

一种基于支持向量机技术的癫痫脑电棘尖波识别方法

邱天爽¹, 郑效来¹, 鲍海平², 赵庚申¹

(1. 大连理工大学电子与信息工程学院, 大连 116024; 2. 大连医科大学附属第二医院, 大连 116023)

摘要: 支持向量机是一种基于统计学习理论的新型学习机。文章提出一种基于支持向量机的癫痫脑电特征提取与识别方法, 充分发挥其泛化能力强的特点, 在与神经网络方法的比较中, 表现出较低的漏检率和较好的鲁棒性, 有深入研究的价值和良好的应用前景。

关键词: 脑电信号; 癫痫; 支持向量机; 特征识别
中图分类号: Q959.7

1 引言

脑电图 (EEG) 是通过仪器记录脑细胞的电活动而形成的, 对许多神经系统疾病的诊断、病情监测及疗效观察有十分重要的意义, 脑电图检查是诊断癫痫的一个非常重要的手段。癫痫的自动检测与识别不仅能够克服传统视觉检测方法的种种弊端, 降低医疗工作者的劳动量, 而且在一定程度上有助于统一判断标准, 减少主观因素的负面影响。

癫痫的脑电图特征是发作性高波幅电活动, 其波形和频率各异, 其中某些形式的电活动, 具有特殊的诊断意义^[1,2]:

1) 棘波: 周期在 83 ms 以下, 波幅多在 100 μV 以上的一种快波, 是癫痫的特异性放电, 波形较陡成为棘状波。

2) 尖波: 又称锐波, 波形相似于棘波, 周期通常在 83~200 ms, 波幅在 100 μV 以上。

3) 棘慢波综合: 随棘波之后有 200~500 ms 的慢波, 是一种特殊类型的癫痫发作波, 多见于失神发作。

棘波和尖波是癫痫在脑电图中的主要表现, 在自动识别时, 对棘波和尖波两种波形不作区分, 统称为癫痫瞬变现象。目前, 提出的大部分方法都是针对棘波和尖波, 只有少数方法对慢波进行识别, 所以本文主要考虑对棘、尖波的识别。

2 支持向量机原理和算法

支持向量机 (support vector machine, SVM) 是一种基于统计学习理论的新型学习机, 是由前苏

联教授 Vapnik 最早提出的^[3]。与传统的学习方法不同, SVM 遵循基于 VC 维 (Vapnik-Chervonenkis dimension) 理论结构风险最小化的原则^[4]。其内容是: 通过某种事先选择的非线性映射 $\varphi(\bullet)$ 将输入向量 X 映射到一个高维特征空间 Z , 在这个空间中构造最优分类超平面, 从而使正确和错误样本之间的分离界限达到最大^[5]。从概念上说, 支持向量是那些离决策平面最近的数据点, 它们决定了最优分类超平面的位置, 如图 1 所示。经过分析证明, 最大化分离界限等价于使权向量的 Euclidean 范数最小^[4-6]。

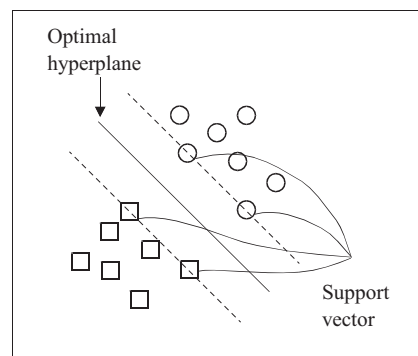


Fig.1 Illustration of the idea of an optimal hyperplane for linearly separable patterns.

