Modeling EEG micro-states using Mixtures of the Watson distribution

Modeling EEG data with Watson distributions

Martin Simon, Arturo Arranz & Manuel Montoya Cognitive Systems at DTU Compute

Danmarks Tekniske Universitet

Abstract

The aim of this project is to express the statistical distribution of electroencephalographs (EEG) data using mixture models of the Watson distribution (moW). In EEG data, the key information is believed to be scale and polarity invariant so an axially symmetric directional distribution is a good candidate for the data modeling. We use the Expectation-Maximization algorithm (EM) to compute such mixtures.

Introduction

Brain activity can be modeled by micro-states obtained from the spatial distributions of electric potential measured using EEG. This neural activity oscillates in polarity and the micro-state information is independent of the intensity of the signal.

Current state-of-the-art methods use K-means clustering as it can efficiently classify a large number of continuous numerical data of high-dimensions. A modified version of K-means (modified predictive residual variance) is used in EEGlab, a Matlab toolbox for EEG data analysis. This version generates directional clusters but it lacks a statistical framework. It can be considered a special case of the moW if we impose hard decision on the clusters and same concentration parameter (kappa).

Since EEG data is a time-series, the temporal information of the samples could contain information about the distribution. Due to this fact, 2 different models are used, EM with independency assumption between the samples, and EM with a Markov Chain of order 1 assumption between the samples, thus modelling the data as an Hidden Markov Model.

We aim to learn the number of states needed to express the data by means of crossvalidation and analyze and compare the clusters found for different EEG data.

The EEG data used to evaluate the algorithms comes from multi-subject, multi-modal human neuroimaging recordings. The volunteers performed a simple perceptual task on pictures of famous and scrambled faces during two visits to the laboratory [4]. The mean value of the time series for the 70 recorded channels can be observed in Figure 6.

Watson Distribution

The Watson distribution models directional and scale invariant data, this implies that every sample lays in the hypersphere, $\mathbb{S}^{p-1} = \{x | x \in \mathbb{R}^p, ||x||_2 = 1\}$. The distribution is governed by the *concentration parameter*, $\kappa \in \mathbb{R}$, and the *mean direction*, $\mu \in \mathbb{S}^{p-1}$.[3]

$$W_p(x|\mu,\kappa) = c_p(\kappa) e^{\kappa(\mu^T x)^2, x}, \qquad x \in \mathbb{S}^{p-1} \tag{1}$$

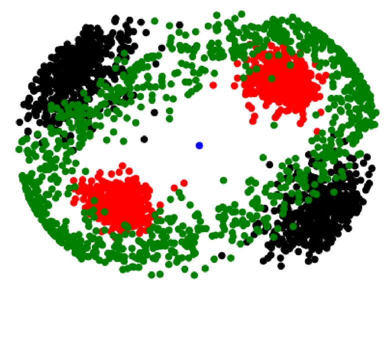

Figure 1: moW with 3 clusters

EM algorithm for independent moW

EM is an iterative algorithm to calculate mixture models. Under the the i.i.d data assumption, the expectation and maximization steps for moW are the following:

E-step:

$$\beta_{ij} = \frac{\pi_j W_p(x_i|\mu_j,\kappa_j)}{\sum_l \pi_l W_p(x_i|\mu_l,\kappa_l)} \tag{2}$$

M-step:

$$\begin{aligned} &\mu_j = s_1^j \quad if \quad \kappa_j > 0 \qquad \mu_j = s_p^j if \kappa_j < 0 \\ &\kappa_j = g^{-1}(1/2, p/2, r_j), \quad where \quad r_j = \mu_j^T S^j \mu_j \\ &\pi_j = \frac{1}{n}\sum_i \beta_{ij} \end{aligned} \tag{3}$$

where s_1 and s_p is the first and last eigenvalues of weighted scatter matrix, respectively.[1]

EM algorithm for HMM moW

In this version of moW, we assume that the data follows a Markov Chain of order 1. In this case, the probability of a sample being generated from a given cluster, depends on the previous cluster.So we have a transition probability matrix, P , and an initial probability vector π.[1]

Results with generated data

In order to assess the correctness of the algorithms we test them with the artificial generated data in Figure 1.

