

用基于平行坐标图的可视化人机交互技术提取 EEG 信号特征的方法^①高海波^{②*} 崔建新^{**} 洪文学^{*} 郝连旺^{*}

(* 燕山大学电气工程学院 秦皇岛 066004)

(** 河北省测试计量技术及仪器重点实验室 秦皇岛 066004)

摘要 应用可视化人机交互(HCI)方法进行了脑电图记录(EEG)信号特征提取技术的研究。该研究一方面在脑机接口(BCI)技术领域提出了一种新的特征提取技术方法,同时通过可视化人机交互的专家智慧参与,实现了面向对象领域和面向数据模式识别的有效结合,克服了单一机器学习的局限性。首先介绍了多元图表示的基本理论,然后提出了基于平行坐标图的可视化人机交互技术,接着进行了单通道和多通道 EEG 信号特征提取的可视化人机交互技术的研究,最后采用第二届国际脑机接口竞赛中的数据集 IV 进行了数据实验。实验表明,本文提出的方法的识别结果优于实验数据集国际竞赛最优结果和文献报道中的当前国际最优结果。

关键词 脑机接口(BCI), 模式识别, 特征提取, 可视化, 人机交互(HCI), 平行坐标图

0 引言

人的大脑是一个极其复杂的系统,研究人的思维机理,实现人与周围环境的信息交换是一项重要的科研课题。脑机接口(brain-computer interface, BCI)是一种通信系统,它不依赖于由外围神经和肌肉组成的正常输出通路^[1]。BCI 将人脑和计算机有机地整合在一个系统中,其研究成果在医疗仪器领域中将能够帮助患者(如脑干或脊髓严重损伤)恢复机体控制和外界交流^[2],还可以在传统控制方式不能满足要求时(如手术、航空领域)提供一种新的控制方式^[3]。为实现 BCI 功能,需要实时监测大脑的意识活动。目前可用的手段很多,但脑电图记录(electroencephalography, EEG)是目前实现 BCI 的唯一可行方式^[4]。因此如无其他特殊说明,当前的 BCI 系统通常就是指基于 EEG 的 BCI 系统。EEG 检测技术相对比较成熟,目前基于 EEG 的 BCI 核心技术研究主要是围绕 EEG 的信号处理、模式识别和特征提取等方面。

特征提取是 BCI 分类研究中的关键问题。当前研究中提出了很多 EEG 信号特征提取方法,其中最突出的是事件相关电位。其它如文献[5]在脑机接口领域应用小波包分解进行脑电信号的特征提取;

文献[6]针对脑电信号的非高斯(non-Gaussian)特性采用谱分析方法进行特征提取,并应用线性分类器、支持向量机分类器和神经网络分类器进行脑机接口模式分类实验研究。基于机器学习的计算机自动分类是当代模式识别领域研究的主题,在其倡导的最小用户参与准则下形成了分类效率和效果、精度和推广性之间的矛盾^[7]。这暴露了单一机器学习固有的局限性,计算机自动模式识别过程中适当的人工参与是必要的和重要的。本文进行了基于多元图表示的可视化人机交互(human-computer interaction, HCI) BCI 特征提取技术研究,通过模式识别过程的专家智慧参与,实现了面向对象领域和面向数据模式分类方法的有效结合,克服了单一机器学习认知的局限性^[8]。这里的技术研究及数据实验均以 2003 年国际脑机接口竞赛中的数据集 IV^[9]为对象。在实验中,本文提出的特征提取技术和方法取得了较好的分类识别结果。

1 基于平行坐标图表示的可视化人机交互技术

数图结合是科学研究的重要思想和方法。模式识别中的样本空间从数学上可以描述为实数矩阵,矩阵的行代表样本,列代表特征。传统方法一般把

① 国家自然科学基金(60605006),河北省教育厅科研计划自然科学重点项目(ZH200802),河北省科技支撑计划项目(072135220)和燕山大学博士基金(2010498)资助项目。

