

面向癫痫脑电图信号识别的径向基最小最大概率分类树

邓赵红* 陈俊勇 刘解放 王士同
(江南大学数字媒体学院 无锡 214122)

摘要: 脑电图(EEG)信号检测和识别是癫痫病的重要诊断手段。径向基函数神经网络具有出色的逼近能力和泛化性能,能直接识别出不同状态的脑电信号,但其透明性和可解释性差,忽视了不同类别数据间可分性的不同。对此,该文提出一种基于径向基函数神经网络和最小最大概率决策技术的分类树,采用一对一策略和排除法,更多考虑了类间可分性的不同。针对脑电信号识别的实验表明,所提方法结构清晰,分类能力强,可解释性更好。

关键词: 脑电信号; 径向基函数神经网络; 最小最大概率; 分类树

中图分类号: R741.044; TP183

文献标识码: A

文章编号: 1009-5896(2016)11-2848-08

DOI: 10.11999/JEIT160082

Radial Basis Minimax Probability Classification Tree for Epilepsy ElectroEncephaloGram Signal Recognition

DENG Zhaohong CHEN Junyong LIU Jiefang WANG Shitong
(School of Digital Media, Jiangnan University, Wuxi 214122, China)

Abstract: ElectroEncephaloGram (EEG) signal detection and recognition is an important diagnostic method for the epilepsy. Radial Basis Function (RBF) neural network has excellent performance on approximation and generalization, and can directly recognize EEG signals in different states. However, its transparency and interpretability are low, and it also ignore the different separabilities between different classes of data. In this paper, a classification tree based on RBF neural networks and minimax probability decision technique is proposed, using one-against-one and exclusive method and paying much attention to the different separabilities among classes. Experiments on EEG signals show that the proposed method has clear structure, strong classification ability and better interpretability.

Key words: ElectroEncephaloGram (EEG) signal; Radial basis function neural network; Minimax probability; Classification tree

1 引言

癫痫是由大脑神经元突发性异常过度放电引发的大脑短暂性功能障碍,年发病率较高。脑电图信号的检测和识别是诊断癫痫病最为重要的手段。多种特征提取和智能识别的方法已被应用于癫痫脑电信号的识别,主要包括傅里叶变换^[1,2]、小波分解^[3-5]等特征抽取方法和决策树^[2]、模糊系统^[3]、人工神经网络^[6-9]等分类方法。

文献[9]将径向基函数神经网络用在癫痫(ElectroEncephaloGram, EEG)信号的分类识别

中。径向基函数(Radial Basis Function, RBF)神经网络具有一般神经网络的优点,如很强的非线性拟合能力及鲁棒性,此外,它还具有全局最佳逼近特性,不存在局部极小问题。但径向基函数神经网络的不足是其可解释性差,不易解释其推理过程和依据,因而不够透明。在处理多分类问题时,径向基函数神经网络一般不考虑各类之间不同的可分性,利用回归的方法直接进行多分类,从而导致分界面也过于复杂、难以描述。

文献[2]采用基于决策树和快速傅里叶变换的混合模型来对EEG信号进行分类。决策树分类器的优势正在于其可解释性和透明性,易于理解。然而,决策树是层次结构的,不可避免地存在自上而下的“误差积累”^[10],即如果在某个节点上发生了分类错误,则会把错误延续到后续节点上。为了降低积累的误差,应当优先使用可分性强的分类器,即先做把握大性的决策。

本文针对上述癫痫脑电信号识别方法存在的不足,通过引入最小最大概率学习机制来探讨相应的

收稿日期: 2016-01-19; 改回日期: 2016-06-08; 网络出版: 2016-09-01

*通信作者: 邓赵红 dengzhaohong@jiangnan.edu.cn

基金项目: 江苏省杰出青年基金(BK20140001), 上海市科学技术委员会扬帆项目(14YF1411000), 上海市教委创新项目(14YZ131)

Foundation Items: The Youth Fund of Jiangsu Province (BK20140001), YangFan Project of Shanghai Municipal Science and Technology Commission(Grant No. 14YF1411000), The Innovation Program of Shanghai Municipal Education Commission (Grant No. 14YZ131)

改进方法。文献[11,12]中提出的最小最大概率机在完成模型参数学习的同时，可实现错分概率上界的最小化。理论上说，此错分概率上界的指标越小，表明两类数据间的可分性越强，分类模型的可靠性越好，即更有把握得到正确的分类结果。文献[13]将之用于 TSK 型模糊系统分类器的学习，得到了透明性、可解释性更好的分类模型。考虑到 RBF 神经网络与模糊推理系统在一定的条件下存在等价性^[14]，可采用一对一的策略和排除法，将最小最大概率机用于 RBF 神经网络分类器的学习，进而基于错分概率上界可构建更为透明和可靠的分类树。

本文首先介绍了 RBF 神经网络，最小最大概率机和分类决策树的相关知识，而后提出了一种基于 RBF 神经网络和最小最大概率决策技术的多分类决策树，并将之用于癫痫 EEG 信号的分类与识别。实验部分，通过 3 组不同分类目标的对比实验，证实了所提方法有着更好的可靠性和可解释性。

2 相关工作

2.1 RBF 神经网络

RBF 神经网络(Radial Basis Function Neural Network, RBFNN)^[9]能够以任意精度逼近任一连续函数，具有良好的泛化能力。一个最简单的 RBF 神经网络是具有单隐层、单输出的 3 层结构的前馈神经网络，如图 1 所示。对于分类问题，通过隐层的非线性映射，通常可将原始特征空间上线性不可分的问题变换为高维特征空间上易于线性可分的问题。RBFNN 的输出可用隐节点输出的线性组合来表示。

$$y = \sum_{k=1}^K f_k^T(\mathbf{x})w_k - b = \varphi^T(\mathbf{x})\mathbf{w} - b \quad (1)$$

