

基于信息约简的脑电信号特征提取**

穆振东, 肖 丹

Feature extraction of electroencephalogram based on data reduction

Mu Zhen-dong, Xiao Dan

Institute of
Information and
Technology, Jiangxi
Blue Sky University,
Nanchang 330098,
Jiangxi Province,
China

Mu Zhen-dong★,
Master, Lecturer,
Institute of
Information and
Technology, Jiangxi
Blue Sky University,
Nanchang 330098,
Jiangxi Province,
China
zd_mu@sohu.com

Supported by: the
Scientific Research
Fund of Jiangxi
Provincial Education
Department, No.
GJJ08477*

Received: 2009-10-15
Accepted: 2009-12-18

Abstract

BACKGROUND: Reaction velocity for brain imagination is a major criterion in measuring quality of brain computer interface (BCI) system. Therefore, electroencephalogram (EGG) analysis, especially the feature extraction screening, is very important. Reduction the number of features is an important way to improve the speed.

OBJECTIVE: To screen EGG features using reduction algorithm, and to decrease the number of EGG features.

METHODS: Firstly, by various analytical methods, the EGG features were extracted and classified; then, discretizing the continuous EGG to establish an EGG information table; At last, reduction theory was used to reduce EGG features, and to sort the data according to reduced attributes, the accuracy of sorting was validated.

RESULTS AND CONCLUSION: Using reduction algorithm and feature marking, discretizing the continuous EGG to establish an EGG information table, and then choose the features from the discrete data. The results show that classification accuracy has not been reduced but the number of features was reduced. However, this method can be used marking features in two sorts, how to marking features of multi-sorts and to perform data reduction is a key point in further study.

Mu ZD, Xiao D. Feature extraction of electroencephalogram based on data reduction. Zhongguo Zuzhi Gongcheng Yanjiu yu Linchuang Kangfu. 2010;14(9): 1642-1644. [http://www.crter.cn http://en.zgckf.com]

摘要

背景: 脑机接口系统中, 对大脑想象的反应速度是衡量脑机接口系统质量的一个重要标准, 因此关于脑电信号的分析, 对脑电信号特征进行筛选就显得尤为重要, 尽可能少的特征来标示相应的动作是提高速度的一个重要途径。

目的: 利用约简方法对脑电信号特征进行特征筛选, 降低脑电特征数目。

方法: 首先利用不同的脑电特征分析方法, 提取特征, 并进行分类; 其次把连续的脑电信号离散化, 利用脑电信号组建脑电信号信息表, 最后利用约简理论, 对脑电信号进行特征约简, 并利用约简特征进行分类, 验证分类准确率。

结果与结论: 利用信息约简的方法及特征标示, 把连续的脑电信号特征离散化, 利用离散化的脑电信号特征构建信息表, 并进行特征筛选, 在保证分类准确率的情况下降低了特征的数量。不足之处在于, 该方法对分类特征进行标记只能在两类之间进行, 对多类型分类特征进行标记和分类特征进行约简将是下一步工作的重点。

关键词: 脑机接口; 特征提取; 信息约简; 脑电信号; 数字化影像技术

doi:10.3969/j.issn.1673-8225.2010.09.028

穆振东, 肖丹. 基于信息约简的脑电信号特征提取[J]. 中国组织工程研究与临床康复, 2010, 14(9):1642-1644. [http://www.crter.org http://cn.zgckf.com]

0 引言

脑机接口是建立在脑电信号分析基础上的一种生物技术和计算机技术相结合的应用型研究, 它提供了一种新型的人机交互方式, 通过制定的脑机接口系统, 利用相应的外部设备, 绕过人体外部肌肉骨骼组织的帮助, 直接产生人脑所想象的相应动作。脑机接口系统的核心是对脑电信号的分析, 特别是相应的想象所产生的脑电信号特征提取。当大脑产生一定的思维活动时, 会对产生相应的脑电信号; 同样, 对特定脑电信号进行分析, 能还原相应的大脑思维, 这就形成了脑电信号-思维活动的对应关系。