② 男,1975年生,博士,讲师;研究方向:模式识别,多元信息处理,脑机接口;联系人,E-mail:hhghb@ysu.edu.cn (收稿日期:2009-04-01)

一个样本视为多维欧式空间中的一个点,点与点的关系描述了样本之间的关系(如距离关系等)。这种方法存在的明显问题是高于 3 维的数据不能可视化,这样很难直观地揭示高维样本点的内部数据结构。多元图表示提供了一种将高维数据向二维平面上映射表示的一种方法,这为高维特征的可视化提供了可能。多元图表示的具体方法很多,如散点图、星座图、雷达图、脸谱图等,这里只介绍本文技术研究中采用的平行坐标图。

1.1 多元图表示数学模型

设所要研究的对象样本数为 N , 样本包含的变量数为 M , 整个多维数据可以表示为 N 行 M 列的数据矩阵

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix} \quad (1)$$

其中 x_{nm} 为矩阵元素 ($n = 1, 2, \dots, N; m = 1, 2, \dots, M$)。多元图表示过程在数学上可以描述成将多维数据的 M 个笛卡尔坐标值(分量)用二维复平面中的 M 个复数(M 个点, 或者 M 个矢量, 或者连接这 M 个点的曲线等)来表示。这样, 多元图映射就成为从实数域 \mathbf{R} 到复数域 \mathbf{C} 的一种映射集 $\phi = \{f_1, f_2, \dots, f_M\}$, 其中 M 为数据样本变量的个数。该过程可表达为

$$f_m(x) = f_{m1}(x) + f_{m2}(x) \cdot j \quad m = 1, 2, \dots, M \quad (2)$$

其中 f_{m1} 和 f_{m2} 为实变函数, j 表示虚数单位。这样, 可以基于以上多元图映射公式得到多元图表示矩阵

$$\mathbf{G} = \begin{bmatrix} f_{11}(x_{11}) + j \cdot f_{12}(x_{11}) & f_{21}(x_{12}) + j \cdot f_{22}(x_{12}) \\ f_{11}(x_{21}) + j \cdot f_{12}(x_{21}) & f_{21}(x_{22}) + j \cdot f_{22}(x_{22}) \\ \vdots & \vdots \\ f_{11}(x_{N1}) + j \cdot f_{12}(x_{N1}) & f_{21}(x_{N2}) + j \cdot f_{22}(x_{N2}) \\ \cdots & f_{M1}(x_{1M}) + j \cdot f_{M2}(x_{1M}) \\ \cdots & f_{M1}(x_{2M}) + j \cdot f_{M2}(x_{2M}) \\ \vdots & \vdots \\ \cdots & f_{M1}(x_{NM}) + j \cdot f_{M2}(x_{NM}) \end{bmatrix} \quad (3)$$

\mathbf{G} 中的行表示样本, 每行的 M 个分量对应二维复平面中 M 个点的复数(矢量)表示形式。矩阵中每个元素(复数)对应二维复平面中的一个点, 该元素的实部为该点的横坐标, 其虚部为该点的纵坐标。

对于公式(1)所示多维数据矩阵 \mathbf{X} , 其平行坐标图表示的映射公式为

$$f_{m1}(x_{nm}) = md, f_{m2}(x_{nm}) = x_{nm} \quad (4)$$

$$f_m(x_{nm}) = md + x_{nm} \cdot j \quad (5)$$

其中 d 为一个正的常数, 表示坐标间的间隔距离。

1.2 基于平行坐标图的可视化人机交互技术

根据式(1)-(5), 容易对样本特征进行平行坐标图表示。在平行坐标图中对不同类别样本用不同线条(可以用线条颜色或线型来区分)表示, 那么很容易直接对每个变量特征进行单变量类别可分性分析^[10]。如图 1 为 Iris 数据集的平行坐标图表示, 其 3 类样本的期望和变化方差分别用不同符号表示并用不同粗细的实线和虚线表示, 其中类别 1 用 * 表示并用 2.5 磅粗线连接, 类别 2 用 ● 表示并用 1.5 磅中度粗细线连接, 类别 3 用 ○ 表示并用 0.5 磅细线连接。显然四个特征变量中变量 3 和变量 4 具有较好的类别区分特性。

