

# 基于支持向量机的酗酒脑电信号分类研究

丁尚文<sup>1</sup>, 王纯贤<sup>2</sup>

(1. 合肥工业大学基础部, 安徽 宣城 242000; 2. 合肥工业大学机械工程系, 安徽 宣城 242000)

**摘要:** 酗酒会对脑认知功能产生严重损伤。为了检测长期饮酒人员是否有酗酒倾向问题, 提出一种基于支持向量机的酗酒脑电信号特征分类识别方法。借助三类评估参数, 实现了对健康者和酗酒者脑电信号定量分析。研究结果发现, 能量评估参数是一种新的最佳评估分类因子, 采用 $\theta$ 能量评估参数利用支持向量机能够对两类不同特征的脑电信号进行分类识别, 识别精度最高达到90%。该方法能够为长期饮酒人员是否有酗酒倾向提供一种新的辅助检测方法, 有助于提醒长期饮酒群体对健康给予关注。

**关键词:** 酗酒; 脑电信号; 支持向量机; 功率谱; 近似熵; 径向基; 极值; 对偶

中图分类号: TH-78 文献标志码: A DOI: 10.16086/j.cnki.issn1000-0380.2019020018

## Research on Classification of Alcoholism EEG Signals Based on Support Vector Machine

Ding Shangwen<sup>1</sup>, Wang Chunxian<sup>2</sup>

(1. Department of Basic Course, HeFei University of Technology, Xuancheng 242000, China;  
2. School of Mechanical Engineering, HeFei University of Technology, Xuancheng 242000, China)

**Abstract:** Alcohol abuse can cause serious damage to brain cognitive function. In order to detect whether long term drinkers have a tendency to alcohol abuse, a feature classification and recognition method of alcoholic EEG signals based on support vector machine (SVM) is proposed. Quantitative analysis of EEG signals for healthy and alcoholics people is realized by using three kinds of evaluation parameters. The results show that  $\theta$ -energy evaluation parameter is a new optimal classification factor. With theta-energy evaluation parameter, by using support vector machine, two kinds of EEG signals with different characteristics can be classified and recognized, and the recognition accuracy can reach 90%. This method can provide a new auxiliary detection method for the tendency of alcoholism of long-term drinkers, and is helpful for reminding long-term drinkers pay attention to health.

**Keywords:** Alcoholism; EEG signals; Support vector machine (SVM); Power spectrum; Approximate entropy; Radial basis; Extremum; Duality

## 0 引言

饮酒作为一种社会习俗和文化已经流传了几千年。酒文化的巨大作用和影响推动了社会的发展, 同时也给社会带来了巨大损失。长期过量饮酒会产生高血压、糖尿病及心脑血管等疾病。高剂量摄入酒精会引发脑细胞毒性水肿, 并且长期酗酒会导致脑部形态在额叶、胼胝体等部位发生特异性损伤<sup>[1]</sup>。

对于过量饮酒危害是否对脑认知功能产生损伤, 国内外学者开展了相关研究工作。Michael 等采用威斯康星卡片分类对急性饮酒志愿者进行测验, 结果表

明急性酒精中毒可导致大脑执行控制能力下降, 其相应的持续语言功能由于受到酒精抑制而导致语言表达不连贯<sup>[2]</sup>。谢成娟等对酒精依赖患者进行爱荷华博弈测试(Iowa gambling test, IGT)。试验结果表明, 对照组的 IGT 净得分显著高于酒精依赖组, 并推测认为患者的 IGT 表现受损, 可能与眶额叶皮质、杏仁核受损等有关<sup>[3]</sup>。Ehler 等采用脑电(electroencephalogram, EEG)技术研究发现酒精依赖患者在静息态下额叶部位 $\beta$ 功率明显高于对照组<sup>[4]</sup>。刘桂青等<sup>[5]</sup>对酗酒者脑皮层 EEG 信号的同步性开展研究, 结果发现酗酒者大脑不同区域间的功能连接强度受到一定程度的损伤。

收稿日期: 2019-02-18

基金项目: 合肥工业大学学术新人提升计划 A 项目基金资助项目(JZ2016HGTA0685)