□: The first class; ○: The second class

收稿日期: 2004-12-14

基金项目: 国家自然科学基金项目 (30170259, 60172072, 60372081); 辽宁省科学技术基金资助项目 (2001101057)

通讯作者: 邱天爽, 电话: (0411)84706005-8309,

E-mail: qiutsh@dlut.edu.cn

2.1 支持向量机的原理

对于一个两模式分类问题（第一类表示为 1，第二类表示为 -1），分类超平面可表示为：

$$\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) + b = 0 \quad (1)$$

其中， \mathbf{w} 为权矢量， $\boldsymbol{\varphi}(\mathbf{x})$ 是输入矢量 \mathbf{x} 映射到高维空间的特征矢量， b 为阈值。

那么，寻找最优超平面可以归纳为求解一个二次规划问题，使(2)式最小。

$$\Phi(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} \quad (2)$$

约束条件为

$$y_i(\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b) \geq 1, \quad i=1, \dots, N \quad (3)$$

其中， y_i 表示第 i 个样本的期望输出， N 为输入的样本数。

(2)式和(3)式给出的是可分模式下的解决方法，在实际应用中，大部分问题是不可分的，所以需要引入松弛变量 ξ_i ($\xi_i \geq 0, i=1, \dots, N$)，则约束条件(3)可以写成

$$y_i(\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad i=1, \dots, N \quad (4)$$

为了保证不可分问题的分类误差最小，代价函数表示如下：

$$\Phi(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \quad (5)$$

其中， C 称为惩罚因子。约束条件为

$$(1) y_i(\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b) \geq 1 - \xi_i$$

$$(2) \xi_i \geq 0, \quad i=1, \dots, N$$

利用 Lagrange 乘子法得到此问题的对偶问题^[4]，即求下式的最大化问题

$$J(\boldsymbol{\alpha}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \boldsymbol{\varphi}(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x}_j) \quad (6)$$

其中 $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]$ 表示 Lagrange 乘子。约束条件为

$$(1) \sum_{i=1}^N \alpha_i y_i = 0$$

$$(2) 0 \leq \alpha_i \leq C, \quad i=1, \dots, N$$

最优权值定义为

$$\mathbf{w} = \sum_{i=0}^N \alpha_i y_i \boldsymbol{\varphi}(\mathbf{x}_i) \quad (7)$$

令 $K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}^T(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x}_j)$ ，(6)式可以写为

$$J(\boldsymbol{\alpha}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (8)$$

约束条件为

$$(1) \sum_{i=1}^N \alpha_i y_i = 0$$

$$(2) 0 \leq \alpha_i \leq C, \quad i=1, \dots, N$$

其中 $K(\mathbf{x}_i, \mathbf{x}_j)$ 称为内积核函数，解优化问题(8)得到 Lagrange 乘子矢量 $\boldsymbol{\alpha}$ ，结合(7)式可以得到判决函数如下

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^N \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \right) \quad (9)$$

其中， $\text{sgn}(\cdot)$ 为符号函数。通过分析发现， $\alpha_i \neq 0$ 对应的向量 \mathbf{x}_i 就是支持向量，因此，判决函数可以写为

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^M \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \right) \quad (10)$$

其中， M 表示支持向量的个数。而对于 $0 < \alpha_i < C, i=1, \dots, M$ ，支持向量满足下式

$$y_i(\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b) = 1 \quad (11)$$

可以由(11)式确定阈值 b ，最终得到判决函数。

采用不同的核函数，就能够造出不同类型的 SVM，常见的 SVM 主要有以下三种^[4,5]：（1）多项式 SVM；（2）径向基函数 SVM；（3）两层感知机 SVM。使用较多的是前两种 SVM。

2.2 SVM 相关算法

通过以上讨论可以看出，SVM 的训练问题是一种解二次规划（quadratic programming, QP）的问题。在其发展的十几年中，出现了许多学习算法，其中比较经典的如 Vapnik 提出的 chunking 方法，Joachims 提出的 SVM^{lib} 算法，Platt 提出的序贯最小优化方法（sequential minimal optimization, SMO）^[7,10-12]等。经过比较，本文采用的是 SMO 方法，其主要思想是将规模很大的 QP 问题分解为一系列规模较小的 QP 问题，每次迭代中利用优化算法求解一个 QP 问题，更新 $\boldsymbol{\alpha}$ 的一个分量子集，即工作集，其容量大小为 2。本文采用 SMO 算法有以下几个原因：（1）工作集的容量为 2，使每次迭代过程中，分解的子 QP 问题存在解析解^[7]；（2）一般 SVM 的训练样本较多，此方法解决了普通计算机容量限制的问题；（3）得到的解是全局最优的；（4）收敛速度较快。

