

基于机器学习的癫痫脑电数据分析方法研究^{*}

周 霖

韦晓燕

(中山大学数据科学与计算机学院 广州 510006) (中山大学中山医学院生物医学工程系 广州 510080)

陈子怡

周 毅

(中山大学附属第一医院 广州 510080) (中山大学中山医学院生物医学工程系 广州 510080)

〔摘要〕 详细阐述基于机器学习的癫痫脑电数据处理和分析方法, 包括监督学习、无监督学习、半监督学习、强化学习和其他类型的机器学习方法, 评价这些方法应用于癫痫脑电数据检测的效果。

〔关键词〕 癫痫; 脑电数据; 机器学习; 信号分析

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.02.013

Study on Epileptic EEG Data Analysis Methods Based on Machine Learning ZHOU Lin, School of Computer and Data Science, Sun Yat-sen University, Guangzhou 510006, China; WEI Xiao-yan, School of Biomedical Engineering, Sun Yat-sen University, Guangzhou 510006, China; CHEN Zi-yi, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China; ZHOU Yi, School of Biomedical Engineering, Sun Yat-sen University, Guangzhou 510080, China

〔Abstract〕 The paper dilates upon epileptic EEG data treatment and analysis methods based on machine learning, including supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning and machine learning methods of other types, and evaluates the application effects of the methods on inspection of epileptic EEG data.

〔Keywords〕 Epilepsy; EEG Data; Machine learning; Signal analysis

〔修回日期〕 2018-01-08

〔作者简介〕 周霖, 硕士研究生, 发表论文 1 篇; 通讯作者: 周毅, 副教授, 发表论文 110 余篇。

〔基金项目〕 国家精准医学重点研发项目 (项目编号: 2016YFC0901602); NSFC-广东大数据科学中心联合基金项目 (项目编号: U1611261); 广东省前沿与关键技术创新专项 (项目编号: 2014B010118003、2015B010106008); 广州市产学研协同创新重大专项 (项目编号: 201604016136)。

1 引言

癫痫是一种常见的慢性脑部疾病, 在全球范围内约有 5 000 万人患有癫痫, 其中 80% 人群出现在发展中国家。每年全球约有 240 万新病例发生。癫痫患者过早死亡几率是常人的 2~3 倍^[1], 给患者工作、生活和家庭带来很大的负担。因此癫痫的研究在生物医学领域一直占有重要地位。癫痫发作特

征为大脑中突发性的异常放电，其具体表现为一段时间内反复出现的发作性脑功能障碍，而发作时间和次数不定，其临床表现大多不易于察觉。通过脑电图来检测癫痫是一个可行的手段。通常癫痫脑电的检测和分析是由经验丰富的神经科医生或神经电生理学专家人工进行。人工成本高昂，效率较低，难以有效干预或控制癫痫疾病，且人工分析易出现误诊和漏诊。因此需要自动的方法对癫痫发作进行检测和衡量，以减少癫痫发作带来的严重后果。基于机器学习的方法能够快速、自动、准确地对癫痫进行检测和测量。

2 机器学习方法

2.1 概述

机器学习涉及众多学科领域，借助学习算法能够从输入的经验数据中获取潜在工作机制的模式或特征。机器学习既可用于开发临床应用的计算机辅助诊断工具，又能探索疾病的病理生理相关机制。癫痫脑电数据的分析过程包括数据采集、预处理、特征提取选择和模型构建等环节，机器学习法均可应用于这些环节。根据学习环境的不同，可将机器学习的算法分为监督学习、无监督学习、半监督学习、强化学习和其他，机器学习分类结构，见图 1。

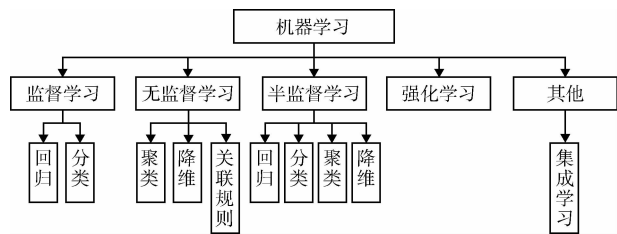


图 1 机器学习分类结构

2.2 监督学习方法

2.2.1 概述 监督学习的理念是建立在监督器的基础上，指导学习系统将未标记的数据与使用训练集类标签相关联。换言之，监督学习提供一组输入和输出变量之间的映射，随后可用于预测未知的数据。在监督式学习下，每组训练数据都有 1 个标识值或结果值。对于监督学习主要有两类问题：回

