Data set IVa ‹*motor imagery, small training sets*›

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, one has to have labelled training data to teach the classifer. To this end, the user usually performs a boring calibration measurement before starting with BCI feedback applications. One important objective in BCI research is to reduce the time needed for the initial measurement. This data set poses the challenge of getting along with only a little amount of training data. One approach to the problem is to use information from other subjects' measurements to reduce the amount of training data needed for a new subject. Of course, competitors may also try algorithms that work on small training sets without using the information from other subjects.

Experimental Setup

This data set was recorded from five healthy subjects. Subjects sat in a comfortable chair with arms resting on armrests. This data set contains only data from the 4 initial sessions without feedback. Visual cues indicated for 3.5 s which of the following 3 motor imageries the subject should perform: (L) *left* hand, (R) *right* hand, (F) right *foot*. The presentation of target cues were intermitted by periods of random length, 1.75 to 2.25 s, in which the subject could relax.
There were two types of visual stimulation: (1) where targets were indicated by letters appearing behind a fixation cross (which might nevertheless induce little target-correlated eye movements), and (2) where a randomly moving object indicated targets (inducing target-uncorrelated eye movements). From subjects *al* and *aw* 2 sessions of both types were recorded, while from the other subjects 3 sessions of type (2) and 1 session of type (1) were recorded.

Format of the Data

Given are continuous signals of 118 EEG channels and markers that indicate the time points of 280 cues for each of the 5 subjects (*aa*, *al*, *av*, *aw*, *ay*). For some markers no target class information is provided (value NaN) for competition purpose. Only cues for the classes 'right' and 'foot' are provided for the competition. The following table shows the respective number of training (labelled) trials "#tr" and test (unlabelled) trials "#te" for each subject.

	#tr	#te
aa	168	112
al	224	56
av	84	196
aw	56	224
ay	28	252

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
  - `pos`: vector of positions of the cue in the EEG signals given in unit *sample*, length #cues
  - `y`: vector of target classes (1, 2, or `NaN`), length #cues
  - `className`: cell array of class names.
- `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 (splitted into three files for each subject):

- `*_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= right, 2=foot, or 0 for test trials).
- `*_nfo.txt`: contains other information as described for the matlab format.

Requirements and Evaluation

Please provide for each subject an ASC II file (named 'result_IVa_aa.txt', 'result_IVa_al.txt', ...) containing 280 lines of your estimated class labels (1 or 2) for every cue. (For training trials this should be the respective value of `mrk.y`, and for test trials the output of your algorithm.)
You also have to provide a description of the used algorithm (ASC II, HTML or PDF format) for publication at the results web page.

The performance measure is the overall classification accuracy (number of correct classified test trials divided by the total number of test trials).

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 the data sets as provided in this competition and presenting the feedback results will appear soon.

[BCI Competition III]