

# 融合互信息和支持向量机的癫痫自动检测算法

沈洋洋<sup>1</sup>, 黄丽亚<sup>1</sup>, 郭迪<sup>1</sup>, 笪钺璐<sup>1</sup>, 陈志阳<sup>1</sup>, 戴加飞<sup>2</sup>

(1. 南京邮电大学 电子科学与工程学院, 江苏 南京 210003;

2. 南京军区南京总医院神经内科, 江苏 南京 210002)

**摘要:**脑电图(Electroencephalogram, EEG)是通过电极记录下来的脑神经细胞群的自发性、节律性电活动,是癫痫诊断中重要的一项检查工具。文中提出了一种新的基于互信息(Mutual Information, MI)和支持向量机(Support Vector Machine, SVM)的特征提取和分类的方法,可以高效地区分正常脑电信号和癫痫脑电信号,并分别对比了相同维度下不同特征向量组合以及不同维度的特征向量组合的分类效果。除此之外,还对比了文中算法与其他常用算法的分类效果和算法效率。实验结果表明,由两类脑电信号的互信息序列提取的以均值、方差组成的二维特征向量,具有运算简单、分类准确率高的优点,同时文中算法比其他常用算法具有更快的运算速度,这对于临床实时监控癫痫是否发作具有积极的指导意义。

**关键词:**互信息;支持向量机;脑电信号;特征提取;癫痫自动检测

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2016)06-0133-05

doi:10.3969/j.issn.1673-629X.2016.06.029

## An Automatic Detection Algorithm for Epilepsy EEG Based on MI and SVM

SHEN Yang-yang<sup>1</sup>, HUANG Li-ya<sup>1</sup>, GUO Di<sup>1</sup>, DA Cheng-lu<sup>1</sup>, CHEN Zhi-yang<sup>1</sup>, DAI Jia-fei<sup>2</sup>

(1. School of Electronics Science and Engineering, Nanjing University of Posts and Telecommunications,  
Nanjing 210003, China;

2. Department of Neurology, Total Hospital of Nanjing Military Region, Nanjing 210002, China)

**Abstract:** Electroencephalogram (EEG) is the most important tools for seizure detection by recording spontaneous and rhythmic electrical activity of brain cells through electrodes. A new method for feature extraction and classification is proposed based upon Mutual Information (MI) and Support Vector Machine (SVM), which can distinguish epilepsy EEG from normal EEG quickly and efficiently. Then the comparison on the classification results is made using various combinations of feature vector in the same dimension and in the different dimension. In addition, the classification results and efficiency are compared between proposed algorithm and other common algorithm. The experiment shows that the two-dimensional feature vectors combining mean and variance extracted from MI sequence of epilepsy EEG and normal EEG, has advantages of simple operation and high classification result, and this algorithm is also faster than others, which is useful for clinical seizure detection in real time.

**Key words:** MI; SVM; EEG; feature extraction; automatic epilepsy detection

## 1 概述

癫痫是困扰人类健康的常见疾病之一,是一种严重的慢性疑难性脑部疾病。癫痫发作时常表现为突然、短暂的运动、感觉、意识、行为、自主神经或精神症状等异常,影响患者的身体和智力发育,严重时甚至危及生命。

目前诊断癫痫最有效的方法就是脑电图检查。众所周知,人类的大脑是目前已知的一个最复杂的系统,

数十亿神经元连接构成了一个非常复杂的大脑网络。脑电图作为一个临床诊断的工具,可以记录丰富的大脑神经元活动信息。医疗工作者根据经验对患者的脑电图进行视觉检测,查找神经元的异常活动信息,从而对癫痫发作进行诊断<sup>[1]</sup>。然而,纯人力的检测不仅耗时,并且,由于分析的主观性和非实时性,极易造成误诊。因此,寻找一个实时、准确、快速的自动检测识别技术,就显得十分必要。

收稿日期:2015-08-16

修回日期:2015-11-26

网络出版时间:2016-05-05

基金项目:国家自然科学基金资助项目(61003237);江苏省高校自然科学基金项目(10KJB510018)

作者简介:沈洋洋(1991-),女,硕士研究生,研究方向为智能信息处理;黄丽亚,教授,硕士研究生导师,研究方向为智能信息处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160505.0829.086.html>