归和分类。本研究主要介绍分类问题。监督学习中典型的算法种类，见图 2。对分类问题常用的算法主要有 K 近邻（K Nearest Neighbors, KNN）、决策树、朴素贝叶斯和支持向量机等。

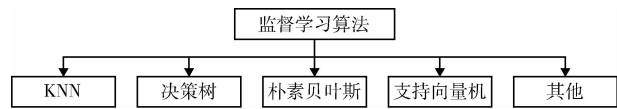


图 2 监督学习中典型的算法种类

2.2.2 KNN 算法 是一种简单而有效的非参数算法。在数据处理方面，Ghaderyan 等使用 KNN 选择癫痫脑电发作样本的子集进行训练，以解决不平衡的数据集的学习问题^[2]。KNN 不仅可以应用于癫痫发作脑电分类和预测，还可以用于算法的优化，Behnam 等利用 KNN 和散射搜索的混合方法作为癫痫发作检测的优化算法^[3]。

2.2.3 决策树 是一种数学分类规则，简单高效且具有很好的解释性。Chaibi 等提出一种基于决策树分析的高频振荡检测方法应用于痫区检测，设计 6 条过滤规则的决策树过滤出不同波段的信号，决策树方法具有更低的错误检测率^[4]。

2.2.4 朴素贝叶斯 贝叶斯分类的理论基础是贝叶斯定理，朴素贝叶斯是贝叶斯分类中最为常见的一种，它还基于特征条件独立假设。Yuan 等使用贝叶斯线性判别分类器对癫痫脑电发作时期进行分类^[5]。

2.2.5 支持向量机（Support Vector Machine, SVM）是目前应用最广泛的机器学习技术，它以统计学习理论为基础，利用不同的核函数将非线性可分样本映射到另一个更高维的空间。随着 SVM 使用日益广泛，基于内核方法的 SVM 受到更多人的关注。在分类方面 Altaf 等提出一种对数高斯基函数核的非线性 SVM 用于癫痫发作分类，在硬件上能够减少能耗和噪声^[6]。在关于支持向量机的核函数评价方面，Satapathy 等评估基于不同类型核函数的 SVM，在实证分析中表明基于径向基核综合性能最高^[7]。

2.2.6 其他算法 此外监督学习方法里还有神经网络中的学习矢量量化（Learning Vector Quantiza-

tion, LVQ) 和极限学习机 (Extreme Learning Machine, ELM), 循环神经网络 (Recurrent Neural Network, RNN) 中的 Hopfield 网络, 可用于癫痫脑电分析模型的构建。Wang 等提出非线性稀疏 ELM 用于癫痫发作检测^[8]。深度学习作为机器学习方法的一个子集, 部分深度学习算法也可用于监督学习, 如深度信念网络 (Deep Belief Network, DBN)、循环神经网络和卷积神经网络 (Convolutional Neural Network, CNN)。Movahedi 等系统地总结 DBN 在脑电图各种应用中包括癫痫检测在内的具体情况^[9]。Vidyaratne 提出一种保留细胞神经网络结构的扩展双向 RNN, 用于癫痫患者的头皮脑电特异检测^[10]。Acharya 等通过使用 CNN 分析癫痫患者脑电图信号的特征来检测癫痫发作^[11]。

2.3 无监督学习方法

2.3.1 概述 又称为知识发现, 用于将对象分成不同的簇。与分类不同的是无监督学习利用的是未标记训练数据集。该技术广泛应用于医学影像学。无监督学习可分为聚类、关联规则和降维。无监督学习算法种类, 见图 3。

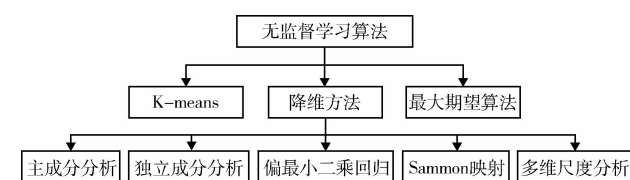


图 3 无监督学习算法种类

2.3.2 聚类算法 主要有 K-means 和最大期望算法, 而 K-means 算法是聚类算法中最简单的一种。Liu 等开发基于 K-means 聚类方法的分析工具来改进和实现癫痫中高频振荡模式的检测和分离^[12]。它常常与其他方法结合使用, Rai 等将 K-means 与经验模式分解进行结合以检测癫痫发作^[13]。最大期望算法 (Expectation Maximization, EM) 是统计模型参数估计的一种方法, 可以简单地求解最大似然估计问题。Prabhakar 等主要利用 EM 和改进的 EM 对癫痫脑电信号的风险程度进行分类和检测^[14]。

