

Review of epileptic EEG feature extraction methods

Yongqiang Yu

School of Qingdao University, Qingdao, China.

Abstract: Epilepsy is a typical non-contagious chronic disease of the brain. Electroencephalogram (EEG) is an important tool for clinical epilepsy diagnosis, so the use of machine learning and deep learning models to diagnose and treat epilepsy has become a popular method, of which epilepsy EEG feature extraction is the basis for establishing a model. In this paper, several methods of epilepsy EEG feature extraction are summarized from four aspects: time domain analysis, frequency domain analysis, time frequency analysis and nonlinear dynamics, and each method is systematically summarized.

Keywords: Epileptic EEG; Feature extraction; Frequency domain; Nonlinear mechanics.

1. Introduction

EEG signaling is a reflection of the overall effect of electrical activity in nerve cells in the brain, and is a kind of physiological signal. EEG is a time-varying, non-stationary signal with a very weak amplitude, about 50uV. Epilepsy is a non-contagious chronic disease of the brain. According to the World Health Organization (data as of February 9, 2022), epilepsy affects approximately 65 million people worldwide, as many as 10% of people experience at least one seizure in their lifetime, and at a given time, the proportion of the general population with active epilepsy (i.e., persistent seizures or need treatment) is estimated to be 4 to 10 per 1000 people.

EEG is one of the most important and accurate tests in the field of epilepsy diagnosis and treatment, which can determine whether the patient's condition is a seizure through the characteristics of the EEG, but also determine the type of seizure, judge the epilepsy syndrome, and also help to find the predisposing factors of epilepsy. When researchers conduct relevant research on EEG, they usually preprocess it first and extract its features for model training. For EEG signals, which are physiological time-domain signals, the direction of extracting their characteristics is roughly four kinds of time-domain analysis, frequency-domain analysis, time-frequency analysis, and nonlinear dynamics.

Feature extraction is part of feature engineering, and features are mined and then put into machine learning or deep learning classifiers. The classification process is shown in Figure 1.

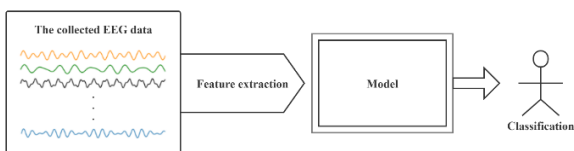


Fig. 1 Classification of epileptic electroencephalogram

2. Feature extraction method

2.1. Time-domain analysis

EEG signals are essentially a time-series signal, and time domain analysis focuses on how the EEG voltage amplitude changes over time. The method of time domain analysis mainly extracts the time domain waveform characteristics of the signal in the time domain, such as waveform analysis,

analysis of variance, amplitude peak detection, etc. The advantage of time domain analysis is that the physical meaning is clear, with strong intuitiveness, simple and fast calculation; The disadvantage is that time domain analysis is not sufficient to fully reflect the information contained in EEG signals. In 1982, Gotman calculated and extracted the time domain characteristics such as peak amplitude, mean duration, slope, and coefficient of variation of epileptic EEG, and proposed an automatic detection method for epileptic seizures based on EEG signals.

2.2. Frequency-domain analysis.

The main idea of the frequency domain analysis method is to convert the EEG signal from the time domain to the frequency domain through methods such as Fourier variation, and extract the frequency domain characteristics. After the frequency domain transformation, the medical community generally believes that there are at least 5 different brain waves in the EEG signals, which are arranged in order of frequency from low to high: Delta(δ), Theta(θ), Alpha(α), Beta(β), Gamma(γ), as shown in Table 1.

Table 1. Three Scheme comparing

Brain wave	Frequency (Hz)	Voltage(μ V)
Delta(δ)	1-4	10-20
Theta(θ)	4-8	20-150
Alpha(α)	8-12	20-100
Beta(β)	12-30	5-20
Gamma(γ)	30-50	< 2

Power spectrum analysis is a common frequency domain analysis method, the main idea is to calculate the energy of the signal in the frequency domain, and produce a power spectrum that can reflect the change of signal power with frequency in the unit frequency band.

The periodic graph method is a common way of calculating the power spectral density, and the calculation formula for the power spectral density $PSD(x)$ of the EEG signal $x(n)$ is $PSD(x) = U^2(k)/N$, and $U(k) = \sum_{n=0}^{N-1} x(n)W_N^{kn}$, N is the period of $U(k)$, and $W_N = e^{-j2\pi/N}$, $k = 0, 1, \dots, N - 1$. Wagh et al. extracted the power spectral density characteristics of epilepsy EEG and put it into the epilepsy EEG automatic classification model based on graph convolutional neural network, and achieved good results. Yunyuan Gao et al. proposed a deep learning-based method for classifying epilepsy EEG signals. This method first

converts the epilepsy EEG signal into a power spectrum density energy map (psdes), then automatically extracts features from psdes using deep convolutional neural networks and transfer learning, and finally performs epilepsy classification.

2.3. Joint time-frequency domain analysis.

EEG signals have the characteristics of non-stationary and randomness, and the Fourier transform cannot provide EEG information in both the time domain and the frequency domain, so the application of frequency domain analysis methods based on the Fourier transform is limited. To address this drawback, time-frequency analysis was proposed. The time-frequency analysis method can accurately reflect the local time-varying frequency characteristics of the signal, map the one-dimensional time-domain signal to the two-dimensional time-frequency plane, and describe the energy density and intensity of the signal at different times and frequencies by designing the combined function of time and frequency.

