Data set IVc ‹*motor imagery, time-invariance problem*›

Data set provided by Fraunhofer FIRST, Intelligent Data Analysis Group (Klaus-Robert Müller, Benjamin Blankertz), and Campus Benjamin Franklin of the Charité - University Medicine Berlin, Department of Neurology, Neurophysics Group (Gabriel Curio)

Correspondence to Benjamin Blankertz ‹benjamin.blankertz@tu-berlin.de›

The Thrill

When taking a machine learning approach to Brain-Computer Interfacing, the user usually has to perform a calibration measurement in the beginning of a BCI experiment which provides the training data. After that the user should be able to control BCI feedback applications as long as s/he wants. With powerful algorithms the classification performance on the training data is ofter better when using complex, i.e., high-dimensional features. But these features may be affected by signal characteristics that slowly change over time, and accordingly the quality of the BCI feedback can degrade over time. When training data is only available from a short time span at the beginning of a measurement it is difficult to make the algorithm invariant to such disturbances. This data set poses the challenge of finding a classification that works on test data that was recorded several hours after the training session.

Experimental Setup

This data set was recorded from one healthy subject. He sat in a comfortable chair with arms resting on armrests. The training data set consists of the first 3 (non-feedback) sessions. (It is the same as the training data of data set IVb). Visual cues (letter presentation) indicated for 3.5 seconds which of the following 3 motor imageries the subject should perform: (L) *left* hand, (F) right *foot*, (Z) *tongue* (=Zunge in german). The presentation of target cues were intermitted by periods of random length, 1.75 to 2.25 seconds, in which the subject could relax. The test data was recorded more than 3 hours after the training data. The experimental setup was similar to the training sessions, but the motor imagery had to be performed for 1 second only, compared to 3.5 seconds in the training sessions. The intermitting periods ranged from 1.75 to 2.25 seconds as

before. The other difference was that the class *tongue* was replaced by the class *relax*. The reason for including the *relax* class into the test data without having training examples for it is the same as for data set IVb, see there.

Format of the Data

Given are continuous signals of 118 EEG channels and, for the training data, markers that indicate the time points of 210 cues and the corresponding target classes. Only cues for the classes *left* and *foot* are provided for the competition (since tongue imagery was not performed in the test sessions).

Data are provided in **Matlab** format (`*.mat`) containing variables:

- `cnt`: the continuous EEG signals, size [time x channels]. The array is stored in datatype `INT16`. To convert it to uV values, use `cnt= 0.1*double(cnt);` in Matlab.
- `mrk`: structure of target cue information with fields (the file of test data contains only the first field)
  - `pos`: vector of positions of the cue in the EEG signals given in unit *sample*, length #cues
  - `y`: vector of target classes (-1 for *left* or 1 for *foot*), length #cues
- `info`: structure providing additional information with fields
  - `name`: name of the data set,
  - `fs`: sampling rate,
  - `clab`: cell array of channel labels,
  - `xpos`: x-position of electrodes in a 2d-projection,
  - `ypos`: y-position of electrodes in a 2d-projection.

As alternative, data is also provided in zipped **ASC II** format:

- `*_cnt.txt`: the continuous EEG signals, where each row holds the values for all channels at a specific time point
- `*_mrk.txt`: target cue information, each row represents one cue where the first value defines the time point (given in unit *sample*), and the second value the target class (-1 for *left* or 1 for *foot*). The file of test data only contains time points.
- `*_nfo.txt`: contains other information as described for the matlab format.

Requirements and Evaluation

Please provide an ASC II file (named 'result_IVc.txt') containing <mark>420</mark> lines of your classifier output (real number between -1 and 1) for each trial of the test data set. The performance criterium is the mean squared error with respect to the target vector that is -1 for class *left*, 1 for *foot*, and 0 for *relax*, averaged across all trials of the test set. Note that there are no training samples for the class *relax*, see also the introductory paragraph of the [description of data set IVb](). This class must be defined by the absence of mental states *left* and *foot*. The motivation for this performance measure is that the idea is to have a system that is suitable for one dimensional cursor control, see also the [description of data set IVb]().

You also have to provide a description of the used algorithm (ASC II, HTML or PDF format) for publication at the results web page.

Technical Information

The recording was made using BrainAmp amplifiers and a 128 channel Ag/AgCl electrode cap from ECI. 118 EEG channels were measured at positions of the extended international 10/20-system. Signals were band-pass filtered between 0.05 and 200 Hz and then digitized at 1000 Hz with 16 bit (0.1 uV) accuracy. We provide also a version of the data that is downsampled at 100 Hz (by picking each 10th sample) that we typically use for analysis.

References

- **Guido Dornhege, Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Müller**. Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Trans. Biomed. Eng.*, 51(6):993-1002, June 2004.

Note that the above reference describes an older experimental setup. A new paper analyzing data sets similar to the one provided in this competition and presenting feedback results will appear soon.

[[BCI Competition III]()]