式(1)中， K 为神经元个数， b 为阈值， $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_d]^T$ 为输入向量， $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_K]^T$ 为输出层权值。

$$\varphi(\mathbf{x}) = [f_1(\mathbf{x}) \ f_2(\mathbf{x}) \ \dots \ f_K(\mathbf{x})]^T \quad (2)$$

这里 f_k 表示隐层中第 k 个神经元的径向基函数。一般常取高斯型函数作为径向基函数，即

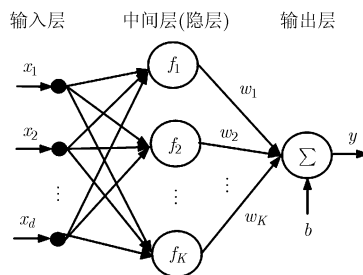


图 1 单隐层单输出 RBF 神经网络

$$f_k(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_k\|^2}{2\sigma_k^2}\right) \quad (3)$$

其中， \mathbf{c}_k 和 σ_k 分别表示径向基函数的中心和宽度。通常可通过聚类等划分技术确定上述参数，例如，可用 FCM^[15]算法对输入数据集进行聚类求得划分矩阵 $\mathbf{U} = [u_{ik}] \in \mathbf{R}^{N \times K}$ ，其中 $u_{ik} \in [0, 1]$ 表示第 i 个输入数据 $\mathbf{x}_i = [x_{i1} \ x_{i2} \ \dots \ x_{id}]^T$ 隶属于第 k 类的程度，进而式(3)中的中心和宽度分量可估计如式(4)：

$$\left. \begin{aligned} c_{kj} &= \frac{\sum_i u_{ik} x_{ij}}{\sum_i u_{ik}} \\ \delta_{kj} &= h \cdot \frac{\sum_i u_{ik} (x_{ij} - c_{kj})^2}{\sum_i u_{ik}} \\ \sigma_{kj} &= \sqrt{\delta_{kj}} \end{aligned} \right\} \quad (4)$$

式(4)中， h 是一个用户可调的超参数。

2.2 最小最大概率机

最小最大概率机 (Minimax Probability Machine, MPM)^[11,12]是基于错分概率上界最小化而设计的二分类模型。设有两类服从均值和协方差分别为 (\mathbf{u}_+, Σ_+) 和 (\mathbf{u}_-, Σ_-) 分布的 d 维向量 \mathbf{x}_+ 和 \mathbf{x}_- 。最小最大概率机定义了如式(5)的优化目标来确定一个分类超平面 $\mathbf{w}^T \mathbf{x} = b$ 。

$$\begin{aligned} \max_{\alpha, \mathbf{w}=0, b} \alpha, \quad \text{s.t.} \quad & \inf_{\mathbf{x} \sim (\mathbf{u}_+, \Sigma_+)} \text{pr}(\mathbf{w}^T \mathbf{x} \geq b) \geq \alpha, \\ & \inf_{\mathbf{x} \sim (\mathbf{u}_-, \Sigma_-)} \text{pr}(\mathbf{w}^T \mathbf{x} \leq b) \geq \alpha \end{aligned} \quad (5)$$

其中， $\inf_{\mathbf{x} \sim (\mathbf{u}_+, \Sigma_+)} \text{pr}(\mathbf{w}^T \mathbf{x} \geq b)$ 表示当条件为 $\mathbf{w}^T \mathbf{x} \geq b$ ， $\mathbf{x} \sim (\mathbf{u}_+, \Sigma_+)$ 时概率的下确界。式(5)显示出，对于服从不同分布的两类数据，存在一个最优的超平面 $(\mathbf{w}^*)^T \mathbf{x} = b^*$ ，使得正确分类概率的下界 α 达到最大，也即错分率的上界 $1 - \alpha$ 达到最小。

进一步，为了便于求解式(5)对应的优化问题，文献[11,12]推出了如下的定理。

定理 如果 $\mathbf{u}_+ = \mathbf{u}_-$ ，那么式(5)的解无意义，最优的最坏情况错分率上界 $1 - \alpha^* = 1$ ；否则，通过求解如下的凸集优化问题，得到 \mathbf{w} 的最优解 \mathbf{w}^* 。

$$\begin{aligned} \kappa(\alpha)^{-1} &= \min_{\mathbf{w}} \sqrt{\mathbf{w}^T \Sigma_+ \mathbf{w}} + \sqrt{\mathbf{w}^T \Sigma_- \mathbf{w}}, \\ \text{s.t.} \quad & \mathbf{w}^T (\mathbf{u}_+ - \mathbf{u}_-) = 1 \end{aligned} \quad (6a)$$

进一步， b 的最优解可置为

$$b^* = (\mathbf{w}^*)^T \mathbf{u}_+ - \kappa(\alpha) \sqrt{(\mathbf{w}^*)^T \Sigma_+ (\mathbf{w}^*)} \quad (6b)$$

式(6a)中， $\kappa(\alpha)^{-1} = \sqrt{(1 - \alpha)/\alpha}$ 。利用得到的最优的

$\kappa^*(\alpha)$, 即可求得正确分类概率的下界, 即

$$\alpha^* = \frac{1}{1 + \left(\sqrt{(\mathbf{w}^*)^T \boldsymbol{\Sigma}_+ (\mathbf{w}^*)} + \sqrt{(\mathbf{w}^*)^T \boldsymbol{\Sigma}_- (\mathbf{w}^*)} \right)^2} \quad (6c)$$

