EEG Pipeline Preprocessing
Research and Suggestions of Specific Methods

Team Orange Panda

September 26, 2016

1. Electrode Interpolation

1.1 Introduction

Interpolation of bad EEG channels is important for recovering spatial resolution which would have been lost due to noise or electrode malfunction. A popular and accurate way to interpolate electrodes is to use the approximation of the scalp as a sphere, and then use well developed spherical interpolation methods from other branches of science.

The problem at hand can be thought of as follows: Given the electrode measurements (control points values) $e_1, e_2, ..., e_n$ of electrodes at spherical locations (control point locations) $(\theta_1, \phi_1), (\theta_2, \phi_2), ..., (\theta_n, \phi_n)$, what is a function f defined over the sphere where $f((\theta_i, \phi_i)) \approx e_i \ \forall i \in \{1, ..., n\}$?

1.2 Methods for Interpolation

1.2.1 Inverse Distance Weighting

A common method for interpolating points $f(\theta, \phi)$ is to take a weighted sum of n_j nearby control points (e_j), e.g.:

$$\frac{f(\theta,\phi) = \sum_{j=1}^{n_j} d((\theta,\phi),(\theta_j,\phi_j))e_j}{\sum_{j=1}^{n_j} d((\theta,\phi)}$$

where d is a distance function.

This method is non optimal because the weighting function does not depend on the spatial covariance of the data.

1.2.2 Optimal Statistical Objective Analysis

This is an improvement on Inverse Distance Weighting which takes into account spatial covariance.

This method introduces the idea of a 'first-guess field', g, which is essentially a prior belief about the value of every point on the sphere (e.g. perhaps the mean electrode value on the sphere, or the value of the closest electrode). This gives a prior estimation for each point we want to interpolate.

Let Σ be the spatial covariance matrix of the data points with the data points and $\vec{\sigma}$ be the covariances of the point we are interpolating with each data point. We then define the weight vector $\vec{w}$ as:

$$\vec{w} = \Sigma^{-1}\vec{\sigma}$$

Then we can just do a similar summation as we did in Inverse Distance Weighting, but taking the prior estimation and new weights into account, e.g.:

$$f(\theta,\phi) = \vec{w} \cdot \vec{e} + g(\theta,\phi)$$

Where $\vec{e}$ is a vector of the values of each data point, and $\vec{w}$ is the weight vector of each data point as defined above.

This method may be problematic because it has some issues with temporal non-stationarity.

1.2.3 Kringing

Kringing is a method that is similar to OSOA. It is a local interpolant which that accounts for the spatial decay of the observed variable.

Kringing uses a function called the semi-variogram, which measures how observations vary with increasing separation distance.

$$\gamma_h = (f(\theta,\phi) - f(\theta+h,\phi)^2 + f(\theta,\phi) - f(\theta,\phi+h)^2)/2n$$

γ_h is the semi-variogram of f at a separation distance of h.

With irregularly spaced data, you can fit a variogram.
Weights for the kringing method are computed in almost the exact way as in the OSOA method, except a lagrangian multiplier is introduced to ensure that the weights sum to one.

This method assumes stationarity, but there are methods such as "universal kringing" which attempt to overcome this limitation by modeling the "drift" of the field.

1.2.4 Spline Methods

A one dimensional spline of degree m fitting a set of n points $< x_1, x_2, ..., x_n >$ with values $< v_1, v_2, ..., v_n >$ is a function $s(x)$ defined piece-wise in each sub-interval $(x_i, x_i + 1)$ as an order-m polynomial. $s(x_i) = v_i \; \forall i$ and $s(x)$ is smooth up to its m-1st derivitive.

There is no elegant way to generalize this idea to two dimensions, so often a two-dimensional surface will be split up into patches on which splines are to be fit. There are many methods for carrying out this subdivision and each one will produce a different interpolation. Also, at the end, surface patches must be sewn together in a way which creates a smooth surface.

One specific method is that of Thin Plate Splines, which is computationally expensive.
Less expensive alternatives are partial spline models, and to partition the control points into subsets and find splines for each, then overlap and weave together the splines in a smooth fashion.

In the eeg_interp.m function, the method of spherical splines from Perrin Et. Al 1989 is used.

1.2.5 Spherical Harmonics

There are infinitely many smooth functions which can be defined over a sphere. A useful computational tool would be some sort of basis which allows us to represent any smooth function on a sphere. This was the idea behind the Fourier transform for functions on the unit circle.

Such a basis exists, and they are called the Spherical Harmonics.

Harmonics are functions f which satisfy Laplace's equation $\nabla^2 f = 0$. In spherical coordinates, this is:

$$\nabla^2 f = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{r^2\sin\theta}\frac{\partial^2 f}{\partial\phi^2} = 0$$

This equation can be factored into a part dependent on r, and a part dependent on θ and ϕ. If we assume that solutions to the θ, ϕ dependent part are of the form $Y = \Theta(\theta)\Phi(\phi)$, you can carry a second separation of variables, which reveals Legendre's differential equation under a change of variables $t = \cos\theta$. Thus, the solutions to Laplace's equation, the Spherical Harmonics, are of the form:

$$Y_l^m(\theta,\phi) = Ne^{im\phi}P_l^m(\cos\theta)$$

Where Y_l^m called a spherical harmonic function of degree l and order m, P_l^m is the associated Legendre polynomial, N is a normalization constant, and θ and ϕ are colatitude and longitude respectively.

