

# 基于随机森林癫痫患者脑电数据的分析研究

盛晓欣<sup>①</sup> 田翔华\* 周毅<sup>②</sup>

**摘要** 目的: 脑电信号的研究对癫痫状态的判别具有重要的临床意义, 利用癫痫患者的脑电信号数据所做出模型来判别预测。方法: 该分类将随机抽样66.7%的脑电信号数据(4 084个样本)作为训练集。将33.3%的脑电信号数据(2 042个样本)作为测试集, 脑电信号数据输入到随机森林分类器。进行正常脑电和癫痫脑电信号的预测判别, 并证实此方法的可行性和可靠性。结果: 随机森林模型总的预测误差为2.94%。结论: 所提出的利用随机森林的方法研究癫痫患者的正常脑电和癫痫脑电信号的预测误差较低, 具有良好的参考价值。

**关键词** 脑电信号 癫痫发作 随机森林分类器

Doi:10.3969/j.issn.1673-7571.2020.01.014

[中图分类号] R319 [文献标识码] A

Analysis of EEG Pattern Recognition of Epileptic Patients Based on Random Forest / SHENG Xiao-xin, TIAN Xiang-hua, ZHOU Yi // China Digital Medicine. -2020 15(01): 41 to 43

**Abstract** Objective: The study of EEG signals has important clinical significance for the identification of status epilepticus. The prediction error made by EEG signals of epilepsy patients is used to determine the reference value of the model. Methods: This classification will randomly sample 66.7% of EEG data (4 084 samples) as a training set. 33.3% (2 042 samples) was used as a test set, and EEG signal data was sent to a random forest classifier. The prediction and discrimination of normal EEG and epileptic EEG signals were performed. And confirmed the feasibility and reliability of this method. Results: The total prediction error of the random forest model is 2.94%. Conclusion: The results show that the proposed random forest method for the study of epilepsy patients with normal brain power and epileptic EEG signals has a low prediction error, and has a good reference value in the diagnosis.

**Keywords** EEG, epileptic seizures, random forest classifier

**Fund project** Xinjiang Natural Science Fund Project (No. 2016D01C169)

**Corresponding author** College of Medical Engineering and Technology, Xinjiang Medical University, Urumqi 830011, Xinjiang Uygur Autonomous Region, P.R.C.

## 1 前言

癫痫作为一种常见的神经系统疾病, 特点是脑神经元异常放电引发反复痫性发作。由于神经元的异常放电部位和传播途径上不同, 癫痫的临床表现也不同。癫痫发作对患者构成很大的风险。癫痫发作会导致功能障碍, 如运动, 感觉, 意识和行为。也可能导致大脑细胞死亡, 影响脑功能, 甚至在严

重病例中威胁患者的生命<sup>[1]</sup>。

作为监测、诊断癫痫的重要临床手段, 脑电图在监测大脑皮层的脑活动中提供了一种更为快捷稳定的低成本和无创技术。异常脑电图波形主要见于癫痫发作患者的脑电图检查(如: 棘波、尖波、棘慢复合波等), 并通过脑电设备输出异常波形。采取人工的方式对脑电图进行判

断不仅效率低下, 也会因为专家的主观经验差异造成诊断结果的差异, 因此脑电信号的自动检测仍是当前生物医学研究的热点问题之一<sup>[2]</sup>。

德国科学家Hans Berger在人的头皮进行脑电信号采集, 对脑电信号进行分析研究取得可喜成果。脑电信号的分析处理已成为脑科学研究中心不可缺少的研究方法, 也为脑科学研究

**基金项目:** 新疆维吾尔自治区自然科学基金项目(编号: 2016D01C169)

\*通信作者: 新疆医科大学医学工程技术学院, 830011, 新疆维吾尔自治区乌鲁木齐市新医路393号

①新疆医科大学医学工程技术学院, 830011, 新疆维吾尔自治区乌鲁木齐市新医路393号

②中山大学医学院生物医学工程系, 510080, 广东省广州市海珠区新港西路135号

工作的深入开展提供技术支持<sup>[3]</sup>。学者Wilson等人用时域和频域分析方法进行人工记录并分析癫痫脑电信号,实验结果达到了92%的敏感率。Xu MG等人通过对大鼠癫痫发作过程整个脑电信号特征分析,表明可以采用非线性动力学指标描述脑电信号混沌状态的特征值。Ouyang GX等人提取癫痫脑电信号的样本熵和多尺度熵分析对区分无癫痫发作、癫痫发作之前、癫痫发作三种状态方面取得较好结果<sup>[4]</sup>。Yuan等人用近似熵提取癫痫发作间期和发作期的特征进行分析,并用终端学习机进行分类,实现了96.5%的识别率<sup>[5]</sup>。