根据优化理论可知, 只要 $\boldsymbol{\Sigma}_+$ 和 $\boldsymbol{\Sigma}_-$ 是正定的, 超平面 $(\mathbf{w}^*)^T \mathbf{x} = b^*$ 就是唯一的。文献[12]给出了简洁有效的基于最小二乘法的迭代求解算法。如图 2 所示, 最小最大概率机学习的过程, 就是通过不断迭代来最大化马氏距离(Mahalanobis distance) [12,16] $\kappa(\alpha)$, 直到两个椭圆相切, 切线即为分界, 充分地保证了错分概率最小, 可见其具有良好的几何解释性。 α 的大小充分反映了两类数据间的可分性, α 指标越趋近于 1, 表明两类数据间的可分性越强, 分类模型越可靠。

此外, 由于线性的最小最大概率机的分类性能有限, 文献[11,12]通过引入核技巧, 给出了核化版本的最小最大概率机来实现非线性分类, 求解过程和线性版本大致相同。

2.3 分类决策树

决策树(Decision Tree, DT)^[2,10]是一种非度量方法, 结构形如一棵倒置的树, 其由节点和有向边组成, 其中节点又分作内部节点和叶节点。内部节点表示一次查询操作, 叶节点代表一个结果。决策树被用来分类时, 数据由根节点出发, 通过对一系列问题的回答, 最终到达叶节点, 得到叶节点所代表的类标记。其中, 后一个问题的提法依赖于前一个问题的回答。

相较于神经网络等度量方法, 分类树的优点是可解释性强, 分类速度快, 通过有限的查询, 即可确定数据所属类别。

3 最小最大概率 RBF 神经网络

文献[13]采用最小最大概率机来训练 TSK 型模糊系统分类器, 得到了透明性、可解释性更好的分类模型。考虑到 RBF 神经网络与 TSK 型模糊系统

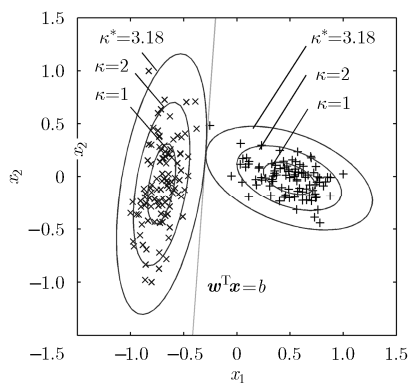


图 2 最小最大概率机的几何解释

在一定的条件下存在等价性^[14], 基于最小最大概率机和 RBF 神经网络, 本文定义了如下形式的优化目标来学习一个可进行二分类的最小最大概率 RBF 神经网络 (Minimax Probability RBF Neural Network, MP-RBFNN)。

$$\begin{aligned} \max_{\alpha, \mathbf{w} \neq 0, b} \alpha, \quad \text{s.t.} \quad & \inf_{\varphi(\mathbf{x}) - (\mathbf{u}_{\varphi(+)} \cdot \boldsymbol{\Sigma}_{\varphi(+)}^{-1})} \text{pr}(\mathbf{w}^T \varphi(\mathbf{x}) \geq b) \geq \alpha, \\ & \inf_{\varphi(\mathbf{x}) - (\mathbf{u}_{\varphi(-)} \cdot \boldsymbol{\Sigma}_{\varphi(-)}^{-1})} \text{pr}(\mathbf{w}^T \varphi(\mathbf{x}) \leq b) \geq \alpha \end{aligned} \quad (7)$$

其中, $\varphi(\mathbf{x})$ 如式(2)定义, 表示 \mathbf{x} 通过 RBF 神经网络映射得到的新特征空间上对应的向量。在此, 可用数据样本集 $\mathbf{X}_\varphi = \{\varphi(\mathbf{x}_i)\}$ 来估算映射后得到的数据样本的均值和协方差。记 $\tilde{\mathbf{u}}_\varphi$ 为样本集 \mathbf{X}_φ 的估计均值, 那么对应的协方差可按式(8)估算,

$$\tilde{\boldsymbol{\Sigma}}_\varphi = (\mathbf{X}_\varphi - \tilde{\mathbf{u}}_\varphi \times \mathbf{1}_N^T)(\mathbf{X}_\varphi - \tilde{\mathbf{u}}_\varphi \times \mathbf{1}_N^T)^T / N \quad (8)$$

其中, $\mathbf{1}_N$ 表示 N 维的全一列向量。

将式(8)以及 $\tilde{\boldsymbol{\Sigma}}_{\varphi(+)}$ 和 $\tilde{\boldsymbol{\Sigma}}_{\varphi(-)}$ 代入到式(6a)中, 得到式(9)的优化目标:

$$\begin{aligned} \kappa(\alpha)^{-1} = \min_{\mathbf{w}} & \sqrt{\mathbf{w}^T \tilde{\boldsymbol{\Sigma}}_{\varphi(+)} \mathbf{w}} + \sqrt{\mathbf{w}^T \tilde{\boldsymbol{\Sigma}}_{\varphi(-)} \mathbf{w}}, \\ \text{s.t.} \quad & \mathbf{w}^T (\tilde{\mathbf{u}}_{\varphi(+)} - \tilde{\mathbf{u}}_{\varphi(-)}) = 1 \end{aligned} \quad (9)$$

如此便可用文献[12]中的方法来求解。

MP-RBFNN 的特点是, 随参数学习所得的 α 指标可刻画两类数据间的可分性, 衡量分类模型的可靠性。根据分类问题的复杂性, 通过调整隐层神经元数目, 可在提升分类精度和控制模型复杂度之间取得平衡。