在脑机接口系统中, 对大脑想象的反应速度

是衡量脑机接口系统质量的一个重要标准, 因此对脑电信号的分析, 对脑电信号特征进行筛选就显得尤为重要, 尽可能少的特征来标示相应的动作是提高速度的一个重要途径。由于对冗余信息约简的重要性, 很多方法被引入用来进行信息约简^[1-10]。本文利用信息系统中对属性进行约简的方法来对脑电信号特征进行特征筛选, 用较少的特征来达到较高的分类准确率。

1 数据约简理论

为了从脑电信号-维活动所组建的规则集合中, 提取出最基本规则, 作者引用了信息系统的概念。一个信息系统是一个二元对 $S=(U,A)$, 其中 U 是一个非空、有穷、被称为全域的个体集, A 为非空、有穷的属性集, 即对

江西蓝天学院信息
技术研究所, 江
西省南昌市
330098

穆振东★, 男,
1975年生, 河南
省上蔡县人, 汉
族, 2004年南昌
大学毕业, 硕士
讲师, 主要从事生
物信号处理研究。
zd_mu@sohu.
com

中图分类号:R318
文献标识码:B
文章编号:1673-8225
(2010)09-01642-03

收稿日期:2009-10-15
修回日期:2009-12-18
(20091117014/GW-Z)

于属性 $a \in A$ 有: $U \rightarrow V_a$, 其中 V_a 被称为属性 a 的值集合, $V = Y_a \in AV_a$ 被说成属性集 A 的值域。如果 $S = (U, A)$, 则 $S' = (U', A')$ 使得 $U \in U'$, $A' = \{a: a \in A\}$, 对 $u \in U$, $a'(u) = a(u)$, 并且对 $a \in A$, 有 $V_a = V'_a$, 则称 S' 为 S 的一个 U' -扩充, S 被称作为 S' 的子系统。若 $S = (U, A)$, 则使得 A 的 $S' = (U, B)$ 被称做 S 的一个 B -扩充。

在脑电信号组成的信息系统中, 属性集合包含有很多冗余属性, 因此必须筛选出其中核心的属性来表示相应的思维活动; 在脑电信号组成的信息系统中, 由于脑电信号的特殊性, 其中包含很多相互矛盾的规则, 只有挑选出正确的规则, 才能提高脑机接口系统的准确率, 为此, 作者应用了信息系统的属性约简的方法, 本文用到了以下几个定义:

定义1: 如果任意一个 $r \in R$ 是 R 中不可约去的, 则等价关系族 R 是独立的, 否则 R 是相关的。

定义2: 设 R 是一个等价关系族, $r \in R$, 若 $IND(R) = IND(R - \{r\})$, 则称 r 在 R 中是可被约去的知识; 若 $P = R - \{r\}$ 是独立的, 则 P 是 R 的一个约简。

定义3: R 中所有的不可约去关系称做为核, 由它们构成的集合称为 R 的核集, 记为 $Core(R)$ 。

定义4: 函数 $dx: A \rightarrow V$, 使得 $dx(a) = a(x)$, 其中 $a \in A, X, U, x \in U$, 称 dx 是决策表 T 上的一条决策规则, 若 $a \in C, A$, 则记 $dx|C$ 是决策规则的条件部分, 若 $a \in D \in A$, 则记 $dx|D$ 是决策规则的结论部分。

定义5: 若任意一个 $y \neq x, dx|C = dy|C \rightarrow dx|D = dy|D$, 则称 dx 是一致的, 否则是不一致的。

定义6: 若对于所有的决策规则都是一致的, 则它们所处的决策表 $T = (U, A, C, D)$ 是一致的, 否则它们所在的是不一致的。

2 脑电特征约简

本文所用数据采用奥地利 Graz University of Technology 生物医学工程研究所公布的一组 Graz-BCI 数据, 本文采用了其中 Data IIIa 数据集中的 K3b, K6b, L1b, 用来验证算法可行性。

为了提取样本特征和对样本提取特征进行验证, 分别从两组数据中提取的想象左右运动脑电信号样本, 并把提取样本分成两类: 学习样本和测试样本。从学习样本中进行特征抽取, 用测试样本对所抽取出的脑电信号特征进行测试。