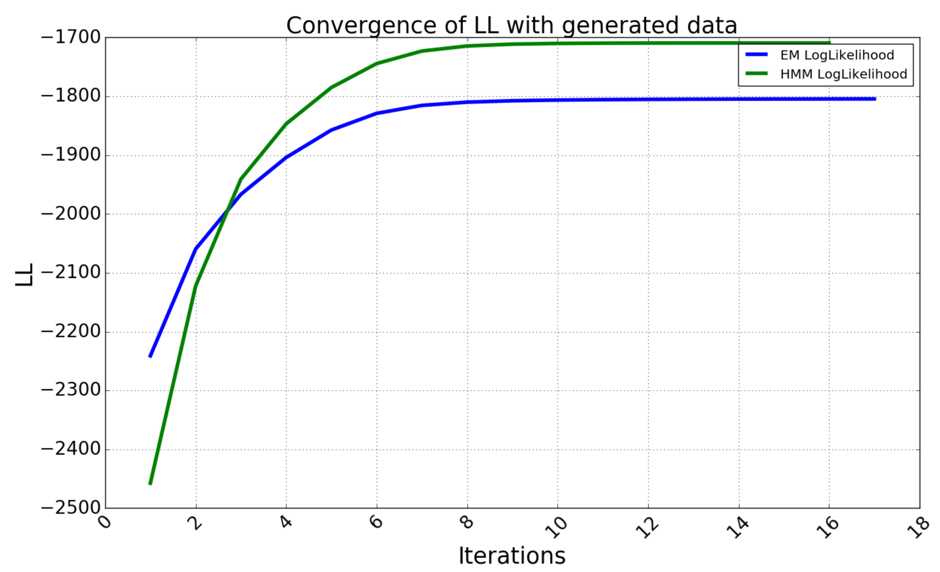

Figure 2: LL convergence for HMM and EM

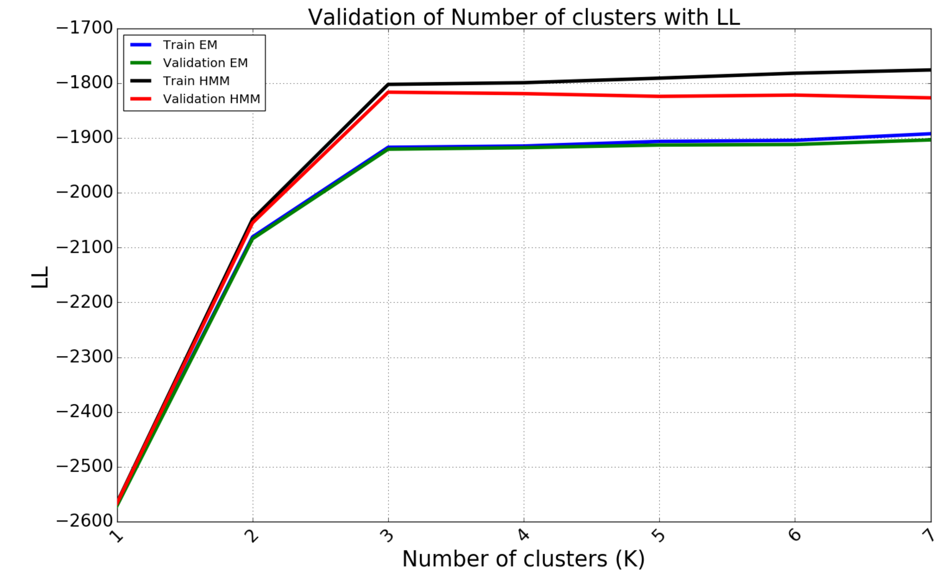

Figure 3: Clusters validation for HMM and EM

As we can observe in Figure 2, the log-likelihood is monotically increasing with the iterations and the HMM likelihood is higher. Figure 3 shows how the cross validation of the number of clusters indicates that 3 clusters is enough to express the data, given the saturation of the likelihood.

Results with EEG data

The 2-fold crossvalidation of EEG data for a single person in Figure 4 shows that 4 clusters are enough to describe the "scramble face" trials. For the "famous face" trials, the optimal number of clusters is 6.

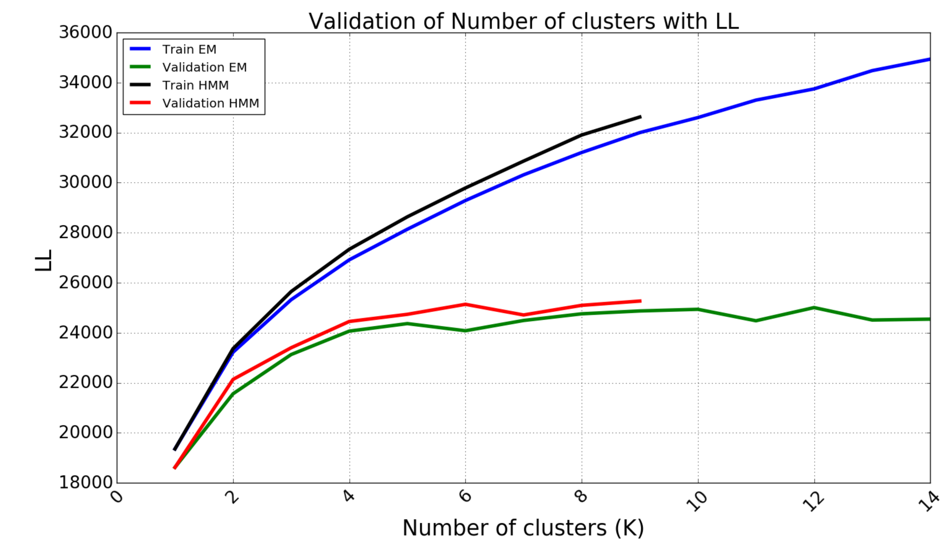

Figure 4: LL HMM-EM for class "Scramble face"