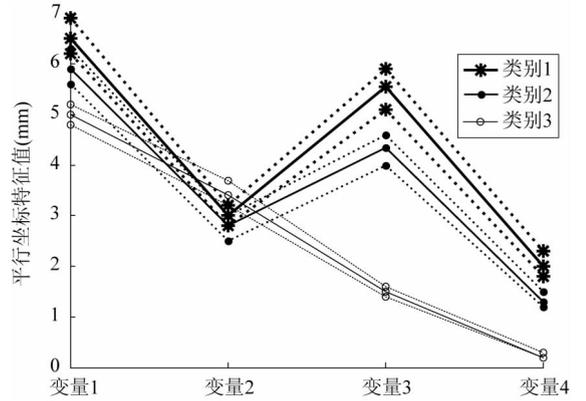


图 1 Iris 数据集的平行坐标图表示

当然, 也可以对样本数据的各种统计特征(如期望、方差等)进行平行坐标图表示实现数据样本的类别可分性分析, 下面将专门介绍, 此处不述。

2 基于平行坐标图表示的人机交互 BCI 特征提取

直接对 EEG 时间序列信号进行模式识别一般很难取得较好的分类效果。应用平行坐标图表示的可视化技术进行了单通道和多通道 EEG 可视化的 BCI 特征提取技术研究。EEG 时间序列信号为高维特征, 经过特征提取后形成的特征空间维数一般还很高, 这里采用改进的遗传算法进行特征选择。

2.1 基于单通道时间序列期望变化异步分析的 BCI 特征提取

要从各通道的 EEG 信号中提取出有利于分类的特征信息, 需从分析各通道的信号规律入手。经过大量可视化分析, EEG 信号的变化非常剧烈, 很难

从各通道样本个体变化中发现任何类别规律。

对各信道脑电时间序列信号(如 F3 通道)的训练样本类别期望进行平行坐标图表示如图 2 所示。对其进行可视化分析发现,在脑电信号变化过程中存在大量的局部异步变化现象。直接通过脑电信号的差异进行分类非常困难,可以通过训练样本类均

值时间序列数据的可视化分析提取脑电信号变化过程中的局部异步变化作为分类特征。基于以上分析提出基于单通道脑电时间序列信号的期望局部变化异步分析的 BCI 特征提取技术。对于图 2 中的 F3 信道脑电时间序列信号进行局部参数为步长 1 的异步变化 BCI 特征提取如该图所示。