作者简介: 丁尚文(1981—), 男, 博士, 讲师, 主要从事仪器信号测量、机器视觉测量等方向的研究, E-mail: dshou1@126.com

Korucuoglu 等采用 EEG 方法研究急性饮酒对脑认知功能的影响,结果发现与安慰剂组相比急性饮酒志愿组的额叶处 EEG 信号  $\alpha$  功率有明显的增强<sup>[6]</sup>。文献[7]采用的 Flanker 范式考察被试者在清醒状态和饮酒状态下的事件相关电位(event-related potentials, ERP)的差异性,研究结果发现饮酒会导致前扣带回处的神经活动兴奋,进而使得觉察、认知功能和控制能力下降。

综上所述可以发现,国内外研究人员采用认知量表、EEG 技术等研究了酗酒成瘾者脑认知功能损伤问题,研究结果给出了饮酒对脑认知功能产生影响的定性结论。进一步研究期望获得研究个体 EEG 信号评估特征参数;利用已经获得的评估参数对酗酒者 EEG 信号识别与分类。

## 1 支持向量机原理

支持向量机(support vector machine, SVM)是以统计学理论为基础的一种模式分类识别技术<sup>[8]</sup>。为了对训练数据进行分类, SVM 的基本原理就是要寻找一个最优分类面,尽可能使得训练数据中样本间距最大,从而将训练数据分类。最优分类面如图1所示。图1中的虚线表示分类的边界线,在边界线上的向量称之为支持向量。边界线之间的距离称之为分类间隔。

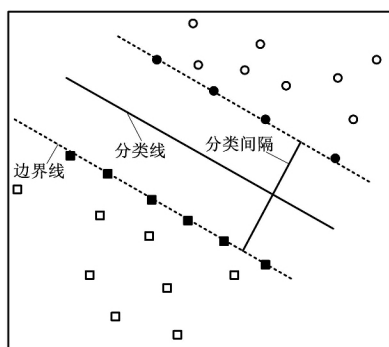


图1 最优分类面示意图

Fig.1 Schematic diagram of optimal classification surface

$w \cdot x + b = 0$  为分类线方程。该方程满足条件:

$$y_i [(w \cdot x_i) + b] - 1 \geq 0$$

若要使得训练数据被分成两类,必须满足分类间隔最大,即满足边长  $\|w\|$  最小条件才能使分类间隔最大。该问题等价于求解下面的优化问题:

$$\min \phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w \cdot w) \quad (1)$$

采用 Lagrange 求极值方法,利用对偶理论将式(1)中目标函数最值求解问题转化为求下列目标函

数最值问题<sup>[8]</sup>:

$$\begin{cases} \max Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ \text{s. t. } \alpha_i \geq 0, i = 1, 2, \dots, n, \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (2)$$

求解式(2)中对偶问题,得到最优分类面的权系数向量以及分类域值为:

$$\begin{cases} w^* = \sum_{i=1}^n \alpha_i^* y_i x_i, \\ b^* = \frac{1}{2} [w^* \cdot x(1) + w^* \cdot x(-1)] \end{cases} \quad (3)$$

$x(1)$  为式(3)中的训练数据中第一类样本点,  $x(-1)$  为第二类样本点。式(4)中的最优分类函数称之为支持向量机,又称之为最优分类函数。

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i^* y_i (x_i \cdot x) + b^* \right\} \quad (4)$$

$(x_i \cdot x)$  在式(4)中表示为内积,且  $(x_i \cdot x)$  可选用核函数  $K(x_i, x)$  替换,核函数的选取满足 Mercer 条件即可,但核函数表达式的不同对 SVM 分类结果有不同影响。另外,对于训练样本不可分情况, SVM 借助松弛变量  $\xi$  以判断对训练数据  $x_i$  的分类程度,惩罚因子  $C$  用来判断最小错分样本点个数和分类间隔,最终获得最优分类面。式(4)中的  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  为已知训练样本。其中,  $x_i \in R^l, y_i \in \{+1, -1\}, i = 1, 2, \dots, n$ 。支持向量算法优化的目标函数为:

$$R(w, \xi) = \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^n \xi_i \right) \quad (5)$$