在不同类型的脑电信号之间, 被抽取出的特征必须是能标示类型的特征。在相同类型的各个样本特征, 到本类型的该特征的距离比到其他类型的距离要近, 因此, 在进行特征约简之间, 通过计算学习样本的类型特征之间的距离, 能够对脑电信号特征进行标记。

具体计算方法如下:

①计算学习样本特征间距离:

$$A_{kij} = \sum_{n=1}^{c \ln} (X_{ij} - Y_{nj})^2 \quad (1)$$

其中 A_{kij} 表示学习样本类型 k 中第 i 个样本的第 j 个特征的标记, $c \ln$ 表示类型 k 中学习样本总数。

②标记学习样本类型:

$$M_{ij} = \begin{cases} 1A_{1ij} < A_{2ij} \\ 2A_{1ij} > A_{2ij} \end{cases} \quad (2)$$

其中 M_{ij} 表示学习样本 i 的第 j 个特征的标记。

在学习样本中, 把样本的运动类型组成集合 D , 结合样本标记集合 M , 组成: $M_i \rightarrow D_i$ 的样本信息, 其中 i 为在学习集合中第 i 个样本。由样本标记集合 M 和运动类型集合 D 组成运动想象学习集脑电特征信息表, 计为 MD 表。在 MD 表中包含了大量的非类型标志的特征, 造成了特征抽取的信息冗余, 影响分类的速度, 为了提高分类速度, 本文针对 MD 表进行数据约简。

脑电信号采集, 受环境和被试者影响较大, 所以在学习样本中, 容易出现样本不一致(定义5), 在进行约简之前, 需要对脑电特征信息表进行一致性判断, 判断脑电信息表一致性算法 $cons$ 描述如下: ①输入: 信息表 $(M_i \rightarrow D_i)$, 不一致集合 CL 。②不一致集合 CL 初始化为空集。③扫描信息表, 若对于 i, j 两个样本, $(M_i = M_j)$ and $(D_i \neq D_j)$, $CL = CL \cup \{(i, j)\}$ 。④若 $CL = \emptyset$, 则输出0, 否则输出1。

为了更好地进行特征筛选, 对学习样本进行样本筛选, 剔除影响脑电特征信息表一致性样本; 具体样本挑选算法 $samsel$ 如下: ①输入: 信息表 $(M_i \rightarrow D_i)$, 不一致集合 CL 。②不一致集合 CL 初始化为空集。③扫描信息表, 若对于 i, j 两个样本, $(M_i = M_j)$ and $(D_i \neq D_j)$, $CL = CL \cup \{(i, j)\}$ 。④扫描不一致集合 CL , 从信息表 $(M_i \rightarrow D_i)$ 中删除不一致样本。⑤输出信息表 $(M_i' \rightarrow D_i')$ 。

信息约简算法 $charred$ 描述如下: ①输入: 信息表 $(M_i \rightarrow D_i)$ 。②核心特征矩阵 $core\{\}$ 初始化为空集。③调用样本挑选算法 $samsel$, 得到一致脑电信息表 $(M_i' \rightarrow D_i')$ 。④扫描 $(M_i' \rightarrow D_i')$ 信息表中特征集 M_i' , 从 M_i' 中依次删除其中特征 m_i 。⑤调用脑电信息表一致性算法 $cons$, 结果赋值 $cout$ 。⑥若 $cout = 1$; 则集合 $core\{\} = core\{\} + \{m_i\}$, 若 M_i' 中特征扫描完毕, 输出 $core$, 否则返回步骤3。

3 算法验证

通过放大器采集到的脑电信号是时域脑电信号, 由

于脑电信号较弱, 因此真正的脑电信号被噪声数据所掩盖, 为了更好的进行分类, 首先对脑电原始信号进行转换。对脑电信号的转换有很多成熟的方法, 例如对脑电信号的同步和去同步关系进行分析^[11-13], 对脑电信号进行小波变换进行分析等^[14-19]。

文献[20-22]分别利用短时傅里叶变换(STFT)、二阶盲辨识(SOBI)和能量熵对脑电信号进行转, 本文分别选取文献20和21中的k3b左右想象脑电信号, 文献22中k3b左右想象的28 33电极对脑电信号为对照。利用相同的分类器, 特征约简前后对照结果如下:

特征约简前后对照表:

项目	STFT	SOBI	能量熵
分类特征数	16	100	86
准确率	93.1%	91.9%	91.6%
约简后特征数	16	35	43
约简后准确率	93.1%	90.5%	92.8%

4 结论

显示利用本方法对特征进行约简, 成功地在不降低分类准确率的同时, 减少了特征数量。不足之处在于, 本方法对分类特征进行标记只能在两类之间进行, 对多类型分类特征进行标记和分类特征进行约简将是下一步工作的中心。

5 参考文献

- [1] Gao J,Wang ST, Dianzi Xuebao.2009; 37(5): 1051-1057. 皋军,王士同. 基于矩阵模式的最小类内散度支持向量机[J]. 电子学报, 2009,37(5): 1051-1057.
- [2] Wang Y.Liaoning Shiyu Huagong Daxue Xuebao.2008; 28(4): 73-77. 王杨. 基于自适应遗传算法的粗糙集属性约简方法[J]. 辽宁石油化工大学学报, 2008,28(4): 73-77.
- [3] Jia XY,Yu SY,Shang L,et al. Guangxi Shifan Daxue Xuebao.2006; 24(4): 83-86. 贾修一,于绍越,商琳,等. 基于Rough集和蚁群算法的属性约简方法[J]. 广西师范大学学报: 自然科学版, 2006,24(4): 83-86.
- [4] Sun T,Dong LY,Li J,et al. Jilin Daxue Xuebao.2006; 44(2): 213-216. 孙涛,董立岩,李军,等. 用于粗糙集约简的并行算法[J]. 吉林大学学报: 理学版, 2006,44(2): 213-216.
- [5] Zhou RQ,Chen WW, Jisuanj Fuzhu Gongcheng.2006; 15(3): 51-53. 周如旗,陈文伟. 基于对称交叉熵的属性约简算法[J]. 计算机辅助工程, 2006,15(3): 51-53.
- [6] Yan Y,Yang HZ,Jisuanji Gongcheng Yu Yingyong.2007; 43(31): 156. 颜艳,杨慧中. 基于遗传算法的粗糙集属性约简算法[J]. 计算机工程与应用, 2007,43(31):156.
- [7] Pan W,Wang YF.Jisuanji Gongcheng Yu Yingyong 2008,44(15): 1-3. 潘伟,王云峰. 基于自适应遗传算法的粗糙集知识约简算法[J]. 计算机工程与应用, 2008,44(15): 1-3.
- [8] Tao Z, Xu BL, Wang DW,et al. Xitong Gongcheng.2003; 21(4): 116-124. 陶志,许宝栋,汪定伟,等. 基于遗传算法的粗糙集知识约简方法[J]. 系统工程,2003, 21(4):116-124.
- [9] Hu QH,Yu DR,Xie ZX Ruanjian Xuebao.2008; 19(3): 640. 胡清华,于达仁,谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简[J]. 软件学报, 2008, 19(3): 640.
- [10] Xia Zhao,Minghu Ha,Aiquan Zhang. Variable Precision Rough Set Model in Incomplete Ordered Decision System[C]. Innovative Computing Information and Control,2008ICIC'08.3rd International Conference. 2008:498-498.