2.3.3 关联规则 主要是 Apriori 算法, 在癫痫中

的应用主要是对患者病历数据的挖掘, 曾勇介绍 Apriori 算法用于脑科电子病历病因挖掘, 获取癫痫患者的病因及其强弱次序^[15]。

2.3.4 降维方法 用于癫痫分析, 主要包括主成分分析、独立成分分析、偏最小二乘回归、Sammon 映射和多维尺度分析。机器学习中的流形学习方法也可用于维数约简, 其典型代表算法为局部线性嵌入算法^[16]。Harikumar 介绍癫痫脑电信号处理可能用到的 14 种降维技术^[17]。在降维方法中使用最广泛的是主成分分析法 (Principal Component Analysis, PCA)。Noertjahjani 等使用 PCA 提取的特征用于癫痫脑电分类和检测^[18]。

2.3.5 其他算法 自组织映射 (Self-organizing map, SOM) 是一种无监督的神经网络算法并已成功应用于各种不同的领域。Azevedo 等测试 SOM 用于癫痫脑电自动分类的可行性, 其实验结果高于其他同类型研究结果^[19]。另外深度学习中的受限玻尔兹曼机和栈式自编码网络 (Stacked Auto Encoder, SAE) 也多用于无监督学习。在癫痫预测方面, Hosseini 等利用 SAE 来预测发作癫痫脑电, 在现有的深度学习结构及 SAE 和 CNN 基础上提出一种优化的方法, 对癫痫发作能够有效预测和定位^[20]。

2.4 半监督学习方法

监督学习与无监督学习相结合的一种学习方法, 在不依赖人工的情况下, 由其算法本身自主完成对未标记数据的利用。半监督学习可分为半监督回归、分类、聚类和降维。癫痫脑电数据需要经验丰富的神经科专家完成数据的标记, 其人工成本高、效率低, 而部分脑电图是未经标记的, 通过半监督学习法可以充分地利用这部分数据。近年来半监督学习算法在癫痫脑电数据分析的方法较少, 而在其他领域的脑电数据分析中, 张娜等提出基于深层堆叠网络的半监督特征学习方法, 通过实验证明其优于使用共空间模式和 DBN 等算法^[21]。

2.5 强化学习方法

一种人工智能方法, 强调个体从与环境的相互作用中学习, 其基本概念是回报函数。强化学习使

计算机程序更倾向于做出正确的决策, 因为可以获得更多的奖励。强化学习常被用作控制技术之一, 特别针对未知环境中自主学习, 在癫痫领域应用中多用于癫痫发作控制。常见的算法主要包括 Q-learning 算法、SARSA、时间差学习和学习自动机。Nagaraj 等提出应用极随机树来优化激励策略选择的强化学习方法用于癫痫控制^[22]。

2.6 其他方法

除上述方法外, 主要还有集成学习 (Ensemble Learning) 算法。集成学习是用多个相对能力薄弱的学习模型组合成强大的学习模型。多个学习模型独立地对同样的样本进行训练, 然后将其结果整合做出整体决策。代表算法有 Boosting、Bagging、AdaBoost、随机森林、梯度推进机 (Gradient Boosting Machine, GBM) 和堆叠泛化。Hassan 等提出一种新的 LPBoost 集成方法用作癫痫发作单通道脑电信号分类模型^[23], 还利用 Bagging 方法对癫痫发作检测^[24]。Mursalin 等使用随机森林分类器检测脑电信号用于癫痫发作分析^[25]。