$$\text{s. t. } y_i (w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

常用的核函数形式主要有线性核函数、内积核函数、径向基核函数等<sup>[8]</sup>。

## 2 数据获取来源

本文分析采用的数据均来自于纽约大学 HenriB 教授在互联网上公开的 EEG 数据库。该数据主要涉及的是酒精中毒病人的相关脑电信号数据。试验记录了两组试验对象,即酗酒者和对照组在3种视觉刺激条件下的 EEG 信号。试验时,按照国际标准在受试者头部放置 64 导电极,设备采样频率为 256 Hz,每次试验记录 1 s 的数据。试验过程施加单一刺激或者复合刺激。数据采集存放在两个数据集 SMNI\_CMI\_TRAIN 和 SMNI\_CMI\_TEST。数据选择:本文数据选择来源于该试验的大数据集(The Large Data Set),受试者包括酗酒者和正常人。各选择 10 例受试者数据用来测试。从酗酒者和对照组的 EEG 数据中分别随机选择 6 段数据作为分析样本,每组试验数据作为训练样本或者

测试样本, 共计三组数据, 分别简称数据 1、数据 2 和数据 3。

### 3 面向酗酒脑电信号的分类技术研究

本文将酗酒者和健康者 EEG 信号相关参数作为标准, 采用 SVM 方法期望借助评估参数对酗酒者和健康者 EEG 信号进行分类识别, 同时优选出用于识别两类不同 EEG 信号的最佳评估因子。

#### 3.1 酗酒脑电信号参数选择

能量参数: 在 EEG 节律特征研究中, 大多选择特征波的能量这一特征值对信号进行特征识别。EEG 信号的低频率节律相关能量  $E$  可用式 (6) 计算得到:

$$E = \sum_{i=1}^N |x(n)|^2 \quad (6)$$

式中:  $x(n)$  为 EEG 信号的幅值;  $N$  为采集到数据点个数。

通过式 (6), 能够计算出低频率节律波 EEG 中相关的  $\alpha$ 、 $\beta$  和  $\theta$  节律波段的能量<sup>[9]</sup>。

EEG 信号功率谱 AR 参数: 在信号频谱分析方法中, AR 谱估计 (简称 AR 模型) 由于可用较短时长数据获得较高频率分辨的优点而被研究者广泛使用。AR 算法详见文献 [10]。该算法是建立在数据基础上的基于自回归系数求解的有效算法。

EEG 信号近似熵: Pincus 等在研究混沌现象课题时发现熵可以用来描述混沌现象<sup>[11]</sup>, 他们利用信号的时间序列复杂度提出了近似熵模型。通过近似熵判断时间序列中新信息发生的可能性, 以此判断混沌现象中产生新模式的可能性。近似熵具体算法详见文献 [11-12]。

#### 3.2 酗酒脑电信号分类识别结果

将酗酒者和健康者的 EEG 信号的特征参数集合 {‘EEG 信号能量参数’, ‘EEG 信号功率谱 AR 参数’, ‘EEG 信号近似熵’} 中任意元素作为 EEG 信号评估参数, 将该参数对应的信号数据作为支持向量的训练样本。例如将 EEG 信号能量值作为训练样本 ( $x_1, y_1$ ) ( $x_2, y_2$ )  $\dots$  ( $x_n, y_n$ )  $x_i \in R^2$   $y_i \in \{+1, -1\}$   $i = 1, 2, \dots, n$ 。酗酒者的 EEG 信号能量值标签对应设定为 1, 健康者的 EEG 信号能量值标签对应设定为 -1。将 10 例酗酒者和 10 例健康者的 EEG 信号能量值数据作为训练样本, 对 SVM 模型参数进行训练和估计。将对应的两类 EEG 信号能量值各 10 例为测试样本, 分别对应的标签设置为 1 和 -1, 具体流程见图 2。根据图 2 算法流程, 对 EEG 数据采用 SVM 方法 (使用 libsvm 软件包) 进行训练和测试。本文选