Using the Spherical Harmonics as a basis, we can now fit a smooth surface to the data.

1.2.6 Tesselation

Tesselation is the process of using a set of points to divide a surface into polygons. For example, a naive division would use the polygons defined by the data points and the edges between data points.

One commonly used method is to recursively subdivide the convex hull of the data into triangles that are nearly equiangular.
Other methods use spherical triangulation.

1.3 How can we test the differences?

If we were to compare one or more of these methods, we would need a metric to determine if one method is better than another.

One method may be to create a known surface and interpolate based on a sampling of points. Then you can see how much the interpolated and true surface differ. Another method could be to use cross-validation on the actual EEG data.

1.4 Which is the best?

In a somewhat old paper (Robertson 1997) several of the above methods were tested against each-other, and the thin plate splines method had the lowest error and bias. Many new papers say that kringing has the best performance, although these 'comparison' papers (W Luo, M.C. Taylor, S.R. Parker 2008), (D. W. Wong 2004) use some specific field as a testbed. It might be a good idea to do a study similar to these on EEG data.

2. Removing Eye Artifacts

2.1 Introduction

Various regions of the brain are activated to control eye movement. Therefore, ocular activity during the progression of a trial often shows up in the EEG data as artifacts that skew the EEG data. We will focus on removing the artifacts produced by blinking and moving the eye.

2.2 Discarding Trials

Independent component analysis and regression-based techniques are effective in removing oculomotor activities from EEG data. However, there are instances when trials should be discarded even when blink and oculomotor artifacts are successfully removed. For example, though ICA removes long blinks, if a subject closes their eyes for several hundred ms, it's likely there were too tired to focus on the stimuli. Additionally, with flickering stimuli, a poorly timed blink may result in the subject missing the stimulus all together. Blink duration and proximity to the visual stimulus are important considerations.

2.3 Saccades and Microsaccades

A saccade is rapid eye movements between fixation points. A microsaccade is a small, jerk-like, involuntary eye movement. They both have a large influence on EEG data collected from frontal and lateral-frontal electrodes, especially if a nose reference was used. Spatial filtering techniques (such as the surface Laplacian) are helpful in isolating oculomotor artifacts, preventing the spread of this activity at other electrodes.

2.4 Artifact Rejection

Artifact rejection uses a threshold voltage criteria reject trials with large artifacts.

2.4.1 Advantages

1. Relatively simple to implement, and can be used before ICA is run.

2.4.2 Limitations

1. Since certain trials are rejected, useful components of EEG signals that could be used for analysis are lost.

2.5 Regression Correction

Regression correction uses distinct sets of EEG and EOG data to derive parameters characterizing the appearance and spread of EOG artifacts in EEG data. It then subtracts out the characteristics identified in the EOG out of the EEG data.

2.5.1 Advantages

1. Can effectively remove Only ocular muscle activity out of the EEG data.

2.5.2 Limitations

1. EEG and ocular activity mix bidirectionally, so regression causes relavent neural signals to also be removed from the data.

2. Requires a specific channel to use as a point of reference. Therefore, filtering out muscle artifacts, for example, would not be possible without a reference channel specific for the muscle group.

3. Regression based artifact removal of moderately and heavily contaminated trials consistently overcorrected and removed neural activity from electrodes in the frontal and periocular sites.

2.6 Independent Component Analysis (ICA)

ICA decomposes EEG time series into components that try to identify independent sources of variance. Weights are then assigned to all electrodes so each component is the weighted sum of activity at all electrodes. This process is similar to being at a cocktail party where every guest speaks into a microphone. The ICA essentially dampens the input of some microphones so that some people can be heard better than others.

2.6.1 Advantages

1. ICA separates EEG signals into independent components based on the data's characteristics. Minimal neural signal loss. It does not rely on clean reference channels to do so.

2. All data in trials are preserved when using ICA (Unlike artifact rejection method).

3. Data on all scalp channels is preserved (Unlike regression models where frontal and periocular sites are lost)

4. Uses spatial filtering from a brief segment of EEG data, giving it the flexibility to produce artifact-reduced EEG data in real-time.

2.6.2 Limitations

1. Can only break down data into the same number of components as there are electrodes. Usually, there are more sources contributing to the data than there are electrodes.

2. Assumption of temporal impedance cannot be satisfied if training set is too small, or where different topographically distinguishable phenomena always appear together in the data.

3. Assumption that physical sources of artifactual and neural activity contributing to EEG signals are spatially stationary through time.

2.7 Quantifying Success and Comparing Across Methods

The most removal of eye movement artifacts, with the least removal of valuable neural signal data. According to numerous published articles, ICA is the most effective method of filtering out eye movement artifacts in EEG data. This is because ICA identifies more than just eye movement artifacts in the data, allowing it to filter out muscle twitches, blinks, and eye movement from the EEG data while removing little important neural signal data.