Epileptic Seizure Prediction (ESP) Using Non-Linear Dynamics And Machine Learning

Logan Fortune

February 2020

Abstract

The aim of this project is to be able to predict epileptic seizures using non-linear dynamics and machine learning tools on EEG recordings (CHB-MIT database).

1 Introduction

The term epilepsy is used to describe neurological diseases which are all based on repetitive *epileptic* seizures. Epileptic seizures are transient abnormal electrical activity of neurons. It is quite easily visible in the Figure 1 showing an EEG recording with a seizure starting at 2996 seconds and ending at 3036 seconds. Those seizures are of different impacts : screaming, falling, unconsciousness, stiffness, twitching, etc... Sadly, in France, 500000 people are subject to epilepsy. Actually, those seizures are said to be *spontaneous* and thus hard to predict. Our research aims to try to find underlying patterns on EEG recordings that could lead to a prediction model. We based our research on Bartosz Swidersk studies [1] that emphasize how the synchronization of the maximum Lyapunov exponent through many recording channels is an efficient predictor. We will also used extensively the famous '*Nonlinear Time Series Analysis*' book (abbreviated NTSA after) written by Holger Kantz.

```
Data length (number of points) : 921600
Data Sample Frequency (Hz) : 256
```

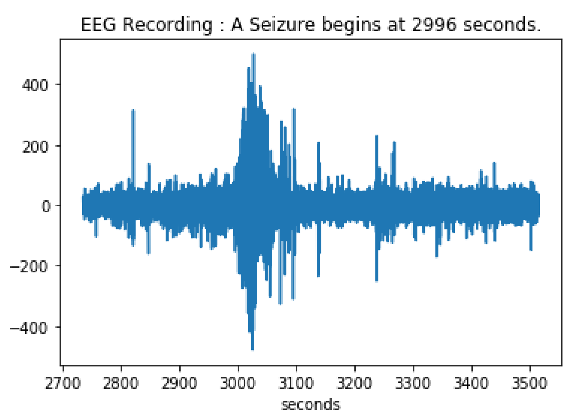

Figure 1: One EEG Recording Channel Data With An Epileptic Seizure

2 About The Database

The *'CHB-MIT Scalp EEG'* Database (`https://physionet.org/content/chbmit/1.0.0`) has been used to make some experimentation. The data consist of many recordings from 22 young people suffering from epilepsy. Each case (called chb01, chb02, etc.) contains between 9 and 42 continuous .edf files (time series signals) from a single subject. For each experimentation, signals are recorded from different parts of the brain during one hour in average. All signals were sampled at 256 samples per second with 16-bit resolution. For each experiment, the epileptic seizure starting and ending time is annotated in a separate file. The data can be read and plotted with Python tools: for instance, the Figure 1 is the representation of the signal measured between the FP1 position and the F7 position on the skull. It appears that the difference between recordings from different parts of the skull are not very important evaluating the shape of the amplitudes as you can see in Figure 3:

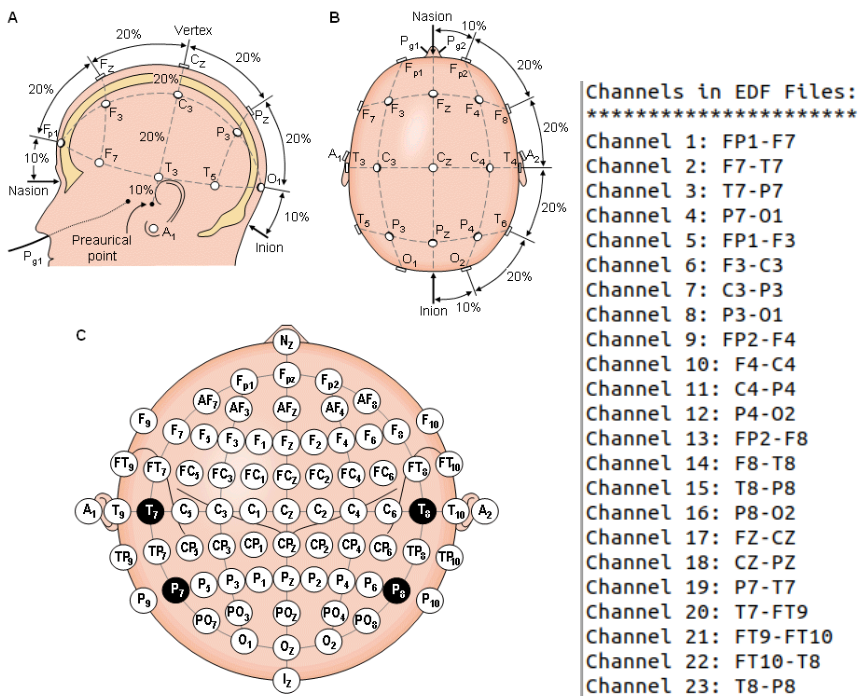

Figure 2: Possible EEG recording channels and information provided for each experimentation

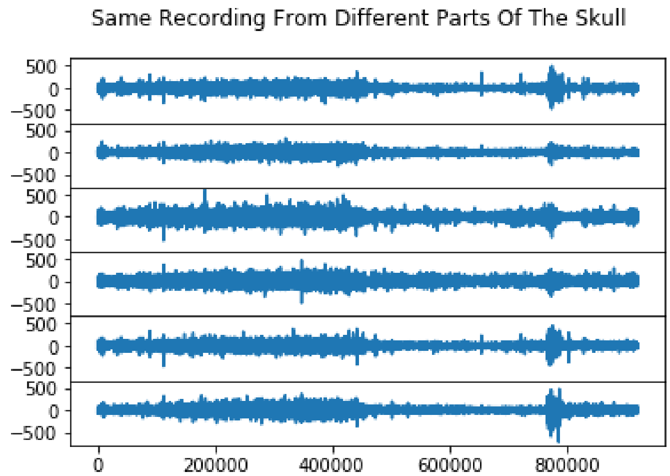

Figure 3: Data from different parts of the skull: (From Top To Bottom) FP1-F7, T7-P7, FZ-TZ, CZ-PZ, FP2-F8, T8-P8

3 Non-Linear Dynamics Analysis And Filtering

The underlying assumption of almost every non linear dynamics tools for time series analysis is that the data are stationary. It means that the statistical property of the process that generates the times series do not change over time. In a more formal way of speaking, it means that the joint probabilities of finding the system in a certain state is shift-invariance (in time) within the observation period (see Chapter 2 and 13 from the NTSA book). Unfortunately, stationary recordings from living beings are impossible because they are sensitive to the environment. Visual, sounds, skin stimuli add time dependency to the neuronal activity. Most stable statistical quantities for stationary time series are the mean and the variance. It is easy to see in Figure 1 and 3 that the variance of EEG recordings is not constant through time which is characteristic of non-stationary time series. Thus, it will be difficult to assume that the results found in this report will characterize the underlying system.

A first requirement for our time series to be considered as valid is that it must have enough data to cover the underlying mechanism over time. Indeed, it would be extremely dangerous to characterize a periodic system with only data recorded over a time under its period. The second problem is the aliasing effect of the sampling period. According to the Nyquist Sampling Rate theorem, the sampling rate must be two times the highest frequency of the input signal. If we consider obviously that neurons are the source of the signals we measure, it is hard to know exactly at which frequency they work but an estimate would be around 2 Hz (`https://aiimpacts.org/rate-of-neuron-firing/`) and 200 Hz as an upper bound. This estimation may be confirmed in Figure 4. Thus, 256 Hz as the sampling frequency respects the Nyquist Sampling Rate theorem. Because the longest relevant time scale can be estimated as the inverse of the lowest frequency which still contains a significant fraction of the total power of the signal, the time series must be longer to that period. The power spectrum is the most accurate tool to evaluate that kind of measurement: Figure 4 shows that the lowest frequency which still contains a significant fraction of the total power is 10^{-1} Hertz and so data are sufficiently sampled (one hour in average) to evaluate correctly the hidden system.