目前,癫痫发作可以通过机器学习 and 数据挖掘方法进行分析检测。首先,从脑电图中提取有效特征信息,进行数据分析的预处理工作;其次,将预处理后的脑电信号数据发送给设定的分类器,进行癫痫发作和癫痫未发作脑电数据的分析检测。在上述实现过程中,设计有效的、能够适用于正常脑电信号和癫痫脑电信号的预测判别方法是研究的关键。

## 2 研究对象

癫痫脑电波的特点是在癫痫发生过程中,患者的脑电波活动在高波幅的情况下,频率和波形与正常情况下的脑电波活动有较大差异,其中的脑电活动存在差异性有着重要的临床学诊断意义。文中实验数据来源于新疆乌鲁木齐市新疆医科大学第一附属医院神经内科患者数据。

**2.1 属性选择** 在临床诊断学上,根据不同频率,正常自发产生的脑电信号主要分为四种节律,即 $\delta$ 节律,在0.3~3 Hz范围内; $\theta$ 节律,其处于4~7 Hz; $\alpha$ 节律,其处于8~13 Hz;

$\beta$ 节律,其处于14~30 Hz。

**2.2 数据划分** 典型的癫痫脑电图主要病态波形有棘波、尖波和棘慢复合波<sup>[6]</sup>,各种波形特征具体如下:棘波是阵发性异常脑电图的最基本形式,时间限制是20~70 ms (1/50~1/14s);棘慢复合波,出现的时限为200~500 ms (大部分为300 ms左右)<sup>[7]</sup>;尖波和棘波相同,但有时时限会比棘波较长。

## 3 研究内容与方法

众所周知,脑电信号具有随机性和非平稳性等特点,受噪声影响较大。脑电图已经作为监测、诊断癫痫的重要临床工具。各种定量定性分析和处理脑电信号均由计算机自动检测和其他方法完成<sup>[2]</sup>。

**3.1 研究内容** 研究主要侧重脑电信号的非线性动力学特征,处理癫痫患者的脑电信号数据,重点借助随机森林分类器方法进行研究分析。将已经预处理后转换的脑电信号数值数据输入到随机森林分类器,进行正常脑电和癫痫脑电信号的预测判别。

**3.2 研究方法** 当今,在各种类型的数据处理和数据挖掘过程中,快速获得有价值信息方法有很多种,机器学习方法是一种理想工具。本文中应用随机森林分类器,将原始脑电信号数据分成训练集和测试集,进而构建分类的回归树(out-of-bag, OOB)<sup>[8]</sup>,构建应用于脑电信号数据的随机森林分类模型。

## 4 分析结果

随机森林已成为比较新的机器学习模型<sup>[9]</sup>。本文数据处理所使用工具为R语言。通过脑电信号数据中m1-AV、F7-AV、……、T3-AV等共23项特征参数输入分类器,采用随机森林的核心RandomForest函数,对脑电信号进行分类预测。在进行数据分析时从脑电信号数据集中共选取6 126条脑电信号数据,从中随机抽取66.7%的脑电信号数据(4 084个样本)作为训练集,剩下33.3%(2 042个样本)作为测试集,构建的随机森林分类模型如表1所示。

随机森林中包含了500棵决策树,从随机森林分类模型混淆矩阵结果中,反馈每个决策树节点选择了4个变量;由混淆矩阵表1所示,模型总的预测误差为2.94%。

由混淆矩阵(Confusion matrix)所示,模型的预测结果同训练集实际结果之间的差别情况。从表1中可以看出,模型将数据集中类别异常(Unnormal)的2 015个训练集样本预测是正确的,将101个样本错误的预测为正常(Normal),在该类样本下预测误判率为4.78%。模型将数据集中类别正常(Normal)的1 949个训练集样本预测正确,将19个样本错误的预测为异常(Unnormal),在该类样本下预测误判率为1.0%。

在随机森林构建之后,进一步可以搜索脑电信号数据中各个自变

表1 随机森林分类模型混淆矩阵

Items	Unnormal	Normal	class.error
Unnormal	2015	101	0.047731569
Normal	19	1949	0.009654472