择线性核函数作为 SVM 的核函数, 借助 EEG 信号不同频率段对应的能量值, 将酗酒者和健康者 EEG 信号分类。

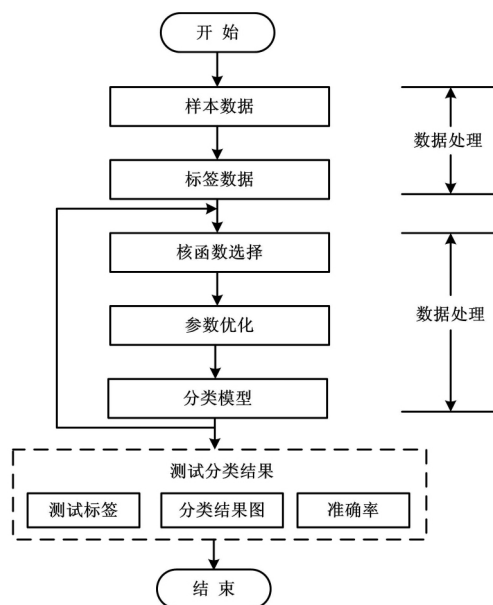


图 2 基于 SVM 的酗酒者和健康者的 EEG 信号分类流程图

Fig. 2 EEG signal classification process for alcoholics and healthy people based on SVM

基于不同数据组的测试效果对比如表 1 所示。

表 1 基于不同数据组的测试效果对比  
Tab. 1 Comparison of test results based on different data groups

特征参数	测试误差率 / %		
	数据 1	数据 2	数据 3
$\alpha$ 能量	41.67	40	25
$\beta$ 能量	41.67	42.5	35
$\theta$ 能量	20	17.5	10
近似熵	23.33	22.5	30
AR 参数	30	30	25

从表 1 可以发现, 采用近似熵、AR 参数作为训练样本, 能够对酗酒者 EEG 信号和健康者 EEG 信号分类识别, 测试误差率最小值分别为 22.5% 和 25%。采用  $\alpha$ 、 $\beta$  和  $\theta$  能量值作为训练样本, 能够对酗酒者 EEG 信号和健康者 EEG 信号分类识别, 测试误差率最小值分别为 25%、35% 和 10%。对应的基于  $\theta$  能量评估参数的酗酒者和健康者 EEG 信号分类结果见图 3。通过以上的分类结果比较, 可以发现最佳的评估因子为  $\theta$  能量评估参数, 且使用  $\theta$  能量评估参数采用 SVM 方法能够对两类不同 EEG 信号分类识别, 识别精度最高达到 90%。

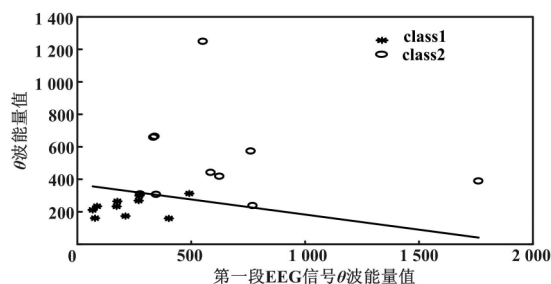


图3 第十通道 EEG 信号  $\theta$  能量评估参数分类结果图  
(分类误差 10%)

Fig.3 The classification result chart of theta energy evaluation parameter of the tenth channel EEG signal (classification error 10%)

#### 4 结束语

酗酒会对脑认知功能产生严重损伤,国内外研究者采用认知量表、EEG 等技术已进行大量的研究,并给出较为可靠的脑认知损伤报告。本文提出基于支持 SVM 方法的 EEG 信号自动分类检测技术。本文的工作能够为当前国内外研究机构对酗酒成瘾者评估检测提供技术参考,能够对健康饮酒者是否有酗酒倾向提供辅助检测。

参考文献:

[1] UHLMANN A, BANDELOW B, STEIN D J, et al. Grey matter structural differences in alcohol-dependent individuals with and without comorbid depression/anxiety—an MRI study. [J]. European Archives of Psychiatry & Clinical Neuroscience 2018(3): 1–10.

[2] Michael Lyvers, Juliette Tobias-Webb. Effects of acute alcohol consumption on executive cognitive functioning in naturalistic settings [J]. Addictive Behaviors 2010 35(2): 1021–1028.

[3] 谢成娟, 袁莉莉, 汪凯. 酒精依赖患者的决策功能障碍 [J]. 安徽医科大学学报 2012 47(12): 1460–1462.