4 MP-RBFNN 与分类树

4.1 多分类问题

经典的 RBF 神经网络通常利用回归的方式进行多分类, 例如, 对于 3 类分类问题, 训练时期望的输出可表示为 $[0 \ 0 \ 1]^T$, $[0 \ 1 \ 0]^T$ 和 $[1 \ 0 \ 0]^T$, 测试时则把实际输出分量最大的下标作为判定的类别, 相当于用到了 3 个“一对多”(One-Against-Rest, OAR)的二元分类器。而文献[17,18]指出多数情况下“一对一”(One-Against-One, OAO)策略要优于 OAR, 因为 OAR 策略将余下的类别混为一类, 易造成分类界面的复杂化。

OAO 策略可将一个完整的多分类问题化解成多个二分类子问题, 最终训练得到 $M(M-1)/2$ 个二元分类器, 相较于 OAR 策略, 每一个二分类子问题难度小些, 易于找到简单有效的分界面, 也易于解释。利用 OAO 策略进行测试时, 一般配合采用的是“投票法”。然而, 投票法的问题是会出现多个类的票数相同的情况(拒分现象), 且对每一个输入数据都需要 $M(M-1)/2$ 次比较。

4.2 径向基最小最大概率分类树

在 OAO 策略下，为了避免投票法的拒分问题和提升分类模型的效率，本文采用排除法来构建分类决策树，树的每个内部节点都是一个二元分类器。即沿着树生长的方向做排除法，只需 $M - 1$ 次比较，即可获得分类结果。同时，为了降低树形结构固有的“误差积累”，本文将充分利用最小最大概率技术所提供的 α 指标，优先选用 α 指标大的二元分类器，即先做把握性大的分类。

本文提出的径向基最小最大概率分类树 (Radial Basis Minimax Probability Classification Tree, RB-MP-CT)，训练和测试过程可如表 1 所示。

表 1 RB-MP-CT 的训练和测试过程

训练过程：	
步骤 1	指定神经元个数，按式(4)得到各径向基函数的中心和宽度。
步骤 2	根据式(2)，式(3)， M 类训练数据通过 RBF 网络映射到新的特征空间中。
步骤 3	按照 OAO 的策略，利用最小最大概率机训练出各二元分类器，得到 $M(M - 1)/2$ 组最优的 w 和 α 指标。
步骤 4	记 $C(i : j)$ 表示 i -Against- j 所得分类器， $\alpha(i : j)$ 表示与之对应的 α 指标。将 α 最大的分类器作为分类树的根节点。
步骤 5	假设当前节点所用分类器为 $C(i : j)$ ，若分类结果为样本不属于第 i 类，则其子节点可用分类器为 $C = C \setminus i$ (排除了所有与 i 有关的分类器)，所用分类器为 C 中与 j 有关的 α 指标最大者；若分类结果为样本不属于第 j 类，则其子节点可用分类器为 $C = C \setminus j$ ，所用分类器为 C 中与 i 有关的 α 指标最大者。重复此步骤，直至用尽所有 $M(M - 1)/2$ 个分类器，构造出一棵 M 层的完全二叉树，其叶节点即为各类的类标签。
测试过程：	
步骤 1	根据式(2)，式(3)，测试数据通过径向基网络的作用转化到新的特征空间中。
步骤 2	数据通过分类树，按照排除法，一层一层往下，经历 $M - 1$ 次比较，由根节点到达叶节点，获得最终的类标签。

在余下的分类器中选择子节点时，通常不会出现两个二元分类器的 α 指标大小相等的情况。而如果出现这样的小概率事件，一般可以简单处理，任选其中一个分类器作为子节点。

为了直观地解释多分类决策树，现举例如下进行说明：设有 4 类数据，利用 OAO 策略共得到 6 个二元分类器， α 之间的大小关系为 $\alpha(1 : 4) > \alpha(2 : 4) > \alpha(1 : 3) > \alpha(3 : 4) > \alpha(2 : 3) > \alpha(1 : 2)$ ，那么 $C(1 : 4)$ 先被选为分类树的根节点。再按照训练步骤 5，先向左分支，认为样本不属于第 4 类，那么接下来子节点待选的分类器为 $C(1 : 2)$ ， $C(1 : 3)$ 和 $C(2 : 3)$ ，从中选择与第 1 类有关的 α 指标最大者，

即为 $C(1 : 3)$ 。再进 1 层，若排除样本属于第 3 类的可能，则最终选择 $C(1 : 2)$ 作为左侧的子节点。如此重复，得到的即是一个 4 层的完全二叉树(如图 3)。

5 实验研究

5.1 实验数据

本部分针对德国波恩大学提供的癫痫 EEG 信号数据设计实验。该数据集分为 5 组，每组数据包含 100 个 23.6 s 长度，采样频率为 173.61 Hz 的 EEG 信号片段(4097 个采样点)。各组数据的具体情况如表 2 所示。图 4 表示的是各组数据中第 1 个 EEG 信号的波形(即 A001, B001, C001, D001 和 E001)，更多细节可参见文献[19]。

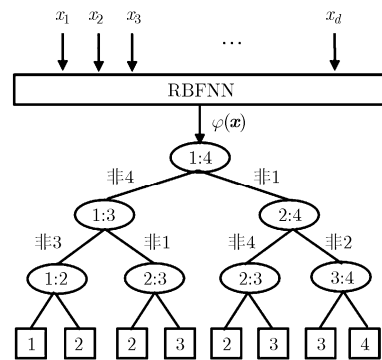


图 3 RB-MP-CT

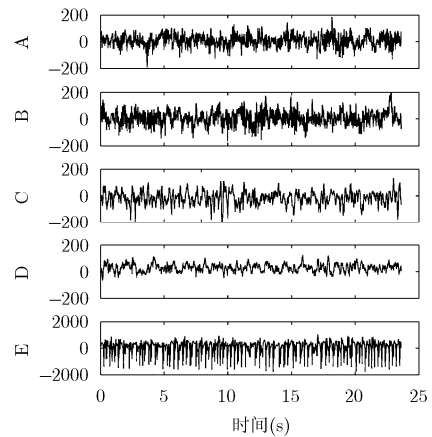


图 4 各组首个 EEG 信号的波形

表 2 EEG 数据集的组成和描述

组	编号范围	描述
健康	A A001-A100	健康者睁眼时测得的 EEG 信号
	B B001-B100	健康者闭眼时测得的 EEG 信号
	C C001-C100	患者发作间歇期大脑海马结构体的 EEG 信号
患病	D D001-D100	患者发作间歇期大脑致痫区的 EEG 信号
	E E001-E100	患者发病期测得的 EEG 信号

