A probabilistic model for segmenting EEG into microstates

Frans Zdyb
Supervisors: Lars Kai Hansen, Andreas Trier Poulsen
DTU Compute,
Technical University of Denmark
folzd@dtu.dk

Abstract

EEG exhibits states during which the spatial map of electric potential changes little, until it rapidly morphs into a different map. Called "microstates", they are thought to reflect states of mental processing. To infer microstates from noisy recordings, researchers have used modified clustering algorithms. This work presents a probabilistic model for inferring microstates using variational inference. The results are competitive with commonly used tools.

1. Introduction

By placing electrodes on the scalp, and recording the fluctuating differences in electric potential between them, one can get a measure of brain activity that is non-invasive and has a high temporal resolution. Because of these properties, EEG is used extensively in the brain and mind sciences, and in medicine.

Good explanations of brain activity must employ representations of the EEG signal that make the studied phenomenon easily quantifiable in terms of those representations.

One way of representing EEG is as a sequence of microstates[2]. The EEG recording is segmented into periods of coherent synchronized activation of large-scale neuronal networks, characterized by a unique topography of electric potentials over the entire channel array. An example of a microstate segmentation is in Figure 1.

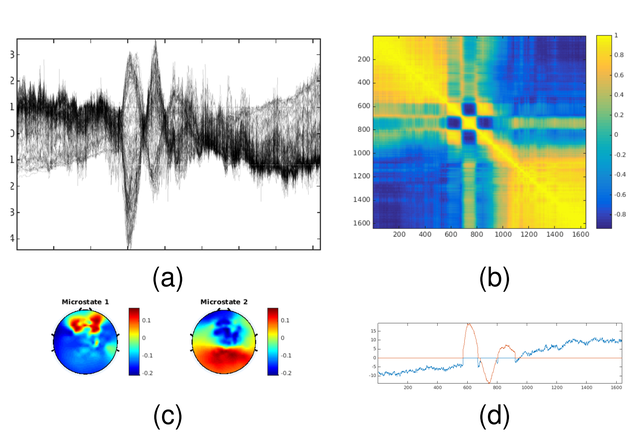

Figure 1: *(a) 0.8 seconds of EEG from 128 electrodes, sampled at 2048 Hz, averaged over 172 epochs, (b) correlation between any two scalp maps, (c) spatial components of a segmentation into two microstates, (d) temporal components of the same segmentation*

Studies have discovered significant changes in the microstate sequence in a variety of neuropsychiatric disorders and behavioral states[2]. CARTOOL, a piece of software for functional brain mapping, implements two algorithms for microstate segmentation[4]. One is a modified agglomerative clustering, the other a k-means, modified to be scale-invariant, called "N-microstates"[1]. N-microstates assumes the same model that forms the basis of this work.

2. Model

The N-microstates model expresses each time-frame in the recording as one of K spatial maps, scaled with some intensity, plus some Gaussian noise.

$$y_{jt} = \sum_k w_{jk} x_{kt} z_{kt} + \epsilon_{jt} \quad (1)$$

where

$$\begin{aligned} j \in \{1..J\} &= \text{channel}, \\ t \in \{1..T\} &= \text{time}, \\ k \in \{1..K\} &= \text{microstate}, \\ y_{jt} &= \text{voltage}, \\ w_{jk} &= \text{spatial component of microstate } k, \\ x_{kt} &= \text{temporal component of microstate } k, \\ z_{kt} &= \text{binary indicator of microstate } k, \\ \epsilon_{jt} &\sim N(0, \beta^{-1}) \\ x_{kt} &\sim N(\mu_{kt}, \alpha) \\ p_k(z_{kt} = 1) &= \frac{1}{K} \end{aligned}$$

This model is a hard clustering due to the prior structure on the indicator variables.

$$z_{kt} z_{k't} = 0, \forall k \neq k'$$

To encourage clustering subsequent frames into the same microstate, we add an assumption of temporal smoothness.