3 癫痫特征波参数提取方法

癫痫棘尖波时域模型如图 2 所示：

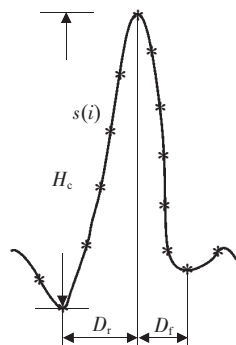


Fig.2 The time-domain model of spike (sharp) waves

棘尖波一般由两部分组成，上升支和下降支，即两个相邻幅度极值之间的部分，Gotman 等^[9]称之为“半波”。由图 2 所示的模型定义四个变量：波形峰 - 峰值 H_c ，离散样本值 $s(i)$ ，以及 D_r 和 D_f 。根据脑电专家的经验：癫痫放电时，有许多棘尖波附着在其后的慢波上，使上升支或下降支成为不明显的切迹。所以，本文使用如下的特征参数来识别癫痫特征波和非癫痫特征波：绝对高度： $H_a = \max_{i \in \text{spike}}(s(i))$ ，其中 $s(i)$ 表示样本值，spike 为单个棘尖波样本点的集合；

一阶差分： $\Delta s = \max_{i \in \text{spike}}(s(i) - s(i-1))$ ，即棘尖波波形中相邻两点差的最大值；

斜度： $k = \max(\frac{H_c}{D_r}, \frac{H_c}{D_f})$ ；

波形的宽度： $W_s = D_r + D_f$ ；

波形的形态： $F_s = \frac{H_c}{W_s}$ 。

从训练样本中提取出特征参数，将这些参数（即向量 x ）输入到 SVM 训练，得出支持向量 ($\alpha_i \neq 0$ 对应的那些样本)，进而求得权矢量 w 和阈值 b ，得到分离癫痫特征波和非癫痫特征波的最优超平面，也就是说，本文通过有监督学习方法来构造 SVM。

4 实验结果及分析

为了验证 SVM 在癫痫特征识别中的作用，本文使用的是来自北京天坛医院和大连儿童医院的真实脑电数据，共 5 个癫痫患者病例（总计 1 024 s），采样频率均为 256 Hz，将其分为两部分：训练样本集（456 s）和检验样本集（568 s）。由脑电专家对这 5 例脑电图进行检查，共发现 720 个棘尖波。为讨论方便，本文将 SVM 与 BP 神经网络的识别结果进行比较（训练样本相同）。目前，国际上还没有一种标准的性能评估方法，常用的方法是以误检率和漏检率来评估算法的性能^[8,9]，误检率为检出的非棘尖波数与检测出的波形总数之比，漏检率为漏检的棘尖波数与实际棘尖波总数之比。两种方法的性能比较如表 1 所示。

Table 1 Comparison of result of the two methods

Comparison between two methods	Spikes detected	Spikes detected with false	Spikes undetected	Undetected rate (%)	False rate (%)
Radials basis function	697	60	23	3.2	7.9
BP neural network	678	9	42	5.8	1.3

从表中可以看出，SVM 方法的漏检率比神经网络的方法低，为医疗工作者提供了较可靠的信息，但误检率比神经网络方法大。通常误检率与漏检率不可兼得，一般期望结果的漏检率要尽可能小，在节省医疗工作者劳动量的同时，保证不漏掉病情的重要信息。在实验过程中发现，癫痫放电后期，棘尖波变得不明显，成为附着在慢波上的切

迹，这些切迹对病情的诊断是很有意义的。图 3 是两种方法对同一段脑电信号（属于检验样本集）的识别结果，用“o”定位检测到的棘尖波，比较图中的结果不难看出，在癫痫放电后期，SVM 对附着在慢波上的棘尖波切迹的检测效果较好，而神经网络的结果漏掉了许多诊断信息，这充分说明了 SVM 泛化能力强，有良好的鲁棒性。