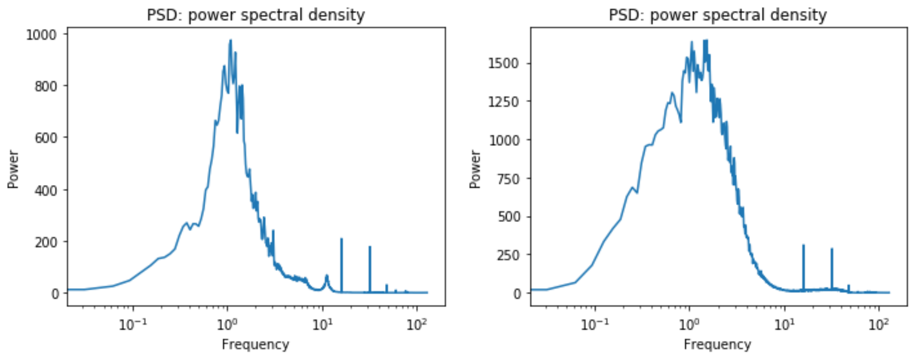

Figure 4: Example of the power spectrum of the channel 1 with an epileptic seizure (Left Figure) and without an epileptic seizure with the same patient: Welch method with the 'Hann' window of length 8192 points

One considerably problem is that our data are strongly non-stationary as we can see with the running mean and the running variance of one recording on one channel without epileptic seizure (weak stationary principle: same mean and variance through time) in Figure 5. In fact, non stationary data cannot be modeled and thus predicted. It is most of the time a real expertise to transform non stationary data into stationary ones. Our challenge lies in extracting properties of non stationary data while "even slight non-stationarity can sometimes lead to severe mis-interpretations" (Holger Kantz, NTSA book p.277).

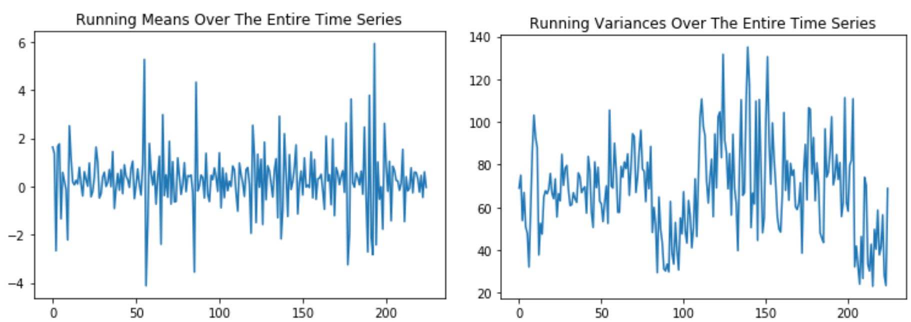

Figure 5: Example of the Running Mean and Variance of non-overlapping segments of length 4096 points (total number of points: 921600)

The power spectrum from Figure 4 is an important tool to extract dominant frequencies and harmonics. Because we are maybe studying a chaotic system, it is hard to distinguish chaoticity and noise. According to the NTSA book (p.21-22), "deterministic chaotic signals may also have sharp spectral lines but even in the absence of noise there will be a continuous part of the spectrum" (see Taylor-Couette flow experiment). When we studied the power spectrum, we lose all time information: in order to preserve this information, the spectogram performs a spectral analysis on consecutive segments of the time series (see Figure 6).

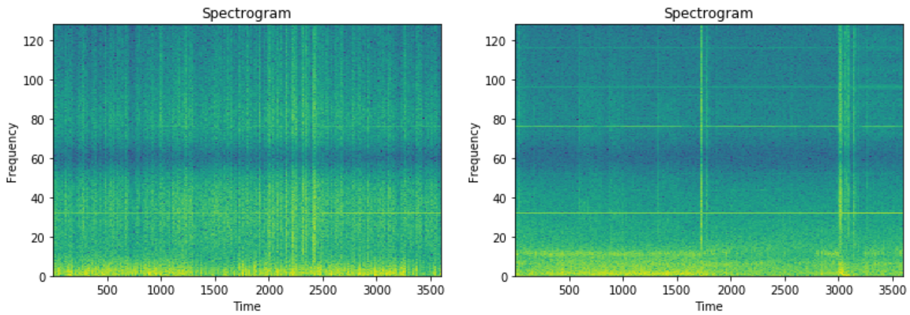

Figure 6: Example of the spectrogram of the channel 1 with an epileptic seizure (Right) and without seizure (Left)

References

[1] Bartosz Swiderski et al. "Epileptic Seizure Prediction Using Lyapunov Exponents and Support Vector Machine". In: vol. 4432. July 2007, pp. 373–381. DOI: 10.1007/978-3-540-71629-7_42.