$$P(\mathbf{Z_t}|\mathbf{Z_{t-1}}) = \prod_t \prod_k \left((p_0 - \frac{1-p0}{K-1}) z_{kt} z_{k\,t-1} + \frac{1-p0}{K-1} \right)$$

3. Inference

The naive approach to inferring the model parameters is to compute and maximize the marginal log-likelihood.

$$\mathcal{L} = \ln p(\mathbf{Y}|\mathbf{W}) = \ln \int p(\mathbf{Y}|\mathbf{W}, \mathbf{X}, \mathbf{Z}) p(\mathbf{X}) p(\mathbf{Z}) \mathrm{d}\mathbf{X} \mathrm{d}\mathbf{Z}$$

where

$$p_k(z_{kt}) = \left(\frac{1}{K}\right)^{z_{kt}}, p(\mathbf{Z}) = \prod_t^T \prod_k^K p_k(z_{kt})$$

$$p(x_{kt}) = \mathcal{N}(x_{kt}|0, \alpha^2), p(\mathbf{X}) = \prod_t^T \prod_k^K p(x_{kt})$$

This is intractable. Instead, we do variational inference, inspired by [3].

1) We introduce the variational distributions that approximate the priors $p(\mathbf{X})$ and $p(\mathbf{Z})$, $q(\mathbf{X}|\boldsymbol{\mu}, \Sigma^2)$ and $q(\mathbf{Z})$,

$$q(\mathbf{X}|\boldsymbol{\mu}, \Sigma^2) = \prod_t^T \prod_k^K \mathcal{N}(x_{kt}|\mu_{kt}, \sigma_{kt}^2)$$

$$q(\mathbf{Z}) = \prod_t^T \prod_k^K m_{kt}^{z_{kt}}, \sum_k m_{kt} = 1$$

2) use Jensen's inequality to obtain a lower bound on the marginal log-likelihood,

$$\begin{aligned} F(w_{jk}, m_{kt}, \mu_{kt}, \sigma_{kt}^2, \beta) = \\ -\frac{1}{2} \sum_t \sum_k \ln 2\pi\sigma_{kt}^2 - \frac{KT}{2} + \sum_t \sum_k m_{kt} \ln m_{kt} \\ + \frac{1}{2} KT \ln 2\pi\alpha^2 + \sum_{t'}^T \sum_{k'}^K \frac{\mu_{kt}^2 + \sigma_{kt}^2}{2\alpha^2} \end{aligned}$$

$$-\gamma_1 KT - \gamma_2 \sum_{t'}^T \sum_{k'}^K m_{kt} m_{k\,t-1} + T \ln K - \frac{JT}{2} \ln \frac{\beta}{2\pi}$$

$$+ \frac{\beta}{2} \sum_{jt} \left(y_{jt}^2 - 2 y_{jt} \sum_k w_{jk} \mu_{kt} m_{kt} + \sum_k w_{jk}^2 (\mu_{kt}^2 + \sigma_{kt}^2) m_{kt} \right)$$

and 3) minimize $-F$ using block-coordinate descent.

$$\arg\min_{w_{\rho\tau}} F = \frac{\sum_t y_{jt} \mu_{kt} m_{kt}}{\sum_t (\mu_{kt}^2 + \sigma_{kt}^2)}$$

$$\arg\min_{m_{\kappa\tau}} F = \mathrm{softmax} \left(\beta \sum_j \left(y_{jt} w_{j\kappa} \mu_{\kappa\tau} - \frac{w_{j\kappa}^2 (\mu_{\kappa\tau}^2 + \sigma_{\kappa\tau}^2)}{2} \right) + \gamma_2 (m_{\kappa\,\tau-1} + m_{\kappa\,\tau+1}) \right)$$

$$\arg\min_{\mu_{\kappa\tau}} = \frac{\beta m_{\kappa\tau} \sum_j y_{j\tau} w_{j\kappa}}{\frac{1}{\alpha^2} + \beta m_{\kappa\tau} \sum_j w_{j\kappa}^2}$$

$$\arg\min_{\sigma_{\kappa\tau}^2} = \frac{1}{\frac{1}{\alpha^2} + \beta m_{\kappa\tau} \sum_j w_{j\kappa}^2}$$

$$\arg\min_{\beta} = \frac{JT}{\sum_j \sum_t y_{jt}^2 - 2 y_{jt} \sum_k w_{jk} \mu_{kt} m_{kt} + \sum_k w_{jk}^2 (\mu_{kt}^2 + \sigma_{kt}^2) m_{kt}}$$