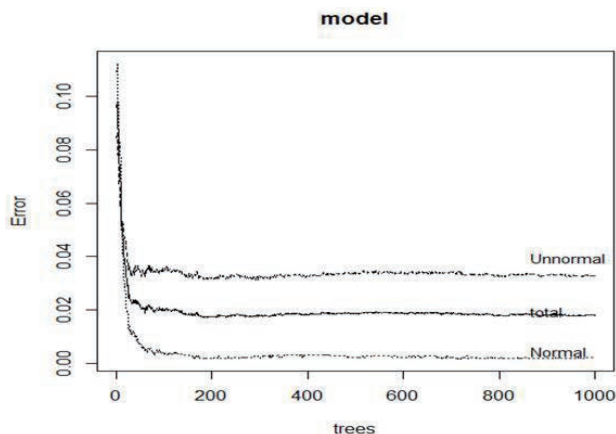


图1 模型误差与决策树数量关系图

量对模型的判别条件有很大的影响变量,对模型判别的重要性可以通过Importance函数计算出来。所有自变量在两个测量标准(标准Mean Decrease Accuracy、标准Mean Decrease Gini)下计算的相应自变量的重要性。自变量的指标值较高表明该变量对模型的判别条件有较大的影响。通过随机森林模型判别的重要性可以计算得出该样本脑电信号数据集中变量Cz对模型的判别条件有很大的影响。

通过构建的模型,如图1可以看出,模型基于OOB数据的总体误判率为2.94%。异常(Unnormal)数据的误判率为4.78%,正常(Normal)数据误判率为1.0%。

## 5 讨论

基于癫痫患者的脑电信号数据利用随机森林分类器进行建模,研究发现模型总的预测误差为2.94%,异常样本估计误判率为4.78%,正常样本误判率为1.0%。该方法在正常脑电和癫痫脑电信号的癫痫患者研究中具有较低的预测误差,分析预测癫痫发作为其提供合适的科学参考。

脑电图所对应的脑电信息数据对

癫痫患者的识别和预测具有重要的临床意义,不仅可以大大减少医务人员的工作量,而且通过数据分析得到的癫痫脑电结果也更加客观。基于癫痫患者的脑电信号分析有望在癫痫治疗领域带来新的突破,具有重要的临床价值和应用意义。本文利用随机森林分类器方法有效降低了预测误差,但仍存在一些问题需要解决,需要继续深入的开展相关工作。

随机森林所构建的模型在计算量方面没有明显改善,但它们的预测精度有所提高。可以很好地预测多达几千个解释变量的作用。但是随机森林所构建的模型对多元共线性不敏感,执行速度虽然比其他Boosting等快,但比单个的决策树慢<sup>[8]</sup>。

本文主要是针对正常脑电和癫痫发作脑电进行分析研究,对于癫痫发作间期没有什么显著研究成果,进一步的研究应围绕癫痫发作间期展开相关工作,使医疗工作者在未来的工作中通过癫痫发作间期的脑电信号能够做到预警,减少对病人的伤害,得到及时、有效的治疗。

## 6 结语

随着研究的深入,人们对算法的

要求也不再仅集中于识别率方面,还包括分类效率和分类多样性。因此,大多数现有算法不符合实际的应用要求。通过相关工作的研究,有望为癫痫的发病机制和病理学研究提供依据。

## 参考文献

- [1] 祁玉.基于脑电的癫痫预警及预警—抑制诊疗系统关键技术研究[D].杭州:浙江大学,2015.
- [2] 胡灵芝.数据挖掘方法及其在医学领域中的应用[J].辽宁中医药大学学报,2010(7):51-52.
- [3] 陈颖萍,林家瑞.基于小波变换和独立分量分析相结合的癫痫脑电分析[J].生物医学工程研究,2006,25(1):9-13.
- [4] Ouyang GX,Li J,Liu XZ,et al.Dynamic characteristics of absence EEG recordings with multiscale permutation entropy analysis[J].Epilepsy Research,2013,104(3):246-252.
- [5] Yuan Q,Zhou W,Li S,et al.Epileptic EEG classification based on extreme learning machine and nonlinear features[J].Epilepsy Res,2011,96(1-2):29-38.
- [6] 王鹏翔,张兆基.小波变换下的脑电信号癫痫特征波识别算法研究[J].信息与电脑(理论版),2017(17):63-65.
- [7] 赵建林,周卫东,刘凯,等.基于SVM和小波分析的脑电信号分类方法[J].计算机应用与软件,2011,28(5):114-116.
- [8] 曹正凤,谢邦昌,纪宏.一种随机森林的混合算法[J].统计与决策,2014(4):7-9.
- [9] 姚登举,杨静,詹晓娟.基于随机森林的特征选择算法[J].吉林大学学报(工),2014,44(1):137-141.

【收稿日期:2018-08-07】

【修回日期:2018-10-16】

(责任编辑:肖婧婧)