Common time-frequency analysis methods are short-term Fourier transform, wavelet variation, and wavelet packet transform. Unlike STFT, which matches by windowing the Fourier transform, wavelet analysis directly replaces the infinitely long trigonometric function basis of the Fourier transform with a finitely long wavelet basis function that decays with frequency, transforming the original function or signal from function space to scale and displacement space. The discrete wavelet transforms of signal A can be defined as:

$$\begin{aligned} WT_f(j, k) &= CWT(2^{-j}, k2^{-j}) = \int_{-\infty}^{\infty} f(t) * 2^{\frac{j}{2}} \psi(2^j t - k) dt \\ &= \int_{-\infty}^{\infty} f(t) * \psi_{j,k}(t) dt = \langle f(t), \psi_{j,k}(t) \rangle \end{aligned} \quad (1)$$

Based on the theory of wavelet analysis, Wavelet Packet Transform (WPT) adaptively selects the best basis function according to the characteristics of the analyzed signal and matches it to the signal to improve the analysis ability of the signal. The wavelet transform only further decomposes the low-frequency part of the signal, while the WPT decomposes both the low-frequency part of the signal and the high-frequency part of the signal. The decomposition process of a three-layer wavelet decomposition tree and a wavelet packet decomposition tree is shown in Figure 2. S is the original signal, A is the low-frequency signal, and D is the high-frequency signal, and the number of layers is numerically indicated. Acharya et al. decompose EEG into subband signals through WFT, and input the higher-order accumulation of subband signals into the model as EEG features to complete classification detection. In 2007, Abdulhamit Subasi et al. used the discrete wavelet transform to decompose the frequency subband of EEG signals. These subband frequencies are then used as inputs to the ME network.



Fig. 2 Diagram of three-level wavelet decomposition tree and wavelet packet decomposition tree

2.4. Nonlinear dynamic analysis.

Nonlinear dynamical system refers to a system in which the state and output variables of the system cannot be described by linear relations under the influence of external conditions.

In recent years, a growing body of research has shown that the brain is a complex, nonlinear chaotic system in which EEG signals can be seen as their output.

Approximate entropy is a nonlinear kinetic indicator used to measure the regularity and unpredictability of time-series fluctuations, proposed by Pincus design in 1991. It uses a non-negative number to represent the complexity of a time series, reflecting the likelihood that new information will occur in the time series, and the more complex the time series corresponds to the greater the approximate entropy. The calculation process of approximate entropy is shown in Equation (2).

$$ApEn = \Phi^m(t) - \Phi^{m+1}(t) \quad (2)$$

A time series $[X(i), i = 1, 2, \dots, n]$ take M as the window size and divide $X(i)$ into $k = n - m + 1$ sequences. $\Phi^m(t) = \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \ln C_i^m(t)$ represents the logarithmic mean of the ratio of the distance between these sequences and all k sequences greater than the threshold to the total number.

Richman and Moorman found that approximate entropy lacked relative consistency and relied heavily on record length, with lower than expected consistency for short records. Therefore, they proposed sample entropy based on approximate entropy as an improved time series complexity measurement method. Sample entropy has two advantages over approximate entropy: the calculation of sample entropy is independent of data length and has better consistency, eliminating self-matching and the algorithm is simpler, requiring only about half the time to calculate. The lower the value of the sample entropy, the higher the sequence self-similarity; The larger the value of the sample entropy, the more complex the sample sequence. Yuedong Song et al. proposed an automatic epilepsy detection model with approximate entropy and sample entropy as the characteristics of EEG signals.

3. Summary

Epilepsy EEG research has always been a hot spot and difficulty in the industry, because the EEG acquisition process mostly uses non-invasive acquisition methods, and the collection process is susceptible to interference from the external environment, resulting in a lot of noise collected EEG data, therefore, on the basis of data preprocessing, the extraction of EEG features is particularly important, only by mining more representative features, the model can produce better results. In this paper, the four dimensions of time domain analysis, frequency domain analysis, time frequency analysis and nonlinear kinetic analysis are reviewed for feature extraction, with the purpose of summarizing and prospecting the research content and possibilities of epilepsy EEG feature extraction, and providing new ideas for clinical epilepsy diagnosis and treatment.

References

- [1] Information on: {"https://www.who.int/zh/news-room/factsheets/detail/epilepsy"}.
- [2] Gotman J. Automatic recognition of epileptic seizures in the EEG[J]. Electroencephalography and clinical Neurophysiology, 1982, 54(5): 530-540.
- [3] Wagh N, Varatharajah Y. Eeg-gcnn: Augmenting electroencephalogram-based neurological disease diagnosis using a domain-guided graph convolutional neural network[C]. Machine Learning for Health, 2020: 367-378.

- [4] Gao Y, Gao B, Chen Q, et al. Deep convolutional neural network-based epileptic electroencephalogram (EEG) signal classification[J]. *Frontiers in neurology*, 2020, 11: 375.
- [5] Acharya U R, Sree S V, Suri J S. Automatic detection of epileptic EEG signals using higher order cumulant features[J]. *International journal of neural systems*, 2011, 21(05): 403-414.
- [6] 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.
- [7] Pincus S M. Approximate entropy as a measure of system complexity[J]. *Proceedings of the National Academy of Sciences*, 1991, 88(6): 2297-2301.
- [8] Richman J S, Moorman J R. Physiological time-series analysis using approximate entropy and sample entropy[J]. *American Journal of Physiology-Heart and Circulatory Physiology*, 2000.
- [9] Song Y, Crowcroft J, Zhang J. Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine[J]. *Journal of neuroscience methods*, 2012, 210(2): 132-146.