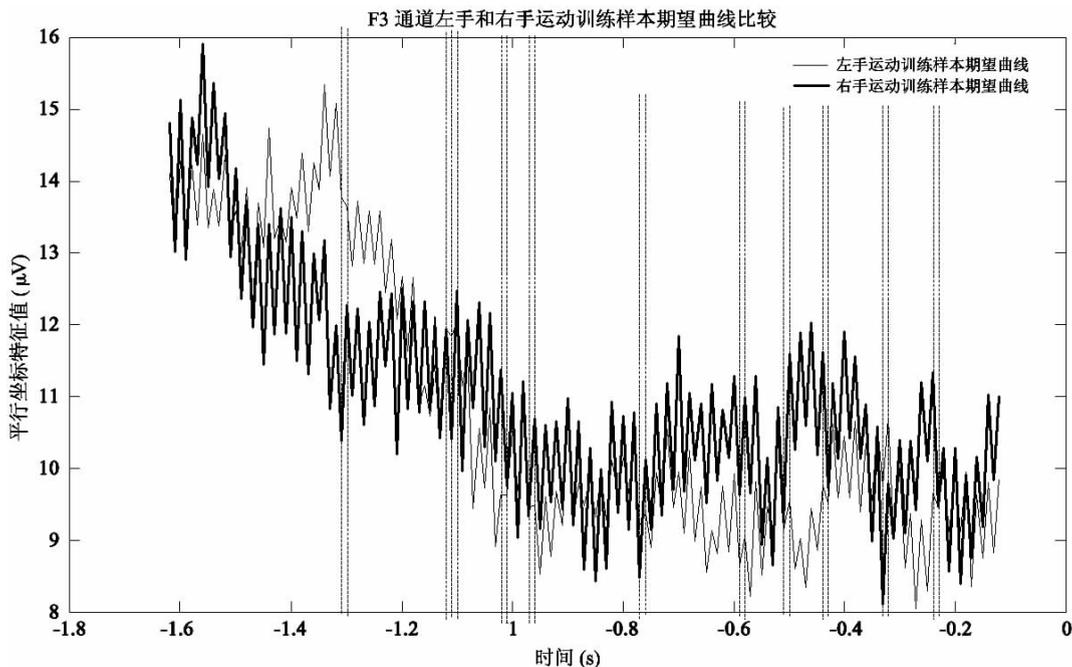


图 2 基于单通道 EEG (F3) 时间序列局部期望异步变化分析的 BCI 特征提取

图中每一对竖线标识处的两类样本均值变换规律具有明显差异,往往一类急剧增大另一类急剧减小,若以标识组中前后信号幅值比作为特征,显然两类数据之间具有明显差异而将有利于分类。当然,这里的类间局部异步变化特征是基于训练样本类均值平行坐标图可视化分析得到,每一点处的均值变化并不一定都能够完全反映所有样本点的变化规律。故并不是该方法提取的所有特征都具有良好的分类性能,所形成的特征空间表示还需要基于类分性准则的特征选择。当然,基于单通道脑电时间序列信号的期望局部变化异步分析的 BCI 特征提取过程中局部参数不一定为步长 1。该特征提取过程容易通过算法实现,此处不述。

2.2 基于多通道数据融合的可视化 BCI 特征提取

通过对各通道脑电时间序列数据样本的可分性进行可视化分析发现,各通道脑电原始特征的分类能力是非常有限的。当进行不同的思维活动时头部皮层不同的(两类样本不同通道)脑电信号之间的关系也可能具有有利于分类的作用。基于此开展基于

多通道数据融合的可视化 BCI 特征提取研究。将各通道脑电时间序列数据两两做差并将结果进行样本可分性的可视化分析,发现该方法能够获得分类性能非常好的融合特征表示。

例如将 C3 通道与 C2 通道的脑电时间序列数据做差得到的融合特征的可分性进行基于平行坐标图表示的可视化分析如图 3 所示,其中左手运动样本用细线表示,右手运动样本用粗线表示;实线表示两类样本的期望特征曲线,虚线表示两类样本距期望特征一倍标准差范围曲线。由图 3 中可见,由 C3 通道与 C2 通道脑电数据通过数据融合的方法得到的特征表示具有相对于任何通道的原始脑电数据更为优秀的类别可分性,而且时间序列中数据的可分性随着时间越来越好,即后段的时间序列特征具有更好的分类性能,这为分类研究中的进一步特征选择提供了参考依据。应用该方法还可以对其它的多通道融合结果进行可视化分析,进而挖掘更有利于模式辨识分类的特征。