2.7 机器学习性能评价

大部分研究着眼于机器学习算法在癫痫脑电应用的性能评价方面。就癫痫检测而言, Fergus 等研究多种监督学习的机器学习算法使用并评价其对于检测癫痫的能力^[26]。Qazi 等使用人工神经网络和 SVM 与 K-means 算法用于癫痫发作阶段的分类, 将这些方法与传统分类方法进行性能比较, 结果表明神经网络和 SVM 都表现出良好的鲁棒性, 且神经网络方法在有噪声的数据集上表现优于 SVM 方法^[27]。Bose 等将多层感知机和随机森林用于癫痫 3 期分类对比, 实验结果表明随机森林分类效果最好^[28]。在算法提升上, Nigam 等提出将脑电信号分成 6 个频段的方法来改善机器学习对于癫痫预测的性能^[29]。综合来看, 癫痫分析的机器学习算法还无法摆脱对数据、环境的依赖, 这也阻碍其在临床上的广泛应用。多种机器学习算法相结合的方式是当前癫痫脑电研究的趋势。此外癫痫脑电数据应用机器学习方法也十分依赖于其他方法或机器学习算法

本身对数据进行特征的提取。

3 结语

本研究全面介绍近年来基于机器学习的癫痫脑电处理和分析方法, 包括监督学习、无监督学习、半监督学习、强化学习和其他类型的机器学习方法。在这些方法中监督学习方法应用于癫痫脑电数据分析仍是研究的主流, 适用无监督学习方法的研究也呈增长趋势, 对于半监督学习和强化学习而言还有较大的可研究空间。在癫痫脑电分析方面还面临着过于依赖数据和环境的挑战, 未来所应用的机器学习算法会朝着集成化、通用化的方向前进。总体来说基于机器学习方法的癫痫脑电研究凸显了其不可替代性以及对未来癫痫脑电研究的巨大潜力。

参考文献

- 1 World Health Organization. Epilepsy [EB/OL]. [2017 - 11 - 15]. http://www.who.int/mental_health/neurology/epilepsy/en/index.html.
- 2 Ghaderyan, Peyvand, Ataollah Abbasi, et al. An Efficient Seizure Prediction Method Using KNN - based Undersampling and Linear Frequency Measures [J]. Journal of Neuroscience Methods, 2014, 232 (2014): 134 - 142.
- 3 Behnam M, Pourghassem H. Singular Lorenz Measures Method for Seizure Detection Using KNN - Scatter Search Optimization Algorithm [C]. Tehran: Signal Processing and Intelligent Systems Conference, IEEE, 2016: 67 - 72.
- 4 Chaibi S, Lajnef T, Samet M, et al. Detection of High Frequency Oscillations (HFOs) in the 80 - 500 Hz Range in Epilepsy Recordings Using Decision Tree Analysis [C]. Kuala Lumpur: Image Processing, Applications and Systems Conference, IEEE, 2015: 1 - 6.
- 5 Yuan S, Zhou W, Chen L. Epileptic Seizure Prediction Using Diffusion Distance and Bayesian Linear Discriminate Analysis on Intracranial EEG [J]. International Journal of Neural Systems, 2018, 28 (1): 1750043.
- 6 Altaf, Muhammad Awais Bin, Jerald Yoo. A 1.83 μ J/Classification, 8 - Channel, Patient - Specific Epileptic Seizure Classification SoC Using a Non - Linear Support Vector Machine [J]. IEEE Transactions on Biomedical Circuits and Systems, 2016, 10 (1): 49 - 60.
- 7 Satapathy, Sandeep Kumar, Alok Kumar Jagadev, et al. An Empirical Analysis of Different Machine Learning Tech-