- [11] Zhuang P.Zhongguo Linchuang Kangfu.2004;8(1): 152-154. 庄平. 脑电事件相关去同步化和同步化活动与运动相关性作业[J]. 中国临床康复, 2004, 8(1):152-154.
- [12] Hao DM,Ruan XG,Shengwu Wuli Xuebao 2005;21(1): 39-45. 郝冬梅 阮晓钢. 脑电事件相关去同步化和同步化的神经元群模型[J]. 生物物理学报, 2005, 21(1): 39-45.
- [13] Zhang AH,Zhao YH.Zhongguo Linchuang Kangfu.2005; 9(48): 4-7. 张爱华 赵予晗. 脑电信号相同步分析在识别左右手想象运动中的作用[J]. 中国临床康复, 2005, 9(48):4-7.
- [14] Wan BK,Qi HZ,Zhao L,et al.Tianjin Daxue Xuebao.2006;39(8): 977-983. 万柏坤, 慕宏志, 赵丽, 等. 基于脑电Alpha波的脑-机接口控制实验[J]. 天津大学学报, 2006, 39(8):977-983.
- [15] Li K,Chu LL,Zhu SD,et al.Jisuanji Jishu Yu Fazhan.2006;16(8): 157-159. 李坤, 褚蕾蕾, 朱世东, 等. 基于mu节律能量的运动意识分类研究[J]. 计算机技术与发展, 2006, 16(8):157-159.
- [16] Liao X,Yin Y,Yao DZ.Zhongguo Yixue Wulixue Zazhi.2006; 23(2): 129-132. 廖祥, 尹愚, 尧德中. 基于连续小波变换和支持向量机的运动想象脑电分类[J]. 中国医学物理学杂志, 2006, 23(2):129-132.
- [17] Yang BH,Yan GZ,Yan RG.Shanghai Jiaotong Daxue Xuebao. 2005;39(11): 1879-1882. 杨帮华, 颜国正, 严荣国. 脑机接口中基于小波包最优基的特征抽取[J]. 上海交通大学学报, 2005, 39(11):1879-1882.
- [18] Zou JD,Chen XP,He WX,et al.Zhongguo Zuzhi Gongcheng Yanjiu Yu Linchuang Kangfu.2006; 10(13): 110-112. 邹继东, 陈小平, 和卫星, 等. 小波熵提取脑意识任务特征的脑机接口设计[J]. 中国组织工程研究与临床康复, 2006, 10(13):110-112.
- [19] Yang BH,Yan ZG,Yan B.Zhongguo Shengwu Yixue Gongcheng. 2006; 25(5): 518-522. 杨帮华, 颜国正, 鄢波. 基于离散小波变换提取脑机接口中脑电特征[J]. 中国生物医学工程, 2006, 25(5):518-522.
- [20] Mu ZD,Hu JF,Xiao D. Zhongguo Zuzhi Gongcheng Yanjiu Yu Linchuang Kangfu. 2009; 13(26): 5079-5082. 穆振东,胡剑锋,肖丹. 基于时频域分析的运动想象脑电信号分类[J]. 中国组织工程研究与临床康复,2009,13(26): 5079-5082.
- [21] Hu JF,Xiao D, Mu ZD. Zhongguo Zuzhi Gongcheng Yanjiu Yu Linchuang Kangfu. 2008;12(13): 2481-2484. 胡剑锋,肖丹,穆振东. 基于二阶盲辨识的运动想象脑电信号分类[J]. 中国组织工程研究与临床康复,2008,12(13): 2481-2484.
- [22] Hu JF, Mu ZD,Xiao D. Jisuanji Gongcheng Yu Yingyong. 2008;44(33): 235-238. 胡剑锋,穆振东,肖丹. 基于能量熵的运动想像脑电信号分类[J]. 计算机工程与应用,2008,44(33): 235-238.

来自本文课题的更多信息--

基金资助: 江西省教育厅 2008 年度科技计划项目“基于 Rough 集的脑电信号特征提取和分类方法研究”(GJJ08477) 研究阶段成果。

利益冲突: 无相关利益冲突。

课题的创新点: 主要创新点和特色有以下几点: 将 Rough 集理论应用于脑电信号特征提取和分类研究; 采用新的刺激模式, 使得脑电信号提取不再通过长时间的学习和训练, 提高效率; 降低环境因素对脑电信号分类的影响, 使脑电信号特征达到临床应用的要求, 推动脑电信号朝应用方面发展。

课题评估的“金标准”: 课题未采用“金标准”, 虽然课题是对脑电信号进行分析处理, 但课题申报领域为信号与信息处理。

设计或课题的偏倚与不足: 对连续的脑电信号进行约简方法, 仅适用于两种想象脑电信号; 不同特征提取方法上, 最终特征数目差距较大; 提取出的脑电信号特征没有表现其生理特征。

提供临床借鉴的价值: 课题针对运动想象脑电信号特征提取和分类研究, 利用信息约简能降低脑电信号特征数目, 以此达到提高脑电信号分类速度。