4. Results

To compare N-microstates to the variational version of it, we simulate a sequence of randomly generated microstates, and compute the mutual information between the inferred sequence z_{kt} and the simulated one. We restart both algorithms with random initialization 40 times, to mitigate bad local minima. We repeat the procedure 50 times each for a range of signal-to-noise ratios in the range 0-10 dB.

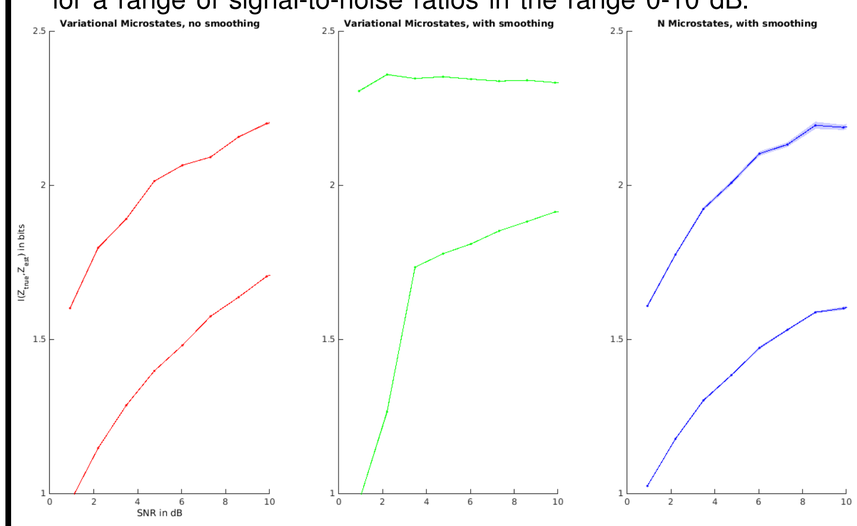

Figure 2: *Mean ± 3sd of 50 best-out-of-40-restarts mutual information between the true segmenation and segmentations estimated by Variational Microstates without smoothing (red), with smoothing (green) and N-microstates (blue), as a function of SNR. Higher curves are for $K = 7$, lower for $K = 4$. Higher is better, best possible mutual information is the entropy of the true segmentation, $H(Z_{true}) = 2.54$*

Both algorithms are implemented in an extension for the MATLAB-toolbox EEGLAB[5], at https://github.com/deoxyribose/microstates.

References

[1] Pascual-Marqui, Roberto D and Michel, Christoph M and Lehmann, Dietrich Segmentation of brain electrical activity into microstates: model estimation and validation *Biomedical Engineering, IEEE Transactions on*, volumne 42, number 7, pages 658–665, 1995

[2] Arjun Khanna and Alvaro Pascual-Leone and Christoph M. Michel and Faranak Farzan Microstates in resting-state EEG: Current status and future directions *Neuroscience & Biobehavioral Reviews*, volumne 49, number 0, pages 105–113, 2015

[3] Kappen, Hilbert J and Gómez, Vicenç The variational garrote *Machine Learning*, volumne 96, number 3, pages 269–294, 2014

[4] Denis Brunet, Micah M. Murray, and Christoph M. Michel Spatiotemporal Analysis of Multichannel EEG: CARTOOL *Computational Intelligence and Neuroscience*, Volume 2011, Article ID 813870, 15 pages doi:10.1155/2011/813870

[5] Delorme, Arnaud and Makeig, Scott EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis *Journal of neuroscience methods*, volume 134, number 1, pages 9–21, 2004