5.2 特征提取

为了有效地识别 EEG 信号,通常需要对原始信号做一些变换来提取特征^[20]。这里部分借鉴了文献[3,8]的特征提取方法,通过离散小波变换(Discrete Wavelet Transform, DWT)和统计方法来提取特征量。采用了二阶 Daubechies 小波^[21],将长为 4097 的 EEG 信号 x 分解成了 4 层(如图 5)。首先,分解 x 得到第 1 层小波的近似系数 $cA1$ (低频分量)和细节系数 $cD1$ (高频分量),然后,进一步分解 $cA1$,得到 $cA2$ 和 $cD2$,以此类推,即得到 5 个子带,如图 6 所示。设 $cA0 = x$, 各层系数的长度为

$$L(cD_j) = L(cA_j) = \lfloor (0.5 \times (L(cA_i) - 1)) \rfloor + 2 \quad (10)$$

其中, $0 < j = i + 1, 2$ 即为阶数。

在此基础上,每个子带的最大值、最小值、均值和标准差被统计出来,得到 20 维的特征向量。实验中,这些特征被统一地归一化到区间 $[0, 1]$ 。

5.3 实验设置

从表 1,图 4 和图 6 中可以看到, E 组信号变化最为强烈, A 组信号波形与 B 组相似, C, D 组信号表现得比 A, B 组信号弱。不同的场景下,会有不同的分类目标,本文设计了 3 组实验: Expt-1 将健康

和患病的 EEG 信号分作 2 类, Expt-2 将健康、患病间隙期和患病发作期的 EEG 信号分作 3 类, Expt-3 将每一组 EEG 信号分作一类。

实验以 MATLAB 8.1 为平台,以线性和核化的 MPM,文献[4,8]中多层感知器神经网络(MLPNN),以及 MATLAB 工具箱中的 RBFNN, DT 为对比算法,与本文提出的 RB-MP-CT 进行了比较。其中 MPM(linear)和 MPM(kernel)用于多分类时采用了 OAO 策略和投票法, RBFNN 采用多输出回归进行直接多分类。各组实验都采用了 5 折交叉验证的策略进行参数选优。

实验以“均值±标准差”的形式给出各算法对数据集分类的测试精度、敏感度和特异度^[3,8],以及 MPM 相关算法对相关模型训练所得的 α 指标。

5.4 结果分析

各组实验结果如表 3~表 5,图 7 所示。从结果来看,各组实验中,本文方法都有很好的表现,在 Expt-1 和 Expt-2 中的分类精度达到 0.95 以上,在 Expt-3 中的分类精度能达到 0.81。Expt-1 和 Expt-2 中,分类任务较为简单, MLPNN 和 RBFNN 同本文方法的分类精度是相当的。但在 Expt-3 中,分类的类别数为 5,这时 RBFNN 没有找到最佳的分界面,而本文方法采用 OAO 策略则比较容易找到多个简单有效的分界面。

同 MLPNN 的结果相比,本文方法在同等条件下体现出了更好的有效性和可靠性。对 AB-CDE 的分类,本文方法实验所得分类精度为 0.9680,略高于 MLPNN。文献[8]中 MLPNN 对 AB-CDE 的分类精度高达 98.8%,文献[4]同样使用了 MLPNN 对 AB-CDE 的分类精度则为 93.2%,主要因为特征提取和实验方案存在不同。对 AB-CD-E 的分类,本文方法实验所得分类精度与文献[8]中取得的 95.6% 的分类精度是相近的,而在敏感性上好于文献[8]中的结果。在分 5 类时,本文方法取得的结果略好于 MLPNN,明显好于 DT 等。

由于决策阶段采用了分类树结构,所以本文方法显得简洁高效,图 7 显示只需 4 次比较即可判定类别,而投票法则需 10 次比较,并可能存在拒分现象。相比传统的 DT 算法,本文方法结合了 RBF 神经网络,因而拥有更好的非线性逼近能力和泛化性能。由于优先选用 α 指标大的分类器,最大程度降低了积累误差。

特别地,各组实验中,基于最小最大概率决策的相关算法的 α 指标充分刻画了不同类之间的可分性。从表 4 可以看到 C 类和 D 类的敏感度普遍偏小,又 $\alpha(C:D)$ 最小, $\alpha(A:B)$ 也较小,说明 C 组信号与 D 组差异性小, A 组信号与 B 组也较为相似,这

