$$

对于协方差矩阵求逆时会遇到奇异的问题,Babuska 等人对于协方差矩阵的计算做了进一步的改进^[16]。

$$F_i = (1 - \gamma)F_i + \gamma \det(F_0)^{1/n} I, \gamma \in [0, 0]。 \quad (12)$$

其中 γ 是调节参数, F_0 是整个数据集的协方差矩阵。

3 实验及结果分析

3.1 数据描述

所用真实脑电数据来自 2003 年第二届国际 BCI 大赛(Graz 科技大学提供)。该实验由反馈信号控制的在线 BCI 系统完成,所执行的意识任务是通过想象左右手运动来控制光标的左右移动。本实验中信号处理的任务就是通过对脑电信号数据的在线分析和分类,在一定的时间内识别出受试者的意识任务是想象左手运动还是右手运动。实验详细内容及参数见文献[14]。

3.2 实验

为验证方法的有效性,对上述脑电数据进行分类。由于人脑想象左右手运动时的 ERD/ERS 在 C_3 和 C_4 两电极比较明显,选取 C_3 、 C_4 左右共 100 组数据分别进行 EMD 分解^[14]。对分解后的 IMF 分量进行傅里叶变换从而获得频谱能量,得到主要能量集中在前 2 个 IMF 分量中,使用此方法避免了传统的单凭经验获取主要成分的缺陷,之后提取前 2 个 IMF 分量的绝对均值作为特征值,表 1 所示为 C_3 、 C_4 电极部分脑电数据的 IMF 分量特征值。

表 1 一部分脑电数据特征值

参量		C_3		C_4	
	1	0.015 78	0.029 78	0.022 19	0.027 13
	2	0.023 49	0.028 50	0.024 71	0.042 41
MAV	3	0.044 94	0.039 64	0.023 78	0.029 92
	4	0.021 79	0.020 99	0.026 96	0.033 33
	5

针对上述特征向量,使用 WFCM-M 进行分类,其中设置模糊加权数 $m=2$,聚类数目为 2,迭代停止阈值 $\epsilon=10^{-5}$,权系数、聚类中心和隶属度矩阵如表 2 所示。

表 2 WFCM-M 权系数、聚类中心和隶属度

权系数	0.009 268 696	0.011 309 466	...	0.009 913 2
聚类中心	V1:0.027 838	0.033 627	0.029 954	0.043 059
	V2:0.026 795	0.032 464	0.029 104	0.041 490
隶属度	0.506 950	0.506 950	...	0.463 300
	0.493 050	0.493 050	...	0.536 700

按照最大隶属度原则进行分类,误分率为 17%。对上述特征向量采用经典 FCM 方法进行分类,误分率为 22%,误分率要明显高于聚类方法。随后对于 BCI 数据不作 EMD 分解而直接提取样本的绝对均值作为特征值分别使用 WFCM-M 和经典 FCM 分类方法对相同的数据集分类,4 种方法分类结果如表 3 所示,在同样使用经验模式分解对相同脑电数据分解提取特征值的前提之下,WFCM-M 方法优于经典 FCM。同时,都

是在非监督的条件下对脑电数据进行分类,经过 EMD 对脑电数据分解后的结果要优于直接提取特征值进行分类,以上结果验证了提出的 BCI 分类方法的有效性。

表 3 脑电数据误分率

方法	EMD WFCM-M	EMD FCM	WFCM-M	FCM
误分率	17%	22%	31%	39%

4 结 论

基于脑电信号非平稳、复杂、信噪比低的特性,使用非平稳、非线性的经验模式分解算法(EMD)对想象左右手脑电数据进行分解,提取包含主要能量的前 2 个 IMF 分量的绝对均值(mean absolute value)作为特征值,并使用非监督型加权的马氏距离模糊 C-均值(WFCM-M)对数据分类,与现有的几种聚类方法进行了对比研究。实验结果表明,提出的方法对于 2003 年第二届 BCI 大赛脑电信号的误分率较低,所提出的改进的 FCM 算法对于生物医学信号处理领域的非平稳、非线性、弱小信号的分类和鉴别有一定的优势,可以扩展至其它信号处理领域。

参考文献:

- [1] Wolpaw J R, Birbaumer N, McFarland D J. Brain-computer interface for communication and control [J]. *Clinical Neurophysiology*, 2002, 113(6): 767-791.
- [2] 徐宝国, 宋爱国, 王爱民. 基于小波包能量的脑电信号特征提取方法[J]. *东南大学学报*, 2010, 40(6): 1203-1206.
XU Baoguo, SONG Aiguo, WANG Aimin. EEG feature extraction method based on wavelet packet energy[J]. *Journal of South East University*, 2010, 40(6): 1203-1206.
- [3] 王攀, 沈继忠, 施锦河. 想象左右手运动的脑电特征提取[J]. *传感技术学报*, 2010, 23(9): 78-82.
WANG Pan, SHEN Jizhong, SHI Jinhe. Feature extraction of EEG for imagery left-right hands movement[J]. *Chinese Journal of Sensors and Actuators*, 2010, 23(9): 78-82.
- [4] Gandhi T, Panigrahi B K, Anand S. A comparative study of wavelet families for EEG signal classification [J]. *Neurocomputing*, 2011, 74(17): 3051-3057.
- [5] Shenoy P, Krauledat M, Blankertz B R, et al. Towards adaptive classification for BCI[J]. *Journal of Neural Engineering*, 2006, 3(1): 13-23.
- [6] 应基光, 王行愚. 基于 EMD 的左右手运动想象脑电信号分析研究[J]. *生物医学工程学进展*, 2009, 30(3): 125-130.
YING Jiguang, WANG Xingyu. Analysis of right and left motor imagery based on EMD[J]. *Progress in Biomedical Engineering*, 2009, 30(3): 125-130.
- [7] Shimo N, Pang S N, Kasabov N, et al. Curiosity-driven multi-agent competitive and cooperative LDA learning [J]. *International Journal of Innovative Computing, Information and Control*, 2008, 4(7): 1537-1552.
- [8] 赵海滨, 王宏. 基于功率谱估计的神经网络的脑-机接口研究[J]. *系统仿真学报*, 2007, 19(19): 4581-4583.
ZHAO Haibin, WANG Hong. Power spectrum estimation based on neural network Brain - Computer Interface [J]. *Journal of System Simulation*, 2007, 19(19): 4581-4583.
- [9] Chen B H, Hu J L. An improved multi-label classification method based on SVM with delicate decision boundary [J]. *International Journal of Innovation Computing, Information and Control*, 2010, 4(6): 1605-1614.
- [10] Hsu W Y, Lin C Y, Kuo W F, et al. Unsupervised fuzzy c-means clustering for motor image EEG recognition [J]. *International Journal of Innovation Computing, Information and Control*, 2011, 7(8): 4965-4975.
- [11] 刘小芳, 曾黄麟, 吕炳朝. 点密度函数加权模糊 C-均值算法的聚类分析[J]. *计算机工程与应用*, 2004, 24: 64-66.
LIU Xiaofang, ZENG Huanglin, Lü Binchao. Clustering analysis of dot density function weighted fuzzy C-means algorithm [J]. *Computer Engineering and Applications*, 2004, 24: 64-66.

- [12] 楼晓俊,李隽颖,刘海涛. 距离修正的模糊 C 均值聚类算法[J]. 计算机应用,2012,32(3):646-648.
LOU Xiaojun, LI Junying, LIU Haitao. Improved fuzzy C-means clustering algorithm based on distance correction[J]. Computer Applications, 2012, 32(3): 646-648.
- [13] Park C, Looney D, Kidmose P, et al. Time-frequency analysis of EEG asymmetry using bivariate empirical mode decomposition [J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2011, 19(4): 366-373.
- [14] 刘冲,赵海滨,李春胜. 基于 CSP 与 SVM 的运动想象脑电信号分类[J]. 东北大学学报:自然科学版, 2010, 31(8): 1098-1101.
LIU Chong, ZHAO Haibin, LI Chunsheng. Based on the CSP and SVM classification the movement of the EEG signals [J]. Journal of Northeastern University: Science Edition, 2010, 31(8): 1098-1101.
- [15] Khorshidtalab A, Salami M J E, Hamed M. evaluation of time-domain features for motor image movements using FCM and SVM[C]// Proceedings of 2012 International Joint Conference on Computer Science and Software Engineering, May 30-June 1, 2012, Bangkok, Piscataway: IEEE Press, 2012: 17-22.
- [16] Babuska R, Veen P J, Kaymak U. Improved covariance estimation for Gustafson-Kessel clustering[C]// Proceedings of 2012 IEEE International Conference on Fuzzy Systems, 2002, Honolulu, Hi. Piscataway: IEEE Press, 2012: 1081-1085.
- [17] 蔡静颖,谢福鼎,张永. 基于自适应马氏距离的模糊 C 均值算法[J]. 计算机工程与应用, 2010, 46(34): 174-176.
CAI Jingying, XIE Fuding, ZHANG Yong. Mahalanobis distance based on adaptive fuzzy C-means algorithm[J]. Computer Engineering and Applications, 2010, 46(34): 174-176.
- [18] 王建英 孙德山 张永. 基于马氏距离的 FCM 图像分割算法[J]. 计算机工程与应用, 2010, 46(1): 147-149.
WANG Jianying, SUN Deshan, ZHANG Yong. Mahalanobis distance-based FCM image segmentation algorithm [J]. Computer Engineering and Applications, 2010, 46(1): 147-149.
- [19] Xiang S M, Nie F P, Zhang C S. Learning a mahalanobis distance metric for data clustering and classification[J]. Pattern Recognition, 2008, 41(12): 3600-3612.

(编辑 侯 湘)