- niques for Classification of EEG Signal to Detect Epileptic Seizure [J]. International Journal of Applied Engineering Research, 2016, 11 (1): 120 – 129.
- 8 Wang Y, Li Z, Feng L, et al. Automatic Detection of Epilepsy and Seizure Using Multiclass Sparse Extreme Learning Machine Classification [J]. Computational and Mathematical Methods in Medicine, 2017, 2017 (1): 1 – 10.
- 9 Movahedi F, Coyle J L, Sejdic E. Deep Belief Networks for Electroencephalography: a review of recent contributions and future outlooks [J]. IEEE Journal of Biomedical & Health Informatics, 2017, (99): 1.
- 10 Vidyaratne L, Glandon A, Alam M, et al. Deep Recurrent Neural Network for Seizure Detection [C]. Vancouver: International Joint Conference on Neural Networks, IEEE, 2016: 1202 – 1207.
- 11 Acharya U R, Oh S L, Hagiwara Y, et al. Deep Convolutional Neural Network for the Automated Detection and Diagnosis of Seizure Using EEG Signals [EB/OL]. [2017 – 11 – 05]. <http://www.sciencedirect.com/science/journal/aip/00104525>.
- 12 Liu S, Ince N F, Sabanci A, et al. Detection of High Frequency Oscillations in Epilepsy With k – means Clustering Method [C]. Montpellier: International IEEE/embs Conference on Neural Engineering, IEEE, 2015.
- 13 Rai K, Bajaj V, Kumar A. Novel Feature for Identification of Focal EEG Signals with K – Means and Fuzzy C – means Algorithms [C]. Singapore: IEEE International Conference on Digital Signal Processing, 2015: 412 – 416.
- 14 Prabhakar S K, Rajaguru H. EM Based Non – linear Regression and Singular Value Decomposition for Epilepsy Classification [C]. Bangkok: ICT International Student Project Conference, 2017: 1 – 4.
- 15 曾勇. 关联规则在脑科电子病历挖掘中的应用 [J]. 医学信息学杂志, 2014, 35 (10): 55 – 58.
- 16 罗新勇. 基于流形学习的脑电特征提取方法及应用 [D]. 北京: 北京工业大学, 2016.
- 17 Harikumar R, Kumar P S. Dimensionality Reduction Techniques for Processing Epileptic Encephalographic Signals [J]. Biomedical & Pharmacology Journal, 2015, 8 (1): 103 – 106.
- 18 Noertjahjani S, Hidayat R, Susanto A, et al. An Epileptic Attack Detection Based on the Principle Components Analysis (PCA) [C]. Irbid: International Conference on Information & Communication Technology and Systems, IEEE, 2016: 105 – 108.
- 19 Azevedo C R, Boos C F, Azevedo F M D. Classification of Epileptiform Events in EEG Signals Using Neural Classifier Based on SOM [C]. Tamilnadu: International Conference on Electrical Engineering and Information Communication Technology, IEEE, 2015: 1 – 5.
- 20 Hosseini M P, Pompili D, Elisevich K, et al. Optimized Deep Learning for EEG Big Data and Seizure Prediction BCI via Internet of Things [J]. IEEE Transactions on Big Data, 2017, (99): 1.
- 21 张娜, 唐贤伦, 刘庆. 基于半监督学习的脑电信号特征提取及识别 [J]. 工程科学与技术, 2017, 49 (S2): 230 – 237.
- 22 Nagaraj, Vivek, Andrew Lamperski, et al. Seizure Control in a Computational Model Using a Reinforcement Learning Stimulation Paradigm [J]. International Journal of Neural Systems, 2017, 27 (7): 1750012.
- 23 Hassan A R, Subasi A. Automatic Identification of Epileptic Seizures from EEG Signals Using Linear Programming Boosting [J]. Computer Methods & Programs in Biomedicine, 2016, 136 (2016): 65 – 77.
- 24 Hassan A R, Siuly S, Zhang Y. Epileptic Seizure Detection in EEG Signals Using Tunable – Q Factor Wavelet Transform and Bootstrap Aggregating [J]. Computer Methods & Programs in Biomedicine, 2016, 137 (2016): 247 – 259.
- 25 Mursalin M, Zhang Y, Chen Y, et al. Automated Epileptic Seizure Selection Using Improved Correlation – based Feature Selection with Random Forest Classifier [J]. Neurocomputing, 2017, 241 (C): 204 – 214.
- 26 Fergus P, Hussain A, Hignett D, et al. A Machine Learning System for Automated Whole – brain Seizure Detection [J]. Applied Computing & Informatics, 2016, 12 (1): 70 – 89.
- 27 Qazi K I, Lam H K, Xiao B, et al. Classification of Epilepsy Using Computational Intelligence Techniques [J]. Caa Trans-actions on Intelligence Technology, 2016, 1 (2): 137 – 149.
- 28 Bose S, Rama V, Warangal N, et al. EEG Signal Analysis for Seizure Detection Using Discrete Wavelet Transform and Random Forest [C]. Paris: International Conference on Computer and Applications, 2017: 369 – 378.
- 29 Sai K A, Nigam L, Karnam D, et al. Machine Learning for Seizure Prediction: a revamped approach [C]. Kochi: International Conference on Advances in Computing, Communications and Informatics, IEEE, 2015: 1159 – 1164.