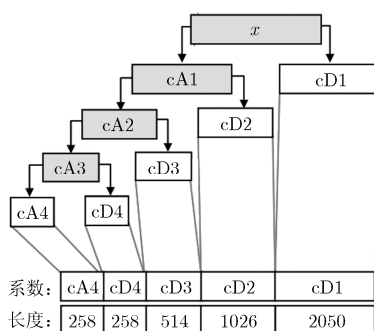


图 5 DWT 分解 EEG 信号过程

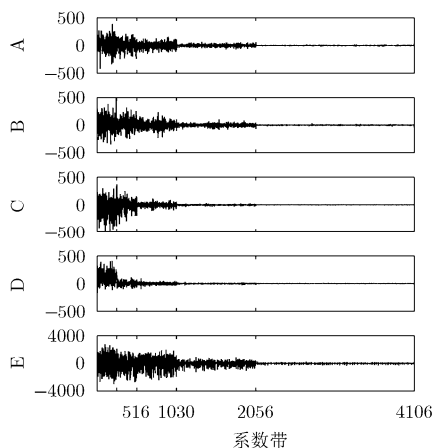


图 6 各组首个 EEG 信号的小波系数带

与图 4, 图 6 和表 2 所反映的信息是一致的。而 Expt-2 和 Expt-3 中与 E 有关的二元分类器的 α 指标都较大些, 且 E 类的特异度都普遍最大, 充分说明 E 组信号的与众不同。 α 指标代表着分类正确率的下界, Expt-1 中, MPM(linear)的 α 指标太小, 表明其分类能力有限(比照 Expt-2, 将 E 组信号与 C, D 组混合造成了 MPM(linear)的误解)。而 MPM(kernel)的 α 指标却大于分类精度, 说明它存在轻微的过拟合现象, 这一情况在文献[12]的实验研究中也讨论过。 相比而言, 本文方法所得 α 指标是正常合理的。

可以看到 α 指标的大小同各类的敏感度、特异度是存在关联的, 而敏感度和特异度与医疗诊断中的漏诊率和误诊率是密切相关的。 在使用决策树分类时, 优先选用 α 指标大的分类器的意义正在于降低积累误差。 显然 E 组信号是较易识别的(α 指标大、特异度高、误诊率低), 因而可以先将发病期的癫痫患者(E)区分开来, 而后再进一步诊断受检者是健康的(AB)还是处在发作间歇期(CD)。 综上所述, 本文方法不但拥有很强的非线性分类能力, 同时因其简洁高效, 可解释性更好。 本文方法可在训练阶段发现数据中各类数据分布关系的差异性, 建立有效的分类树模型, 而不必在测试阶段计算出敏感度和特异度后再行比较分析, 因而本文方法特别适用于 EEG 这一类的数据, 对临床医疗诊断具有指导意义, 对其他类间可分性无明显差异的数据虽然较之已有方法优势不明显, 亦可保持其可用性。

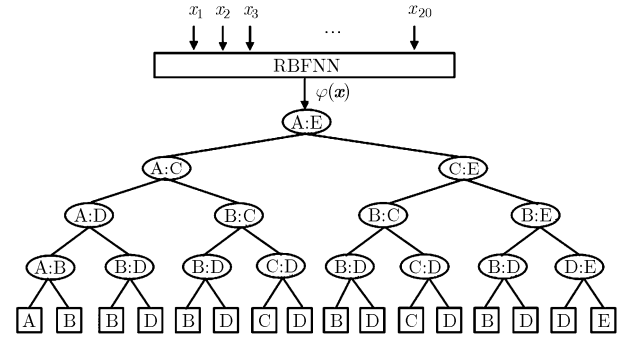


图 7 Expt-3 中 RB-MP-CT 训练所得模型

5 结束语

针对癫痫 EEG 信号识别, 结合已有研究成果, 本文提出了新的解决方案。 为充分利用不同组信号之间的可分性, 本文采用最小最大概率技术来学习 RBF 神经网络的权值参数, 进一步在 OAO 策略和排除法下, 利用学习得到的 α 指标构造可靠的分类树, 得到简洁高效的多分类模型。 实验结果证实了本文方法有很强的非线性分类能力, 同时其可解释性更好, 推理过程清晰。 然而, 由于采用了一对一策略, 类别数多的时候, 模型训练工作会比较繁重, 这是需要改进的地方。 在余下的分类器中选择子节点时, 如果出现两个分类器的 α 指标大小相等的情况, 是否有更好的处理方式, 这在未来工作中也有待深入探讨。

表 3 Expt-1 中各算法性能指标

指标	RB-MP-CT	MPM(linear)	MPM(kernel)	MLPNN	RBFNN	DT
分类精度	0.9680 ± 0.0130	0.8360 ± 0.0152	0.9520 ± 0.0084	0.9660 ± 0.0182	0.9620 ± 0.0164	0.9360 ± 0.0270
敏感度	0.9567 ± 0.0190	0.8500 ± 0.0264	0.9600 ± 0.0253	0.9600 ± 0.0303	0.9733 ± 0.0091	0.9500 ± 0.0312
特异度	0.8850 ± 0.0224	0.8150 ± 0.0379	0.9400 ± 0.0385	0.9750 ± 0.0250	0.9450 ± 0.0371	0.9150 ± 0.0418
$\alpha(AB : CDE)$	0.8222 ± 0.0084	0.4138 ± 0.0076	0.9561 ± 0.0014	-	-	-