目前对脑电信号的时域、频域以及时频域的非线性分析方法主要集中在分析信号复杂度和相关性上。其中,针对癫痫诊断的自动检测方法大多采用小波变换<sup>[2]</sup>、熵<sup>[3-5]</sup>以及一些同步性参数来进行分析。例如,徐永红等于 2015 年提出基于改进多元多尺度熵的癫痫脑电信号自动分类方法,将传统的多元多尺度熵针对单个变量的嵌入模式改为对所有变量的同时嵌入。仿真实验表明,其对波恩癫痫数据的 5 个数据集均具有较好的分类精度<sup>[5]</sup>。Li Shufang 等于 2013 年利用经验模态分解(Empirical Mode Decomposition, EMD)和 SVM 对癫痫患者发作期和发作间期的脑电信号进行特征提取和模式识别。实验结果表明,该算法的敏感性达到 97%、特异性达到 96.25%<sup>[6]</sup>。韩凌等于 2014 年提出空频域特征分析方法预测癫痫,将多变量相位同步参数、希尔伯特边际谱和希尔伯特加权频率组成一个三维特征向量作为特征值输入到支持向量机中进行预测。实验显示,该方法具有较低的错误预报率以及较高的预测敏感度<sup>[7]</sup>。张涛等于 2015 年提出基于 AdaBoost 算法的癫痫脑电信号识别,通过对弱分类器进行筛选、引入平滑因子和权值修正函数三个措施对算法进行优化,并且将优化后的算法与小波包分解结合。仿真实验结果表明,该算法对癫痫信号的识别率为 96.11%,对正常脑电信号的识别率为 99.51%<sup>[8]</sup>,等等。

而近年来渐渐兴起的互信息分析方法在癫痫信号领域的应用还有待探索。互信息可以检测两个时间序列间线性和非线性的统计依赖关系,并可以衡量时间序列之间的信息传输。它作为一种非线性分析方法,可以度量信号之间的相关性,被广泛应用于癫痫<sup>[9]</sup>、精神分裂症<sup>[10]</sup>、静息状态<sup>[11]</sup>等脑电信号的分析与检测中。其中不乏将互信息与 EMD 结合、与小波变换结合,或者对时间序列进行一些时域处理(比如信号重构、信号时延等)提取交叉互信息、自动互信息等等。

文中直接将未经任何时域处理的时间序列之间的互信息应用于癫痫信号,提出一种基于互信息和 SVM 的快速识别算法,利用互信息分析多导联的脑电图,提取互信息的特定统计量做特征,输入到支持向量机进行分类,在对脑电信号识别提速的同时,又保证了识别准确率。

## 2 基本原理

### 2.1 互信息

如果一个离散随机变量  $X = \{x_1, x_2, \dots, x_N\}$ , 其概率分别是  $\{p_{1s}, p_{2s}, \dots, p_{Ns}\}$ , 根据 Shannon 信息熵的概念,可以定义  $X$  的 Shannon 熵为:

$$H(X) = - \sum_{i=1}^N p_{ix} \log_2 p_{ix} \quad (1)$$

随机变量  $X$  和  $Y$  的联合 Shannon 熵可以定义为:

$$H(X, Y) = - \sum_{i,j} p_{ij} \log_2 p_{ij} \quad (2)$$

则随机变量  $X$  和  $Y$  的互信息可以定义为:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) \quad (3)$$

根据互信息的概念,当两个时间序列完全相同时,它们的互信息最大;当两个时间序列相互独立时,它们的互信息值为 0;其他情况下的值介于 0 和最大值之间。

文中就是利用不同导联之间的互信息(无时间延迟),衡量癫痫脑电与正常脑电的不同导联之间信息传输的差异,并使用相应的统计量对其进行量化。

### 2.2 特征提取

文中计算出某一导联的时间序列与所有导联的时间序列之间的互信息,构成该导联的互信息序列。由于有 19 个导联,单纯用互信息序列组成特征向量太过庞大,而单独挑选出互信息序列中的几个互信息值组成特征向量又太过繁琐、片面,故文中选择几个常用的、可以分析互信息序列幅值和波动特性等统计特性的统计量来构成需要的特征向量,即均值、方差、偏差系数和波动系数。