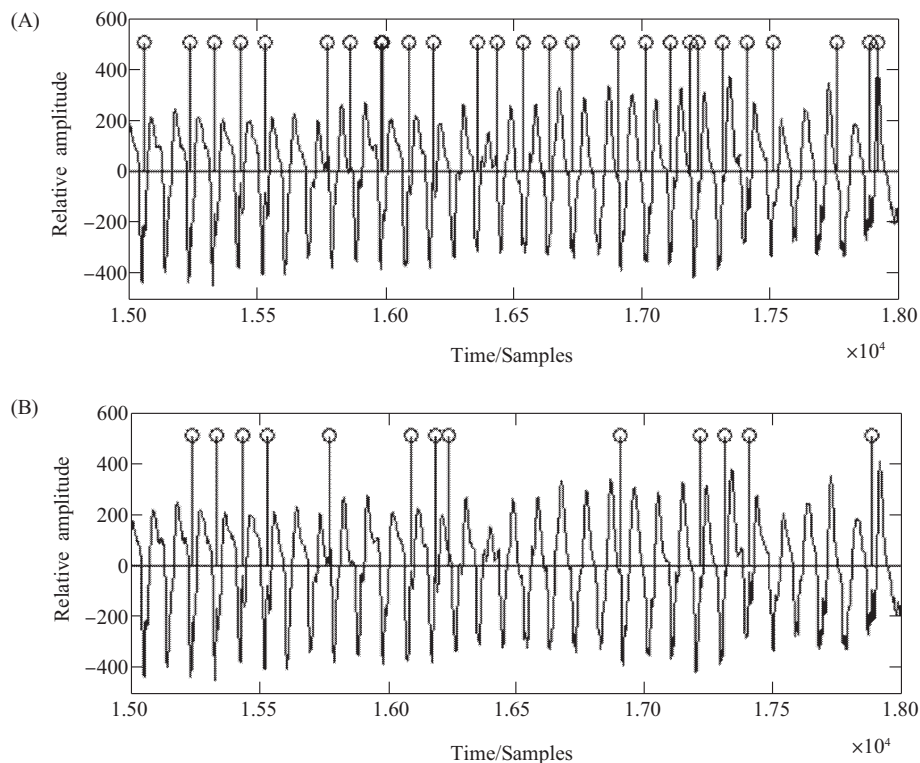


Fig.3 Illustration of the comparison of detection results of the two methods in post-icta.
(A) SVM; (B) Neural network.

5 讨 论

SVM 理论作为一种新的机器学习方法, 具有近似最优地解决模式分类问题的内在能力, 而且在没有问题域相关知识时设计的机器可以具有很好的性能。本文提出一种基于 SVM 的癫痫脑电信号的识别分类方法, 充分发挥其泛化能力强的特点, 在对真实脑电信号处理中表现出良好的性能, 漏检率为 3.2%, 误检率为 7.9%, 充分说明了 SVM 对脑电信号处理及相关问题有良好的应用前景。然而, 对 SVM 来说, 得到良好性能的代价是较大的计算复杂度, 而且算法不能很好地控制支持向量的选择, 这些问题还有待于深入研究。希望通过本文的研究工作, 能给脑电信号处理或相关领域研究提供一些参考。

参考文献:

[1] 周昌贵. 临床脑电图手册. 成都: 四川辞书出版社, 1990
[2] 谭郁玲. 临床脑电图与脑地形图学. 北京: 人民卫生出版社, 2001

[3] Vapnik V. Translated by Kotz S. Estimation of dependences based on empirical data. New York: Springer-Verlag, 1982
[4] Haykin S. Neural networks: a comprehensive foundation. Prentice Hall, Tsinghua University Press, 2001
[5] Vapnik V (著), 张学工 (译). 统计学习理论的本质. 北京: 清华大学出版社, 2000
[6] Schölkopf B, Sung KK, Burges JC, Girosi F, Niyogi P, Poggio T, Vapnik V. Compare support vector machine with gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing*, 1997,45(11):2758~2765
[7] Platt JC. Fast training of support vector machines using sequential minimal optimization. In: Schölkopf B, Burges C, Smola AJ, ed. *Advances in kernel methods-support vector learning*. Cambridge: MIT Press, 1999. 185~208
[8] Wilson SB, Emerson R. Spike detection: a review and comparison of algorithms. *Clinical Neurophysiology*, 2002,113(12): 1873~1881
[9] 宦飞, 王志中, 郑崇勋. 基于时频分析检测 EEG 中癫痫样棘 / 尖波的方法. *生物物理学报*, 2000,16(3):539~546
[10] 王国胜, 钟义信. 支持向量机的若干新进展. *电子学报*, 2001, 29(10):1397~1400
[11] 刘江华, 程君实, 陈佳品. 支持向量机训练算法综述. *信息与控制*, 2002,31(1):45~50
[12] 高学, 金连文, 尹俊勋. 一种基于支持向量机的手写汉字识别方法. *电子学报*, 2002,30(5):651~654

A SUPPORT VECTOR MACHINE BASED METHOD OF EPILEPSY SPIKE AND SHARP WAVES RECOGNITION IN EEG

QIU Tian-shuang¹, ZHENG Xiao-lai¹, BAO Hai-ping², ZHAO Geng-shen¹

(1. Department of Electronic Engineering, Dalian University of Technology, Dalian 116024, China;

2. The 2nd Attached Hospital to Dalian Medical University, Dalian 116023, China)

Abstract: The support vector machine (SVM) is a new learning machine based on the statistical learning theory. An SVM based method of epilepsy feature extraction and recognition is proposed. The data analysis shows that the SVM is very robust in identification of spike and sharp waves compared with that in back-propagation neural network method.

Key Words: EEG; Epilepsy; Support vector machine; Feature recognition

欢迎投稿和订阅 《生理学报》

《生理学报》是由中国生理学会和中国科学院上海生命科学研究院主办、中英文兼登的学术期刊。现任主编为中国生理学会理事长姚泰教授。双月刊，大16开铜版纸印刷，96页。主要刊登生理学和相关生命科学的研究论文、研究快报、实验技术及以本人研究工作为主的综述。该刊的前身是《中国生理学杂志》，于1927年创刊，是我国最早出版的生理学学术期刊。从1953年起，改名为《生理学报》。在国内外享有较高的声誉。在所刊登的论文全面反映了中国生理学界在各个领域的最新成果和研究进展，具有较高的学术水平和创新性。

《生理学报》已被美国《生物学文摘》、《化学文摘》、《医学索引》、MEDLINE/PubMed、BIOSIS previews数据库，荷兰《医学文摘》，俄罗斯《文摘杂志》以及中国的《中国科学引文数据库》，中国期刊网，万方数据网，重庆维普资讯科技期刊数据网等国内外检索期刊或数据库收录，是中国自然科学核心期刊和生物类核心期刊。

《生理学报》1999年荣获首届国家期刊奖；2000年获中国科学院优秀科技期刊特别奖；2001年入选双奖期刊“中国期刊方阵”；2003年又荣获第二届国家期刊奖；2005年荣获第三届国家期刊奖百种重点科技期刊。

《生理学报》提供即期全文上网，E-mail推送下期目录和在线网上投稿、查询的服务。稿件快速处理，1~2个月反馈一审结果。已接受文章摘要提前上网。学术质量好的英文稿件优先发表。

2005年起，本刊全面实行网上投稿、网上审稿。

欢迎您向《生理学报》投送优质的稿件！

热忱欢迎广大读者、作者和商家订阅、投稿和发布广告！

编辑部地址：上海市岳阳路319号31B楼405室，邮编：200031

电话：021-54922832；传真：021-54922833

网址：<http://www.actaps.com.cn>；E-mail:actaps@sibs.ac.cn

国内统一刊号：CN 31-1352/Q

国际标准刊号：ISSN 0371-0874

邮发代号：4-157

订阅：全国各地邮局，也可向编辑部邮购 定价：23元/期，全年138元