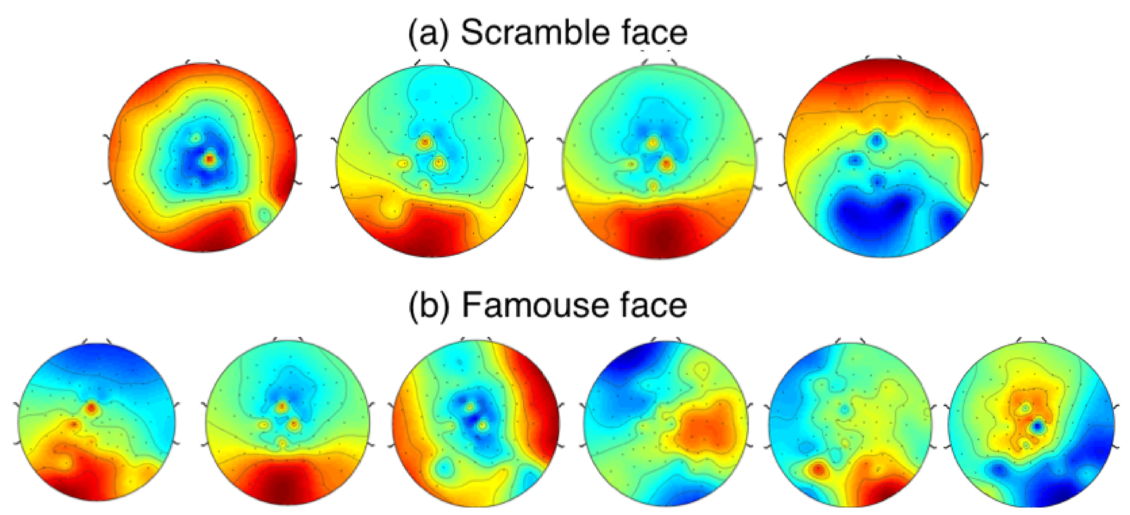

Figure 5: Skullmaps from each class

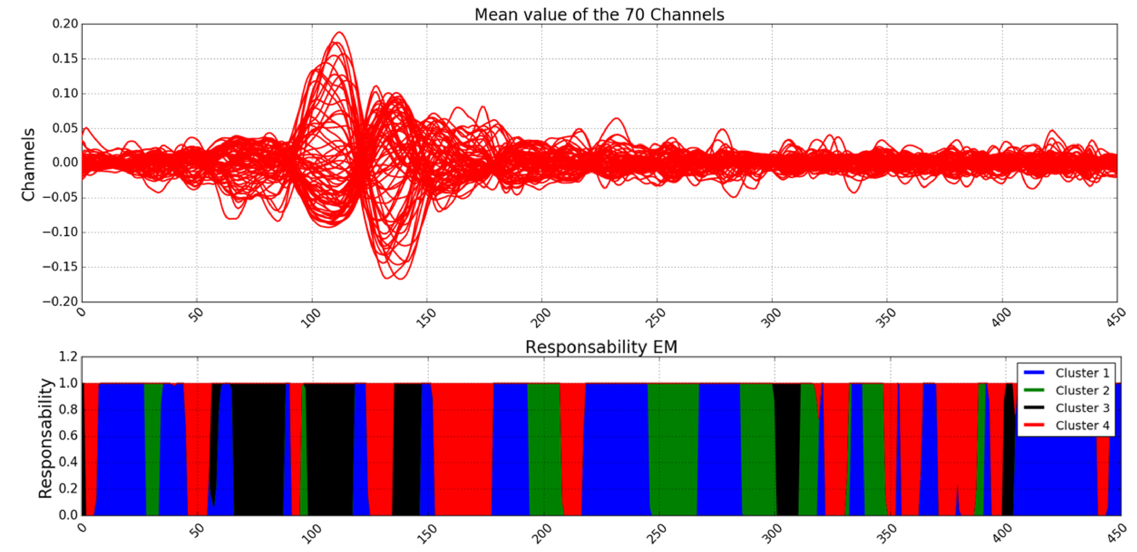

Figure 6: Temporal evolution of EEG data and EM clusters

The scalp maps for each moW are represented in Figure 5 [2]. Figure 6 shows the temporal ocurrence of the clusters for the validation trial of class "Scrambled face", as we can see, the clusters have time dependency, which can be observerd in the estimated transition matrix in the HMM, which has high values in the diagonal, meaning that once we are in a cluster (micro-state), we tend to stay in it.

Conclusions and Further work

- The number of clusters needed to express the EEG data for a single person is in the range 4-6.
- The data reflect temporal correlation indicated by the HMM transition matrix and likelihood of both methods
- The scalp map indicates high visual perception for both classes and "famous face" class has an emotional response.
- Further work would consist on using the moW for classificatio and further analysis on the meaning of the clusters.

References

[1] Christopher Bishop. *Pattern Recognition and Machine Learning*. 1st edition, 2006.

[2] Makeig S. Delorme, A. Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis.

[3] Karp Dimitrii Sra Suvrit. *The multivariate Watson distribution: Maximum-likelihood estimation and other aspects*. 2011.

[4] Henson Richard N Wakeman, Daniel G. A multi-subject, multi-modal human neuroimaging dataset.