Recognition of asynchronous cognitive events from EEG

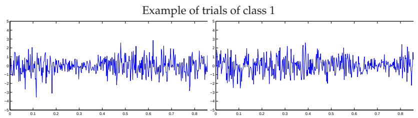
Nicolas Bourdaud, Ricardo Chavarriaga, José Millán
Defitech Foundation Chair on Non-Invasive Brain-Machine Interface (CNBI)

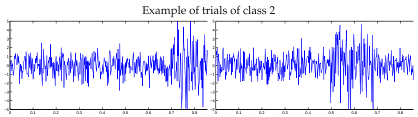
cnbi
Defitech Foundation Chair in Non-invasive Brain-machine Interface

Introduction

We propose a new approach of classification of asynchronous EEG data based on Bayesian estimation of the parameters.

Context of detection

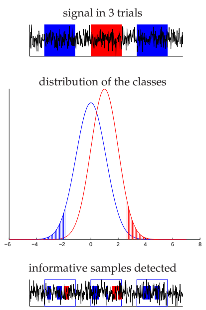




- Classification of trials
- Stationary distribution of features
- Absence of strong time-cues
- discriminant part of the signal is localized (i.e. not all part of a trial is discriminant for classification)

Previous discriminant approach

A previous approach has been developed (*Bourdaud et al.* 2009) based on the definition of two sets of discriminant samples:



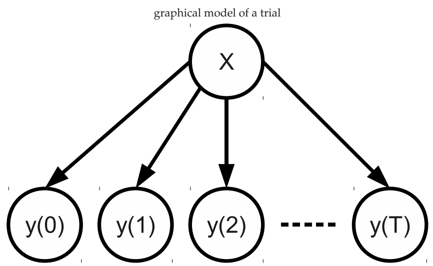
- time information discarded
- opposing tails of the distribution define two informative sets
- informative samples are detected in a new trial
- voting based classification

problems:

- Hard threshold
- Distribution of informative samples not taken into account
- Parameters estimation: heuristics based on percentile

New signal model

In a trial, samples y_i are observations of the same state (i.e. the class x)



Under the condition, we can express the posterior probability:

$$p(x|y_{1:T}) = p(x)\frac{\prod_{i=1}^{T} p(y_i|x)}{\sum_{x=0}^{1} p(x)\prod_{i=1}^{T} p(y_i|x)}$$

Samples are not all discriminant in a trial. In a trial, we assume there is a probability λ_c that a sample belongs to the "informative set" of the class. The likelihood of a sample is then modeled as a mixture of two distributions:

- a distribution of non-informative set (NIS), *same* for both classes.
- a distribution of the asynchronous informative set (AIS) specific to each class.

Then:

$$p(y|x_c) = \lambda_c p_{AIS}(y|x_c) + (1-\lambda_c)p_{NIS}(y|x_c)$$

Training procedure

A distribution model for each informative and non-informative set should be posed according to the problem we want to deal with. Once these models decided, their parameters θ_i will be estimated during the training phase.
This estimation is done using a Bayesian estimator, the *posterior expectation*. This estimates the parameters as the expectation of their posterior probabilities:

$$\hat{\theta} = \int_{\theta} \theta p(\theta|y) d\theta$$

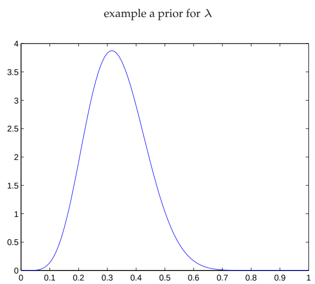
Given the complexity of this integral, we employ a Monte-Carlo Markov Chain (MCMC) algorithm which allows us to simulate the distribution $p(\theta|y)$).
We can note that:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

where $p(y|\theta)$ is given by the signal model described in the previous box and $p(\theta)$ is the prior knowledge we have about the parameters.

By designing the signal as a mixture of distributions, we have introduced a lot of freedom in the distribution of θ. This freedom is counterbalanced by the prior whose a simple and sensible choice allows us to favor discriminant solution in a elegant way. Examples of priors:

- λ_c not too high neither small
- means of the two informative set should be "distant" enough

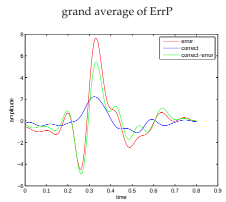


Strength of the approach

- Assuming a sensible choice of the distribution models, the output are real posterior probabilities
- The classification is not based on hard thresholds
- Less sensitive to artifacts/bad samples than discriminant approaches (the signal model embeds a way to penalize artifacts)
- Deals with non-aligned data
- Provides a way to know on which data in the trial, the decision has been based
- Can be easily extended to more classes

Application

Error potentials (ErrP) are EEG potentials elicited when a human subject is aware of an erroneous decision. Given their well-known structure, we use them to test the approach



- Features: amplitude of 7Hz component.
- Model: mixture of Gaussian distribution
- $p(\lambda)$ is a beta distribution centered on 0.3
- The difference between the mean of the 2 informative sets follows a Gaussian distribution

In spite of very few prior knowledge included:

- parameters of informative sets converges towards the peek of ErrP
- λ_{err} converges to 0.1 (length of the actual discriminant part).

This shows that this estimation procedure is able to catch the discriminant epochs of error potential.
This model will then be tested for classification purpose. We will test it on different problems like visual evoked potential or classification of exploratory behavior.

Acknowledgments

This work is supported by the European IP Project FP6-IST-027140 (BACS)