均值可以描述互信息序列幅值的平均水平,第  $i$  个导联互信息序列的均值  $\mu_i$  定义如下:

$$\mu_i = (1/l) \sum_{j=1}^l MI(i, j) \quad (4)$$

其中:  $MI(i, j)$  表示  $i$  导联的互信息序列的第  $j$  个互信息值;  $l$  表示互信息序列的长度。

方差可以量化互信息序列的波动程度,第  $i$  个导联互信息序列的方差  $\sigma_i$  定义如下:

$$\sigma_i = \sqrt{(1/l) \sum_{j=1}^l (MI(i, j) - (1/l) \sum_{j=1}^l MI(i, j))^2} \quad (5)$$

偏差系数可以度量互信息序列幅度的变化,第  $i$  个导联互信息序列的偏差系数  $V_{ci}$  定义如下:

$$V_{ci} = \frac{\sigma_i^2}{\mu_{0i}^2} \quad (6)$$

其中,  $\mu_{0i} = (\frac{1}{l}) \sum_{j=1}^l |MI(i, j)|$ 。

波动系数用于量化互信息序列变化的强度,第  $i$  个导联互信息序列的均值  $F_i$  定义如下:

$$F_i = \frac{1}{l} \sum_{j=1}^l |MI(i, j+1) - MI(i, j)| \quad (7)$$

这四个统计量的分类效果会在后续实验中详论。

### 2.3 支持向量机(SVM)

分类是数据挖掘领域中一项非常重要的任务,它

的目的是学会一个分类函数或分类模型(或者叫做分类器),而支持向量机就是90年代中期发展起来的基于统计学习理论的一种监督式学习方法。

支持向量机<sup>[12-14]</sup>通过寻求最小结构化风险来提高学习机泛化能力,实现经验风险和置信范围的最小化,从而达到在统计样本量较少的情况下,亦能获得良好统计规律的目的。

假设线性可分样本集表示为 $(x_i, y_i), i = 1, 2, \dots, n, x \in R^d$ ,用 $y \in +1, -1$ 来标记分类结果。输入用作测试的特征向量,通过支持向量机算法以距离最大原则从两种类别中得到最优分类超平面(Optimal Separating Hyperplane, OSH)。

分类超平面必须满足如下约束条件:  
$$y_i[(\omega \bullet x_i) + b] - 1 \geq 0, i = 1, 2, \dots, n \quad (8)$$
在非线性可分的情况下,需引入松弛变量 $\xi_i$ 来修正方程(8):  
$$y_i[(\omega \bullet x_i) + b] - 1 + \xi_i \leq 0, i = 1, 2, \dots, n \quad (9)$$
在方程(9)的约束下,最小化方程(10),得到广义分类超平面:

$$\varphi(\omega) = \frac{1}{2} \|\omega\|^2 + C(\sum_{i=1}^n \xi_i) \quad (10)$$

其中, $C$ 是惩罚因子,对分类错误情况施加一定惩罚。

最后,得到广义的分类超平面,如方程(11),该式对线性可分和非线性可分情况均适用。

$$f(x) = \text{sgn}\left\{\sum_{i=1}^n a_i^* y_i K(x_i, x) + b^*\right\} \quad (11)$$

其中, $K(x_i, x)$ 表示核函数,它可以将原始的特征空间映射到一个新的高维特征空间,从而使得在原始空间线性不可分的情况,转变成在高维空间线性可分的情况。

常用的核函数有很多,文中主要使用线性核函数。线性核实际就是原始空间的内积,它将“映射后空间中的问题”和“映射前空间中的问题”两者在形式上统一起来,虽然使用了核函数,但实际还是求解原始空间的问题。

### 3 基于互信息与 SVM 的癫痫脑电信号的自动检测算法

#### 3.1 实验数据

文中使用了6个正常脑电信号和6个癫痫脑电信号,由南京军区总医院从临床诊断中采集,受试者年龄段为19~60岁,均为右利手,由专家对受试者们的脑电信号进行了鉴定。信号记录采用标准10-20系统,包括19个导联信号,采样频率为512 Hz,记录时间均超过1 min。

#### 3.2 实验方法

文中采用七重交叉验证,来获取更可靠、更稳定的算法性能结果。所谓七重交叉验证,即将原始 EEG 数据随机分为7个样本集,将其中的一个样本集作为训练数据训练模型,剩下的六个样本集作为验证数据测试模型。重复上述过程,使得每一个样本集都做过一次训练,这样就可以得到7个分类结果,对这7个结果进行平均,从而得到最终的分类结果。