[4] EHLERS C L, PHILLIPS E, GIZER I R, et al. EEG spectral phenotypes: heritability and association with marijuana and alcohol dependence in an American Indian community study [J]. Drug and Alcohol dependence 2010 106(2): 101–110.

[5] 刘桂青, 曹锐, 相洁. 酗酒者皮层脑电同步性分析 [J]. 中国生物医学工程学报 2016 35(6): 671–676.

[6] KORUCUOGLU O, GLADWIN T E, WIERS R W. Preparing to approach or avoid alcohol: EEG correlates, and acute alcohol effects. [J]. Neuroscience Letters 2014 559(5): 199–204.

[7] RIDDERINKHOF K R. Alcohol consumption impairs detection of performance errors in mediofrontal cortex [J]. Science, 2002, 28(6): 209–211.

[8] 高世伟, 赵力. 一种基于支持向量机的软测量建模方法 [J]. 自动化仪表 2017 38(7): 42–45.

[9] 丁尚文, 钱志余, 李慧韬, 等. 基于低频段光诱发的 EEG 和 fMRI 同步化特性分析 [J]. 天津大学学报: 自然科学与工程技术版, 2012 45(9): 803–807.

[10] 丁玉美, 阔永红, 高新波. 数字信号处理—时域离散随机信号处理 [M]. 西安: 西安电子科技大学出版社, 2002.

[11] PINCUS S M. Approximate entropy as a measure of system complexity. [J]. Proceedings of the National Academy of Sciences of the United States of America, 1991 88(6): 2297–2301.

[12] 杨孝敬, 杨阳, 李淮周, 等. 基于模糊近似熵的抑郁症患者静息态功能磁共振成像信号复杂度分析 [J]. 物理学报, 2016, 65(21): 320–331.

(上接第 94 页)

设定初始 SOC 误差为 25%。仿真结果表明, 锂电池 SOC 估算误差在 10 s 后低于 10%, 在 80 s 后最大误差低于 2%, 全程平均误差为 0.45%。改进 EKF 算法能有效修正 SOC 初值引起的误差, 并防止因为计算机字长有限而引起的滤波发散, 且估算效果良好。

参考文献:

[1] ZHENG Y. J. Investigating the error sources of the online state of charge estimation methods for lithium-ion batteries in electric vehicles [J]. Journal of Power Sources 2018, 37(7): 161–188.

[2] 凡旭国, 周金治. 基于 PNGV 模型和高斯-厄米特滤波的 SOC 估算研究 [J]. 自动化仪表 2017 38(12): 21–26.

[3] GRADITI G. Technical and economical assessment of distributed electrochemical storages for load shifting applications: an Italian case study [J]. Renewable & Sustainable Energy Reviews, 2016(57): 515–523.

[4] 梁奇, 于春梅, 王顺利, 等. 基于 PNGV 电路模型的航空钴酸锂

电池内阻研究 [J]. 电源学报 2017 15(2): 153–158.

[5] LAI X, ZHENG, SUN T. A comparative study of different equivalent circuit models for estimating state-of-charge of lithium-ion batteries [J]. Electrochimica Acta 2018 2(59): 566–577.

[6] 邓磊, 李小谦, 吴浩伟, 等. 基于改进 PNGV 模型的动力锂电池 SOC 精确估计 [J]. 电源技术 2017 41(10): 1461–1463.

[7] 张传伟, 李林阳, 赵东刚. 基于 BP 神经网络法估算动力电池 SOC [J]. 电源技术 2017 41(9): 1356–1357.

[8] CECILIO B, LUCIANO S, MANUELA G. An equivalent circuit model with variable effective capacity for LiFePO4 batteries [J]. IEEE Transactions on Vehicular Technology, 2014, 63(8): 3592–3599.

[9] 杜森, 谢立洁, 徐梓荐, 等. 三阶扩展 GNL 电池模型的研究 [J]. 电子设计工程 2018(20): 110–113.

[10] WANG S L, FERNANDEZ C, ZOU C Y, et al. Open circuit voltage and state of charge relationship functional optimization for the working state monitoring of the aerial lithium-ion battery pack [J]. Journal of Cleaner Production 2018(10): 1090–1104.