表 4 Expt-2 中各算法性能指标

指标	RB-MP-CT	MPM(linear)	MPM(kernel)	MLPNN	RBFNN	DT
分类精度	0.9540 ± 0.0089	0.9280 ± 0.0286	0.9420 ± 0.0130	0.9560 ± 0.0251	0.9500 ± 0.0141	0.9180 ± 0.0356
敏感度	0.9700 ± 0.0119	0.9500 ± 0.0468	0.9750 ± 0.0306	0.9750 ± 0.0306	0.9600 ± 0.0285	0.9200 ± 0.0758
特异度	0.9350 ± 0.0285	0.8800 ± 0.0647	0.9250 ± 0.0250	0.9600 ± 0.0224	0.9550 ± 0.0326	0.9000 ± 0.0306
	0.9600 ± 0.0224	0.9800 ± 0.0274	0.9100 ± 0.0652	0.9100 ± 0.0894	0.9200 ± 0.0758	0.9500 ± 0.0366
	0.9700 ± 0.0139	0.9567 ± 0.0325	0.9500 ± 0.0204	0.9767 ± 0.0149	0.9733 ± 0.0091	0.9533 ± 0.0183
$\alpha(AB : CD)$	0.9700 ± 0.0274	0.9833 ± 0.0167	0.9633 ± 0.0274	0.9600 ± 0.0303	0.9533 ± 0.0183	0.9400 ± 0.0673
	0.9875 ± 0.0088	0.9550 ± 0.0244	0.9925 ± 0.0068	0.9925 ± 0.0068	0.9925 ± 0.0068	0.9775 ± 0.0105
	0.8034 ± 0.0083	0.7027 ± 0.0072	0.7729 ± 0.0123	-	-	-
$\alpha(AB : E)$	0.9601 ± 0.0049	0.7926 ± 0.0123	0.9731 ± 0.0022	-	-	-
$\alpha(CD : E)$	0.9167 ± 0.0079	0.7530 ± 0.0159	0.9482 ± 0.0087	-	-	-

表 5 Expt-3 中各算法性能指标

指标	RB-MP-CT	MPM(linear)	MPM(kernel)	MLPNN	RBFNN	DT
分类精度	0.8120 ± 0.0370	0.7680 ± 0.0356	0.7900 ± 0.0515	0.8080 ± 0.0449	0.7820 ± 0.0554	0.7240 ± 0.0744
敏感度	0.8800 ± 0.0570	0.7900 ± 0.0418	0.8000 ± 0.0612	0.9300 ± 0.0274	0.9000 ± 0.0612	0.8100 ± 0.1140
	0.8800 ± 0.1151	0.9100 ± 0.0418	0.8900 ± 0.1084	0.8400 ± 0.0652	0.8200 ± 0.0570	0.8100 ± 0.0822
	0.6400 ± 0.0418	0.5700 ± 0.1204	0.6500 ± 0.1275	0.6500 ± 0.1369	0.7100 ± 0.1851	0.6000 ± 0.0935
	0.7000 ± 0.0500	0.5900 ± 0.1475	0.6600 ± 0.0652	0.6800 ± 0.1525	0.6100 ± 0.1140	0.5200 ± 0.1304
	0.9600 ± 0.0224	0.9800 ± 0.0274	0.9500 ± 0.0354	0.9400 ± 0.0652	0.8700 ± 0.0570	0.8800 ± 0.0758
特异度	0.9625 ± 0.0125	0.9650 ± 0.0205	0.9400 ± 0.0240	0.9425 ± 0.0288	0.9450 ± 0.0112	0.9525 ± 0.0205
	0.9750 ± 0.0177	0.9575 ± 0.0227	0.9625 ± 0.0153	0.9775 ± 0.0105	0.9825 ± 0.0143	0.9650 ± 0.0347
	0.9350 ± 0.0240	0.9225 ± 0.0185	0.9325 ± 0.0274	0.9250 ± 0.0385	0.8850 ± 0.0445	0.8950 ± 0.0473
	0.9025 ± 0.0205	0.9000 ± 0.0395	0.9100 ± 0.0445	0.9225 ± 0.0311	0.9175 ± 0.0447	0.8625 ± 0.0385
	0.9900 ± 0.0056	0.9650 ± 0.0311	0.9925 ± 0.0065	0.9925 ± 0.0112	0.9975 ± 0.0056	0.9800 ± 0.0112
$\alpha(A:B)$	0.7515 ± 0.0147	0.6134 ± 0.0215	0.6912 ± 0.0210	-	-	-
$\alpha(A:C)$	0.8549 ± 0.0066	0.7646 ± 0.0111	0.7582 ± 0.0062	-	-	-
$\alpha(B:C)$	0.9192 ± 0.0054	0.8387 ± 0.0096	0.8953 ± 0.0069	-	-	-
$\alpha(A:D)$	0.8197 ± 0.0156	0.7535 ± 0.0100	0.7665 ± 0.0182	-	-	-
$\alpha(B:D)$	0.8947 ± 0.0070	0.8046 ± 0.0059	0.8865 ± 0.0074	-	-	-
$\alpha(C:D)$	0.3882 ± 0.0298	0.2725 ± 0.0327	0.3529 ± 0.0351	-	-	-
$\alpha(A:E)$	0.9777 ± 0.0030	0.8050 ± 0.0153	0.9818 ± 0.0024	-	-	-
$\alpha(B:E)$	0.9531 ± 0.0043	0.7912 ± 0.0112	0.9710 ± 0.0040	-	-	-
$\alpha(C:E)$	0.9563 ± 0.0089	0.7878 ± 0.0181	0.9698 ± 0.0048	-	-	-
$\alpha(D:E)$	0.9112 ± 0.0104	0.7461 ± 0.0146	0.9475 ± 0.0074	-	-	-

