Estimating the activity of regions of interest using Expectation Maximization algorithm

Hilkka Hännikäinen
05.12.2018

Advisor: Naryan Puthanmadam Subramarinyam
Supervisor: Prof. Harri Ehtamo
The human brain has two functional principles:
  - Functional segregation
  - Functional integration

  Quantified in terms of functional connectivity which is defined as statistical interdependencies between the functional properties of different brain regions.

  Functional properties can be measured using EEG/MEG.
Due to limited spatial resolution of EEG/MEG, the underlying neural activity is estimated by solving an inverse problem.

The activity of a region is determined by combining the activity of the neural sources within a certain region using some parcellation e.g. Desikan-Killiany.
• Functional connectivity is estimated from the activity of the regions using various methods such as correlation or Granger causality
• Currently used methods for combining the activity of neural sources within a region are quite ad hoc e.g. mean, flipped mean or PCA
• This thesis focused on developing a data driven method for estimating the activity of a region
Connectivity estimation pipeline

- MEG data
  - 100-300 sensors

- Source estimation
  - 8000-20000 vertices

- Lead field matrix

- Parcellation data
  - Desikan-Killiany atlas: 68 parcels, Destrieux atlas: 140 parcels
  - Mean, flipped mean, PCA

- Single time-course for each parcel

- Connectivity between parcels
Measuring the activity

- The activity of the brain can be measured with noninvasive measurement techniques such as EEG and MEG which measure the electromagnetic fields generated by neurons.

Approximation of a neural source

- A current dipole is an approximation for multiple closely related coherently behaving neurons
The inverse problem

- The activity of the neural sources can be estimated by solving the equation
  \[ \bar{y}_t = G \bar{q}_t + \bar{\varepsilon}_t \]
  where
  \( \bar{y}_t \) = measured data
  \( G \) = the lead field matrix
  \( \bar{q}_t \) = activity of the dipoles
  \( \bar{\varepsilon}_t \) = the measurement noise
  → The neural source activity within a region is combined to form a single time series representing the activity of the region
Objective

• Implementing a data driven method to estimate the activity of the regions
• Estimate a matrix, call it L, such that
\[ \bar{q}_t = L \bar{u}_t + \bar{\epsilon}_t \]

where
\[ \bar{q}_t = \text{the dipole activity} \]
\[ L = \text{the weight matrix} \]
\[ \bar{u}_t = \text{the region activity} \]
\[ \bar{\epsilon}_t = \text{additional dipole noise} \]
Methods 1/2

- The activity of the brain is modeled using a state-space model
  \[ \bar{u}_t = A\bar{u}_{t-1} + \bar{\vartheta}_t \]
  \[ \bar{q}_t = L\bar{u}_t + \bar{\epsilon}_t \]
  \[ \bar{y}_t = GL\bar{u}_t + \bar{\eta}_t \]
  where \( \bar{\eta}_t = G\bar{\epsilon}_t + \bar{\epsilon}_t \)

- Enables to obtain the activity of a region using Kalman filter
Methods 2/2

- Expectation Maximization algorithm is used in the estimation of $L$
- The expectation of $\bar{u}_t$ is estimated using Kalman smoother
- The maximum a posterior is obtained by

$$\hat{L}_{k+1} = \arg \max_L (Q(L, L_k) + \ln p(L))$$

where

$$Q(L, L_k) = \int \ln p(u, y|L)p(u|y, L_k)du$$

is the expectation of the joint log-likelihood under a posterior
Simulation setup

- The weights in L are designed to have a Gaussian distribution around the most active dipole
- To introduce spatial smoothness, a penalty term is added
Results 1/2

The graph illustrates the relative error (%) as a function of the parameter r (m) for different initializations. The legend indicates:

- **L** with true initialization
- **L** with random initialization
- **u** with true initialization
- **u** with random initialization

The error values are represented by data points connected by lines, with error bars indicating the variability of the data.
Real data

• Left auditory stimulus MEG-measurement
  – A stronger response in the activity representing the right hemisphere is expected
• From MNE Python library sample data set
Results 2/2

![Graph of source amplitude over time showing different estimation methods: Mean, Flipped mean, PCA, and Maximum likelihood estimation.](image-url)
Conclusions

• Simulation results show that a larger radius was needed to obtain decent results and the performance on smaller radius was not good
  – Simulation setup might not be optimal
  – Design of the penalty term
  – Large number of parameters

• Application on real data produced a higher amplitude in the activity meaning a more dipole-like activity

• Carefully designed and implemented data driven method would provide a more accurate approach for estimating the region activity
References