- 具体算法如下:
- (1)将数据切分,以1 000个采样点为间隔,切分原始 EEG 数据,切分后的数据长度与原始数据的数据长度相关。
  - (2)求每个样本集的互信息序列,分别求取每个导联的时间序列与其他所有导联的时间序列之间的互信息,将其构成一个19×19的互信息矩阵。该矩阵的每一行都表示当前导联对应的互信息序列。
  - (3)选出最佳分类导联,并用该导联的均值、方差、偏差系数以及波动系数等统计量构成特征向量,输入到支持向量机中进行分类。

#### 3.3 实验结果及分析

单靠观察正常脑电和癫痫脑电波形图来诊断癫痫发作是比较困难的,而且诊断是在获取到脑电信号之后,并不能做到对癫痫发作的一个实时诊断。文中提出的算法就是为了实现对癫痫是否发作进行实时快速地诊断。下文将主要围绕以均值和方差构造特征向量进行分类,来论述该算法的具体操作步骤及其性能的优越性。

图1分别为癫痫脑电信号和正常脑电信号的其中一个样本经归一化处理后的互信息矩阵。矩阵中的每一个点,对应的都是两个时间序列之间的互信息值。

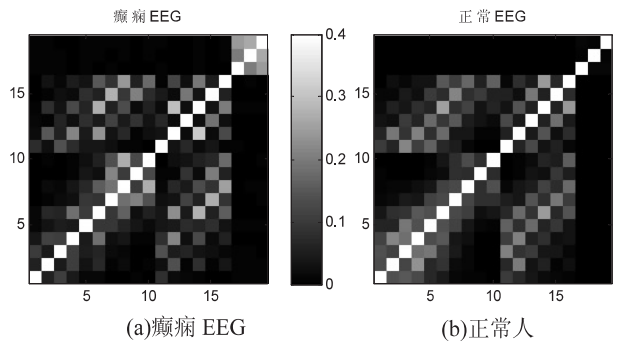


图1 互信息矩阵

从图中明显可以看到,矩阵的左上角,也就是17~18导联之间,癫痫脑电信号的互信息值明显高于正常脑电信号,即癫痫脑电信号在这一区间信息耦合程度更高。

为了更直观地分析这一区域互信息序列的幅度以及波动情况,用均值和方差对其做了一定量化。量化



的结果显示,癫痫脑电信号 17、18 导联的互信息序列的均值高于正常脑电信号,而方差却低于正常脑电信号,即表示癫痫脑电信号 17、18 导联的时间序列与其他导联的平均耦合程度高于正常脑电,且比正常脑电信号的信息传输更集中。通俗的说,癫痫脑电 17、18 导联与其他导联的信息传输比正常脑电更剧烈、高效,以这两个导联的均值和方差构成的特征向量可以使癫痫脑电和正常脑电线性可分。

但并不是所有导联的互信息序列提取的特征向量都与 17、18 导联一样线性可分,因此需要在 SVM 分类前自动选择最佳分类导联。文中最佳导联的选择方法是,将两类脑电信号的每个导联对应的统计量再进行一次平均,并求其平均值差的绝对值,分别选取两组绝对值排在前 1 位的导联,则两组中重合的导联即为最佳分类导联,若没有重合的,则继续选取绝对值排在前 2 位的导联,以此类推,直到选出重合导联。

为了更直观地展示这种自动选择最佳导联的方法及其效果,图 2 不仅画出了两类脑电信号每个导联统计量的平均值曲线,还画出了其统计量的箱须图。该方法最后选出的最佳导联为 18 导联,而由箱须图也可以看出,分类效果比较好的是 17、18、19 导联。