参考文献

- [1] VENEMA V, AMENT F, and SIMMER C. A stochastic iterative amplitude adjusted Fourier transform algorithm with improved accuracy[J]. *Nonlinear Processes in Geophysics*, 2006, 13(3): 321-328. doi: 10.5194/npg-13-321-2006.
- [2] POLAT K and GÜNES S. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform[J]. *Applied Mathematics & Computation*, 2007, 187(2): 1017-1026. doi: 10.1016/j.amc.2006.09.022.
- [3] INAN G and ELIF DERYA U. Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients[J]. *Journal of Neuroscience Methods*, 2005, 148(2): 113-121. doi: 10.1016/j.jneumeth.2005.04.013.
- [4] SUBASI A. EEG signal classification using wavelet feature extraction and a mixture of expert model[J]. *Expert Systems with Applications*, 2007, 32(4): 1084-1093. doi: 10.1016/j.eswa.2006.02.005.
- [5] 王登, 苗夺谦, 王睿智. 一种新的基于小波包分解的 EEG 特征抽取与识别方法研究[J]. *电子学报*, 2013, 41(1): 193-198. doi: 10.3969/j.issn.0372-2112.2013.01.33.
WANG Deng, MIAO Duoqian, and WANG Ruizhi. A new method of EEG classification with feature extraction based on wavelet packet decomposition[J]. *Acta Electronica Sinica*, 2013, 41(1): 193-198. doi: 10.3969/j.issn.0372-2112.2013.01.33.
- [6] SRINIVASAN V, ESWARAN C, and SRIRAM A N. Artificial neural network based epileptic detection using time-domain and frequency-domain features[J]. *Journal of Medical Systems*, 2005, 29(6): 647-660. doi: 10.1007/s10916-005-6133-1.
- [7] VAIRAVAN S, CHIKKANNAN E, and NATARAJAN S. Approximate entropy-based epileptic EEG detection using artificial neural networks[J]. *IEEE Transactions on Information Technology in Biomedicine*, 2007, 11(3): 288-295. doi: 10.1109/TITB.2006.884369.
- [8] ORHAN U, HEKIM M, and OZER M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model[J]. *Expert Systems with Applications*, 2011, 38(10): 13475-13481. doi: 10.1016/j.eswa.2011.04.149.
- [9] ASLAN K and HSAHIN B. A radial basis function neural network model for classification of epilepsy using EEG signals[J]. *Journal of Medical Systems*, 2008, 32(5): 403-408. doi: 10.1007/s10916-008-9145-9.

- [10] 连可, 陈世杰, 周建明, 等. 基于遗传算法的 SVM 多分类决策树优化算法研究[J]. *控制与决策*, 2009, 24(1): 7–12. doi: 10.3321/j.issn:1001-0920.2009.01.002.
LIAN Ke, CHEN Shijie, ZHOU Jianming, *et al.* Study on GA-based SVM multi-class classification decision-tree optimization algorithm[J]. *Control and Decision*, 2009, 24(1): 7–12. doi: 10.3321/j.issn:1001-0920.2009.01.002.
- [11] LANCKRIET G, GHAOUI L E, BHATTACHARYYA C, *et al.* Minimax probability machine[C]. *Advances in Neural Information Processing Systems*, Vancouver, British Columbia, Canada. 2001: 801–807.
- [12] LANCKRIET G R G, GHAOUI L E, BHATTACHARYYA C, *et al.* A robust minimax approach to classification[J]. *The Journal of Machine Learning Research*, 2002, 3(Dec): 555–582. doi: 10.1162/153244303321897726.
- [13] DENG Z, CAO L, JIANG Y, *et al.* Minimax probability TSK fuzzy system classifier: A more transparent and highly interpretable classification model[J]. *IEEE Transactions on Fuzzy Systems*, 2015, 23(4): 813–826. doi: 10.1109/TFUZZ.2014.2328014.
- [14] RUBIO-SOLIS A and PANOUTSOS G. Interval type-2 radial basis function neural network: A modeling framework[J]. *IEEE Transactions on Fuzzy Systems*, 2015, 23(2): 457–473. doi: 10.1109/TFUZZ.2014.2315656.
- [15] 陈聪, 王士同. 基于模糊分组和监督聚类的 RBF 回归性能改进[J]. *电子与信息学报*, 2009, 31(5): 1157–1160.
CHEN Cong and WANG Shitong. Improved RBF regression using fuzzy partition and supervised fuzzy clustering[J]. *Journal of Electronics & Information Technology*, 2009, 31(5): 1157–1160.
- [16] ROTH P M, HIRZER M, KÖSTINGER M, *et al.* Mahalanobis Distance Learning for Person Re-identification [M]. *London: Person Re-Identification*, 2014: 247–267. doi: 10.1007/978-1-4471-6296-4_12.
- [17] Kang S, Cho S, and Kang P. Constructing a multi-class classifier using one-against-one approach with different binary classifiers[J]. *Neurocomputing*, 2014, 149, Part B(PB): 677–682. doi: 10.1016/j.neucom.2014.08.006.
- [18] GALAR M, FERNÁNDEZ A, and BARRENECHEA E. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes[J]. *Pattern Recognition*, 2011, 44(8): 1761–1776. doi: 10.1016/j.patcog.2011.01.017.
- [19] ANDRZEJAK R G, LEHNERTZ K, MORMANN F, *et al.* Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state[J]. *Physical Review E*, 2001, 64(6): 061907. doi: 10.1103/PhysRevE.64.061907.
- [20] PARVEZ M Z and PAUL M. Epileptic seizure detection by analyzing EEG signals using different transformation techniques[J]. *Neurocomputing*, 2014, 145(18): 190–200. doi: 10.1016/j.neucom.2014.05.044.
- [21] ROY V and SHUKLA S. Automatic removal of artifacts from EEG signal based on spatially constrained ICA using daubechies wavelet[J]. *International Journal of Modern Education and Computer Science (IJMECS)*, 2014, 6(7): 31–39. doi: 10.5815/ijmeecs.2014.07.05.
- 邓赵红: 男, 1981 年生, 教授, 博士, 从事人工智能和神经模糊计算的研究.
- 陈俊勇: 男, 1990 年生, 硕士生, 研究方向为人工智能和神经模糊计算.
- 刘解放: 男, 1982 年生, 博士生, 研究方向为人工智能和模式识别的研究.
- 王士同: 男, 1964 年生, 教授, 博士生导师, 从事人工智能和模式识别.