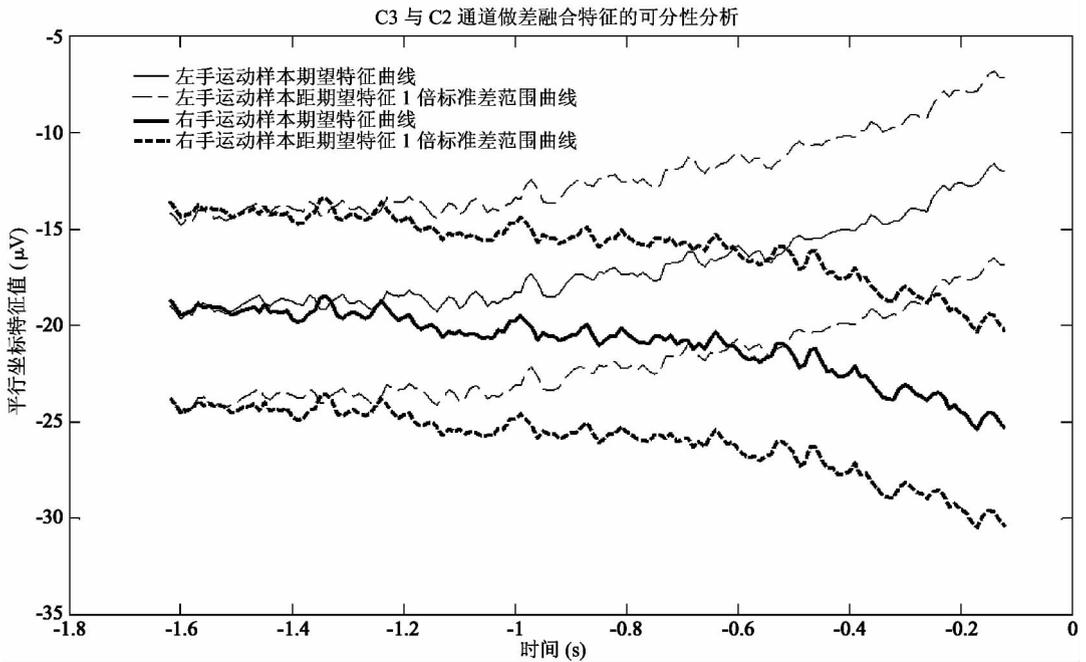


图3 C3与C2通道EEG融合特征样本可分性的可视化分析

2.3 基于改进遗传算法的特征选择

遗传算法(genetic algorithms, GA)是对生物系统里达尔文的自然选择和遗传理论进行直接模拟的一个自适应优化搜索算法,优于传统的启发式搜索方法。GA工作在称为一个种群的一组有效解上,基于达尔文的适者生存的原则,在一系列的迭代遗传操作后,GA获得优化解。

基于遗传算法的特征选择关键问题是如何将问题的解编码为染色体。基于整数编码策略,本文提出一种具体的随机编码方法。假设结果特征提取形成 n 维 BCI 特征空间表示,这样对应遗传算法中的基因个数为 n 。将其作为原始特征进行整数编码并随机生成种群基因,设种群基因数目为 $numgene$, 其范围为 $1 \leq numgene \leq n$ 。将种群基因进行整数编码,随机基因序列生成染色体,设染色体数目为 $numchromosome$ 。若染色体包含基因数目限定为 num , 那么生成的染色体数目范围为 $1 \leq numchromosome \leq P_{numgene}^{num}$;若染色体包含基因数目没有限定,那么生成的染色体数目范围为 $1 \leq numchromosome \leq \sum_{num=1}^{numgene} P_{numgene}^{num}$ 。经过以上两级整数编码和随机生成形成包含染色体数目为 $numchromosome$ 的随机种群。