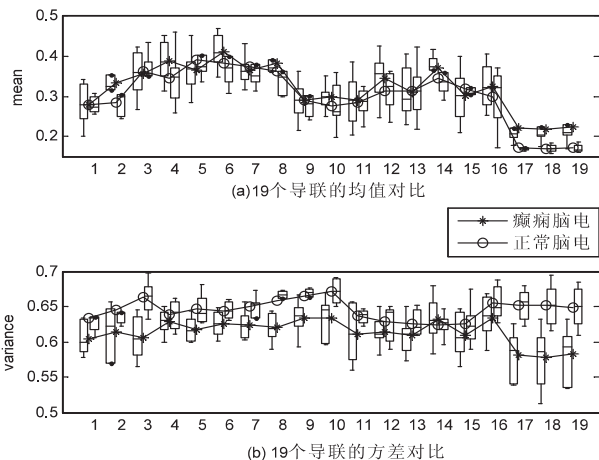


图2 均值、方差的箱须图以及平均值曲线

图 3(a) 给出了 19 个导联以均值、方差作为特征向量的分类准确率。

由图可见,17、18、19 导联的分类准确率最高,均在 99% 以上,因此最佳分类导联的选取方法尽管很简单,但定位却高效准确。

对均值和方差进行一些数学上的变化,便可得到另外两个统计量——波动指数和偏差系数,这两个统计量同样可以量化信号的幅度和波动情况。文献[2]中使用这两个统计量对癫痫信号进行分类获得了较好的结果,故文中在此进一步探讨下。对于文中所提出的算法,是否有必要增加数学运算量,选取波动指数和偏差系数来构成特征向量。图 3(a) 中两条曲线分别

表示选取均值、方差作特征向量的所有导联的分类结果,以及选取偏差系数、波动指数作特征向量的所有导联的分类结果,结果表明,后者的结果与前者区别不大,因此没必要增加不必要的运算量。

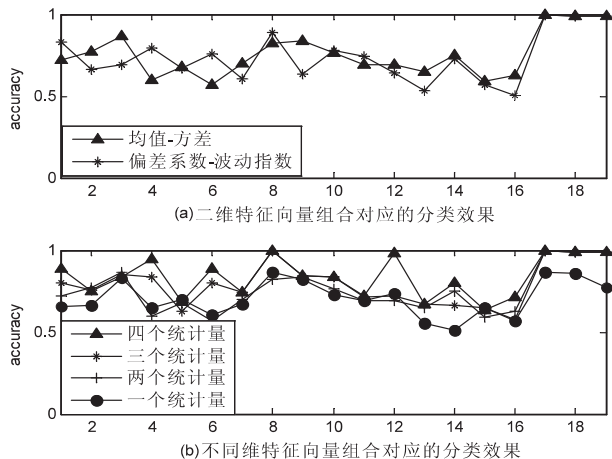


图3 分类效果

其次,SVM 是支持多维分类的,维度的增加通常可以提高分类准确率。图 3(b) 展示了一维、二维、三维和四维特征向量的分类效果。很明显,一维的分类准确率最高的都不到 90%,可见一维特征向量确实不适用于复杂情况。三维和四维特征向量的整体分类效果确实优于二维特征向量,但是三者的最佳分类准确率都达到了 100%,如果是应用于临床实时监测,二维特征向量既可保证算法速度,又可保证算法的准确率。同样,由图 3 可以看出,17、18 导联在所有的特征向量组合下均有最佳分类准确率,与图 1 的互信息矩阵得出的结论一致。

接下来,对文中算法与其他常用算法进行对比。分别为:

(1) 对时域信号时延 0.2 s,求其与未存在时延的时域信号的交叉互信息 (Cross Mutual Information, CMI),然后提取特征向量进行分类。

(2) 采用文献[2]的 EMD 与 SVM 结合的算法,对时域信号进行 EMD 分解,提取特征向量进行分类。

(3) 采用小波变换与近似熵结合的算法,对时域信号进行小波包变换并提取  $\alpha$  波,求其近似熵特征进行分类。

图 4(a) 展示了四种算法的分类结果,三角形表示文中算法的分类结果,星号表示对原始信号进行时延之后使用文中算法的分类结果,圆形表示使用 EMD 和 SVM 结合的算法的分类结果,加号表示小波变换和近似熵结合的算法的分类结果。另外,每种算法处理相同数据量的运行时间也标注在图 4(a) 中。

由图可见,经过时延后,文中算法分类效果大幅降低,这可能是因为时延后信号的相关性变弱;另外两种

常用算法的分类效果很好,但算法耗时却远超文中算法。因此,文中提出的算法兼具速度快和准确率高的优点。

最后,验证算法的健壮性。根据前期实验基础,对两类脑电信号的原始序列添加高斯白噪声进行计算,结果如图4(b)所示。加入高斯白噪声之后,各个导联的整体分类趋势与没有噪声的时候类似,且最佳分类效果依旧高达100%。由此可见,文中提出的算法是可靠有效的。

