

Project specification

Students: Gustav Röhss, Miriam Vall

Tentative project title: Classification of population activity in Parkinson's disease

Supervisor: Arvind Kumar

Introduction

Computer-aided medical diagnosis is a field of study with great potential. Using machine learning, it is possible to predict whether or not a patient has a certain disorder or disease given some information about the patient. There is reasonably potential for machine learning to be applied in the medical sciences for more than a binary diagnosis (Doi, Kunio, 2005). For example, using unsupervised learning techniques, one could possibly classify the severity of a particular condition in an already diagnosed subject.

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects movement. Its causes are genetic, although there are some environmental triggers that may increase the risk (DeMaagd, 2015).

The event that sets off the symptoms of the disease is the death of neurons that release dopamine in the basal ganglia, which is a region of the brain responsible for movement, learning and decision making. Dopamine is a neurotransmitter that sends signals to other nerve cells. When the basal ganglia becomes dopamine-deprived, the patients show symptoms that commonly include tremors (on the extremities and mostly at rest), difficulty of movement initiation or slowness (bradykinesia), and bad posture, among others (DeMaagd, 2015). At its most severe, they can completely lose the ability to move.

There is a characteristic brain signature of PD. When measuring the basal ganglia local field potential over time in patients with PD, the beta-oscillations are more prominent than the ones observed in normal brains and abnormally sustained (Mallet, 2008). The difference in beta-oscillations between a healthy brain and an affected one makes them a good biomarker of the disease, useful when making a diagnosis or administrating treatment (Hammond, 2007).

We have acquired a dataset with several different measurements of brain activity in animals diagnosed with Parkinson's disease (patients) (Cagnan et al., 2019). The data

has measurements over three different regions of the brain simultaneously. The data measures *local field potential* (LFP) and *spiking activity*. In this project we aim to study the variation in simultaneous brain activity in different regions of the brain, and attempt to classify patients into different categories.

Project statement

We have two main research questions.

1. **How different is the population activity in different regions of the brain of a patient with Parkinson's disease?**
2. **Based on available data, can we classify patients into different categories, preferably based on severity?**

Approach

One term commonly used to describe this type of research is *time series clustering* (TSC) (Liao, 2005, p1857).

There are many different ways of clustering data, such as *k-Means*, *Relocation clustering*, and *Self-organizing maps* (Liao, 2005, p1859-1861). Many of these algorithms and techniques, however, are designed to work on discrete data points (DDP), not time series of multiple, possibly thousands, data points.

One approach to counter the difficulties of TSC is to adjust an algorithm meant for clustering DDP “directly”; redesigning it to take a series of data points as a single input, while keeping the core idea of the algorithm intact (Liao, 2005, p1859-1860).

Another approach is to apply some form of preprocessing on the time series data in question, in order to transform it into DDP. This is commonly referred to as *feature extraction*. From there, clustering algorithms and techniques meant for this type of data can be applied (Liao, 2005, p1860; Wang et al., 2006, p338). This is our planned approach.

The primary measurements we will be considering are called *beta-oscillations*. These are oscillations in the LFP in the ranges of about 12-30 Hz. The reason for these being our primary interest is that these oscillations are enhanced in individuals affected by PD (Cagnan et al., 2019). These are the features we are most interested in extracting from our dataset. One way to extract features is to consider time series in the spectral domain. We are planning to initially use the *Fast Fourier Transform* (FFT) (Weisstein, n.d.) to extract spectral measurements from the dataset. We are also planning to consider measurements such as length and severity of beta-oscillations

We are initially planning to use the *k-Means* algorithm to cluster the data in accordance with our research questions, but will research and consider other methods as well.

In order to measure the quality of our results, we will use statistical analysis on the produced clusters. Some measure of *mean* and *variance* over the different features of the clusters will have to be produced, the exact nature of which are difficult to specify in advance. These measures can then be compared and analyzed in order to quantify the statistical relevance of our results. Other measures that we may or may not research and analyze include, but are not limited to, *trend*, *non-linearity* and *skewness* (Wang et al., 2006, p340).

Besides this research, we have identified a software library/package for the programming language and software environment *R* (The R Foundation), called *TSclust* (Montero, Vilar, 2014), a library specifically designed for use in TSC. The exact quality and usability of this library is something we have not yet considered in detail, but will shortly research to determine its' potential to help in this project. It is possible that *TSclust* allows us to focus on analysis rather than software implementation, which could enrich the amount and quality of data analysis produced in the scope of this project.

Time plan



References

Doi, Kunio. (2005). Current status and future potential of computer-aided diagnosis in medical imaging. *The British journal of radiology*. 78 Spec No 1. S3-S19.

Cagnan, H. et al. (2019). *Temporal evolution of beta bursts in the parkinsonian cortical and basal ganglia network*. PNAS, 116(32): 16095-16104.

DeMaags, George. (2015). Parkinson's disease and its management. *Pharmacy and Therapeutics*, 40(8): 504-510, 532.

Hammond, Constance. (2007). Pathological synchronization in Parkinson's disease: networks, models and treatments. *Itrends in Neurosciences*, Issue 7, 357-364.

Mallet, N. (2008). Parkinsonian beta oscillations in the external globus pallidus and their relationship with subthalamic nucleus activity. *Journal of Neuroscience*, 28(52): 14245-58.

Montero, P, and Vilar, J. (2014). TSclust: An R Package for Time Series Clustering. *Journal of Statistical Software*, [online], 62. Available at <https://www.jstatsoft.org/article/view/v062i01> [13-02-20]

The R Foundation. (n.d.). The R Project for Statistical Computing. Available at: <https://www.r-project.org/> [13-02-20]

Wang, X, Smith, K, and Hyndman, R. (2006). Characteristic-Based Clustering for Time Series Data. *Data Mining and Knowledge Discovery*, 13, 335-364.

Warren Liao, T. (2005). Clustering of time series data - a survey. *Pattern Recognition*, 38, 1857-1874

[Weisstein, Eric W.](#) (n.d.). "Fast Fourier Transform." From [MathWorld](#)--A Wolfram Web Resource. <http://mathworld.wolfram.com/FastFourierTransform.html> [13-02-20]