随机种群中的染色体本质上是特征提取的 BCI 特征空间的子空间,以其分类准确率作为种群进化中染色体的生存适应度。基于达尔文的适者生存的

原则,染色体的生存适应度 level 若大于生存极限值 limit 则作为优选染色体进入下一代,若染色体的生存适应度小于生存极限值则被淘汰。若种群中所有染色体在第 1 代就全被淘汰,则种群灭亡,随机生成新的种群重新进化。经过生存竞争种群进入下一代进化,为更大程度保存上一代进化成果,上一代的优选染色体直接成为下一代种群个体,称为克隆。上一代生存竞争淘汰一部分染色体形成种群空区,该部分通过上一代进化中的优选染色体交叉和变异,以及新染色体入侵三种方式补充,该过程称为繁衍。优选染色体交叉是指通过优选染色体之间的基因交换形成新的染色体,该过程没有新的基因引入。优选染色体变异是指优选染色体的部分基因被新的基因替代形成新的染色体,该过程引入新的基因进入种群。新染色体入侵是指完全由新的基因组成的新的随机染色体进入种群进化,该过程最大程度地引入新的种群基因,增加种群生存活力。三种染色体补充方式的控制比例一般为:若上代优选染色体组生存活力值高,交叉方式比例提高,反之提高入侵比例。染色体生存活力用其生存适应度 level 与生存极限值 limit 之比表征,那么优选染色体组生存活力值为全部优选染色体生存活力的平均值。新一代种群通过生存竞争进化选择出新一代优选染色体组。新一代优选染色体组与上一代优选染色体组的生存活力值之比作为进化趋势指标,若该指标 μ 大于 1 则为良性进化,调整生存极限值 limit 为 $\mu^* \text{ limit}$ 继

续进化;若该指标 μ 小于 1 则为不良进化,停止种群遗传进化过程,保存最优染色体作为本次进化选择结果。整个遗传进化可选择种群进化代数或理想染色体目标为中止条件。

3 实验结果及分析

这里的数据实验采用 2003 年国际脑机接口竞赛中的数据集合 IV^[9],该数据集样本的每个样本数据为 27(F3, F1, Fz, F2, F4, FC5, FC3, FC1, FCz, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6, O1)通道脑电信号,包括左、右手手指的运动两类,样本量为 516,其中 3 个样本为类别未知。

为方便和文献中国际上该数据集分类实验结果

对比,实验中同样采用前 413 个为经验样本,其余 103 个为测试样本。分类器选用 3 种最简单的经典分类器,包括贝叶斯分类器(Bayesc),线性判别分析分类器(ldc)和 K 近邻分类器(knnc),其中 K 近邻分类器选择近邻参数为 1,距离参数选择欧氏距离。作为对比,选择所有通道原始 EEG 数据的主成分分析(PCA)特征提取方法作为参考。

实验结果如表 1 所示,由实验结果可以看出,本文提出的基于平行坐标图表示的人机交互 BCI 特征提取方法明显优于经典的主成分分析(PCA)特征提取方法,其中基于多通道数据融合的可视化 BCI 特征提取方法的分类结果优于基于单通道脑电时间序列信号的期望局部变化异步分析的 BCI 特征提取方法。

表 1 BCI 分类实验结果

脑电时间序列	特征提取方法	Bayesc	ldc	knnc
全通道	PCA 特征提取	0.39	0.35	0.43
全通道	基于单通道期望局部变化异步分析的特征提取	0.14	0.17	0.21
C3-C2	两通道数据融合的可视化特征提取	0.07	0.11	0.17

表 2 给出了 2003 年国际脑机接口竞赛中的数据集合 IV 的国际竞赛结果^[11]、文献[12]报道的当前国际上该数据集最佳分类错误率性能和本文基于平行坐标图表示的可视化特征提取的分类实验最佳分类

错误率进行的性能比较。从中可见,本文基于多元图表示的可视化技术在 BCI 分类研究中具有明显的积极作用。

表 2 本文 BCI 分类实验结果和文献报道性能比较

分类实验结果来源	采用方法	分类错误率
2003 年国际脑机接口竞赛 第一名	字空间分解 + fisher 判别准则	0.16
2003 年国际脑机接口竞赛 第二名	探索性数据分析	0.19
2003 年国际脑机接口竞赛 第三名	基于 Fourier 变换系数的 PCA 特征提取	0.23
2003 年国际脑机接口竞赛 前 15 名统计:平均(极值)	--	0.30(0.16-0.49)
文献[12]报道的当前国际最佳分类结果	--	0.09-0.10
本文分类实验	基于平行坐标图表示的可视化特征提取	0.07