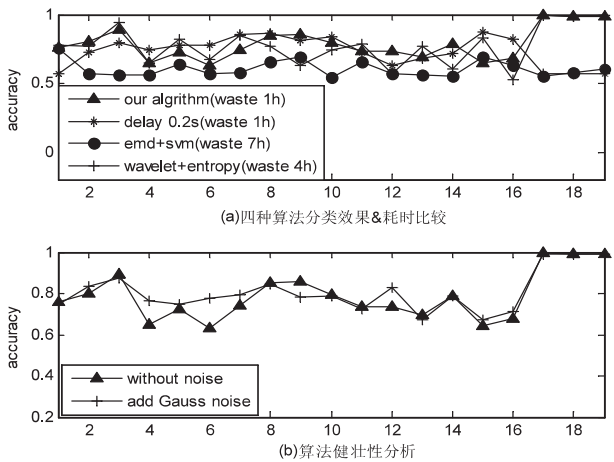


图4 算法分类效果以及健壮性分析

4 结束语

癫痫发作突然、迅猛,严重时可致死。纯人力检测癫痫发作时间较长且极有可能误诊,影响最终治疗。为了提高癫痫患者的生活质量,降低癫痫发作的致死率,文中提出了一种基于互信息和支持向量机的自动检测算法。互信息可以综合考查所有导联之间信号的传输差异,支持向量机可以在统计样本量较少的情况下,亦能获得良好统计规律,两者结合,使得算法更高效准确。实验结果表明,癫痫脑电信号的互信息与正常脑电信号的互信息具有显著差异,提取互信息的相关统计量作特征向量进行分类具有较高的分类准确率,高达100%,且相较于其他主流算法,文中算法在保证准确率的同时,提高了算法速度。

参考文献:

[1] Huang L Y, Zou J, Ma H J, et al. Brain functional network based on mutual information analysis of EEGs and its application to schizophrenia [J]. Advanced Materials Research, 2013, 718-720: 1440-1445.

[2] 黄瑞梅, 杜守洪, 陈子怡, 等. 基于支持向量机的癫痫脑电信号模式识别研究[J]. 生物医学工程学杂志, 2013, 30(5): 919-924.

[3] 汪春梅, 邹俊忠, 张 见, 等. 基于多分辨分析和近似熵的脑电癫痫波自动检测[J]. 计算机应用与软件, 2009, 26(12): 7-9.

[4] 白冬梅, 邱天爽, 李小兵. 样本熵及在脑电癫痫检测中的应用[J]. 生物医学工程学杂志, 2007, 24(1): 200-205.

[5] 徐永红, 崔 浩, 洪文学, 等. 基于改进多元多尺度熵的癫痫脑电信号自动分类[J]. 生物医学工程学杂志, 2015, 32(2): 256-262.

[6] Li S F, Zhou W D, Yuan Q, et al. Feature extraction and recognition of ictal EEG using EMD and SVM[J]. Computers in Biology and Medicine, 2013, 43(7): 807-816.

[7] 韩 凌, 王 宏. 基于空频域特征分析方法的癫痫发作预测[J]. 仪器仪表学报, 2014, 35(11): 2501-2507.

[8] 张 涛, 陈万忠, 李明阳. 基于 AdaBoost 算法的癫痫脑电信号识别[J]. 物理学报, 2015, 64(12): 419-425.

[9] 李红利, 王 江, 邓 斌, 等. 癫痫脑电的互信息和同步性分析[J]. 计算机工程和应用, 2013, 49(6): 19-22.

[10] Na S H, Jin S H, Kim S Y, et al. EEG in schizophrenic patients; mutual information analysis[J]. Clinical Neurophysiology, 2002, 113: 1954-1960.

[11] Melia U, Guaita M, Vallverdu M, et al. Mutual information measures applied to EEG signals for sleepiness characterization[J]. Medical Engineering and Physics, 2015, 37(3): 297-308.

[12] 李志萍. 基于支持向量机的多通道癫痫发作预测[J]. 计算机工程, 2014, 40(2): 199-202.

[13] 张新静, 徐 欣, 凌至培, 等. 基于最大相关和最小冗余准则及极限学习机的癫痫发作检测方法[J]. 计算机应用, 2014, 34(12): 3614-3617.

[14] 林秀晶, 夏勇明, 钱松荣. 基于支持向量机及特征选择的单通道脑电波睡眠分期研究[J]. 生物医学工程学杂志, 2015, 32(3): 503-507.