4 结论

模式识别和特征提取是脑机接口技术研究中的重要环节。实验证明,本文进行的基于多元图表示的可视化特征提取技术对于高维 EEG 时间序列数据的分类特征挖掘具有明显效果,有利于 BCI 技术的发展和提高。

当然,作为一般的模式识别技术方法,基于多元

图表示的可视化特征提取技术的适用领域不仅仅局限于基于 EEG 的脑机接口研究领域,它在其它领域的应用技术研究也是值得期待和必要的。

参考文献

[1] Wolpaw J R, Birbaumer N, McFarland D J, et al. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 2002,113(6): 767-791
 [2] Gao X R, Xu D F, Cheng M, et al. A BCI-based environmental controller for the motion-disabled. *IEEE Trans Neural*

- Syst Rehab Eng*, 2003, 11(2): 137-140
- [3] 杨帮华, 颜国正, 颜荣国. 脑机接口研究进展. 中国医疗器械杂志, 2005, 29(5): 353-357
- [4] Mehrdad F, Ali B, Ward R K, et al. EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, 2007, 118(3): 480-494
- [5] Wu T, Yan G Z, Yang B H, et al. EEG feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement*, 2008, 41(6): 618-625
- [6] Shangming Zhou, John Q G, Francisco S. Classifying mental tasks based on features of higher-order statistics from EEG signals in brain computer interface. *Information Sciences*, 2008, 178(6): 1629-1640
- [7] Ricardo G O. Pattern analysis for machine olfaction: a review. *IEEE Sensors Journal*, 2002, 2(3): 189-202
- [8] 高海波. 基于多元图图形基元和特征基元表示的模式识别方法研究:[博士学位论文]. 秦皇岛:燕山大学电气工程学院, 2009. 85-93
- [9] Fraunhofer. The data set IV of BCI competition II. http://ida.first.fraunhofer.de/projects/bci/competition_ii/berlin_desc.html; Fraunhofer FIRST, 2010
- [10] Gao H B, Hong W X, Cui J X, et al. Pattern recognition of multivariate information based on non-statistical techniques. In: Proceedings of 2008 IEEE International Conference on Information and Automation, Zhangjiajie, China, 2008. 697-702.
- [11] The Organizing Committee for the international BCI Competitions. The Final Results of BCI Competition II. <http://www.bbci.de/competition/ii/results/#beilin>; BBCI, 2010
- [12] Cplee. Official Description: overview of classification task of EEG Classification for Self-paced Key Typing, <http://home.ustc.edu.cn/~cplee/>; 中国科技大学, 2009

Feature extraction of EEG signals using the visualized BCI technique based on parallel coordinates plot

Gao Haibo^{*}, Cui Jianxin^{**}, Hong Wenxue^{*}, Hao Lianwang^{*}

(^{*} Institute of Electrical Engineering Yanshan University, Qinhuangdao 066004)

(^{**} Measurement Technology and Instrumentation Key Lab of Hebei Province, Qinhuangdao 066004)

Abstract

The study on the techniques for extraction of electroencephalography (EEG) signal features was conducted based on human-computer interaction (HCI). It proposes a new way of feature extraction in the technical field of brain-computer interface (BCI), and through the expert intelligence endeavor to feature extraction by the HCI based on graphical presentation of multivariate data, effectively realized the combination of data-oriented pattern recognition and object-oriented domains, and overcame the obstacles of the only machine learning. It firstly introduced the visualized HCI technique based on graphical presentation of multivariate data, then studied the visualized feature extraction techniques for single channel and multi-channel EEG signals. The experiments were performed based on the dataset IV of the international BCI competition II. The experimental results were very superior to that of the international BCI competition II and the previously reported optimal classification performance of the international BCI competition II compared baseline methods. It proved the validity of the research methods in this paper.

Key words: brain-computer interface (BCI), pattern recognition, feature extraction, visualized, human-computer interaction (HCI), parallel coordinates plot