BCI Competition III – dataset V

Julien Kronegg, Douglas Rofes
Computer Vision and Multimedia Lab,
University of Geneva, Switzerland
May 2005

Introduction

In the context of the 3rd BCI competition, we are processing the dataset V in order to compare our current classification method other methods.

Data and features

We used the dataset V provided by the competition organisers[1]. These data come from the IDIAP research institute in Switzerland, where 3 subjects were recorded while doing 3 possible mental tasks (imagination of left and right self-paced hand movement, and word generation) [Millan2004]. A full description of the dataset is available on the competition web site[2]. We used the pre-computed features. The test sets are unlabelled in order to unsure that competitors will not tune their algorithms on already seen data. The classification methods must so be tested on the known datasets only.

Methods

It is of common knowledge that the classifier accuracy depends on the subject. Consequently, we tested 4 classifiers: a SVM[3], a CART[4] decision tree, a LVQ[5], and a Naïve Bayes[6] classifier. Each submitted result will be computed using the supposed "best classifier" for each subject.

Expected results

In order to assess the quality of our classification method, we computed an approximation of the maximum, minimum and typical accuracy.

Maximum accuracy estimation

The classification accuracy upper bound is estimated by a CART 5-folds cross-validation on the training samples. This estimation (see Table 1) shows that Subject1 seems to be the best subject and Subject3 the worst.

Table 1 – Cross-validation accuracy on training set.

TRN/TST	S1	S2	S3
Dataset1/x-val	75.8	73.2	68.9
Dataset2/x-val	80.8	77.1	70.9
Dataset3/x-val	84.2	76.7	68.3

Minimum accuracy estimation

The lower accuracy bound is given by the probability of the most common class. If the classes were equiprobable, this lower bound would be 33.3%, but Table 2 shows that it is not the case.

Table 2 – Minimum accuracy.

max(p(xi))	S1	S2	S3
Dataset 1	41.3	42.4	34.6
Dataset 2	39.2	34.3	34.1
Dataset 3	40.8	41.5	34.0

Typical accuracy estimation

The dataset V reference article [Millan2004] gives the off-line classification rate for some datasets. For the 3 subjects, the classifier has been trained on D1 and D1+D2 and tested on the last 3 minutes of D2 and D3 respectively. We have converted the original error/reject rate to accuracy (see Table 3) using:

$$accuracy=100-(error+reject*2/3),$$

considering that a rejected instance could be classified with a 33.3% accuracy, given that the 3 classes are equiprobable. Of course, this is an approximation and the test set are not strictly the same, so these numbers should be used to rough comparison only.

Table 3 – Approximate accuracy for [Millan2004].

TRN/TST	S1	S2	S3
D1/last3m(D2)	62.2	49.1	47.6
D1+D2/last3m(D3)	69.3	54.6	40.0

Training set and classifier selection

Since it is difficult to evaluate the classifier results (unlabelled test sets), we proceeded by successive steps. Note that all the results

[1] http://ida.first.fhg.de/projects/bci/competition_iii/
[2] http://ida.first.fraunhofer.de/projects/bci/competition_iii/desc_V.html
[3] OSU-SVM Matlab Toolbox 3.0, RBF kernel, γ=2
[4] Matlab Statistical toolbox.
[5] Matlab Neural Network toolbox, S1=size(TRN)/50.
[6] Designed by Douglas Rofes.

presented here are given in percents and for 3 classes.

Effect of the training set size

The CART classifier has been trained on a part of the dataset, and tested on another part of the dataset. Table 4 and Table 5 show that increasing the training set size generally improve the classifier accuracy. Thus, in the next stage, we will use greater training sets.

Table 4 – Accuracy on 50%/50% dataset.

TRN/TST	S1	S2	S3
Dataset 1	51.0	45.2	35.0
Dataset 2	59.1	51.7	45.3
Dataset 3	57.9	50.0	38.1

Table 5 – Accuracy on 74%/26% dataset.

TRN/TST	S1	S2	S3
Dataset 1	54.3	44.5	40.7
Dataset 2	67.0	60.3	48.0
Dataset 3	55.4	57.6	36.8

Effect of the classification method

We have tested four classifiers as specified previously. As shown in Table 6 to Table 9 and in Figure 1, the SVM classifier is generally better than the others classifiers (up to 8%), but interestingly, this is not always the case, especially for bad subjects.

Table 6 – Accuracy for CART.

TRN/TST	S1	S2	S3
D1/D2	60.8	48.4	44.9
D2/D3	64.3	51.2	41.3
D1+D2/D3	64.9	54.2	40.6

Table 7 – Accuracy for SVM.

TRN/TST	S1	S2	S3
D1/D2	69.2	49.9	50.8
D2/D3	73.3	60.3	37.4
D1+D2/D3	72.8	60.3	42.2

Table 8 – Accuracy for Naïve Bayes.

TRN/TST	S1	S2	S3
D1/D2	66.9	42.7	47.9
D2/D3	72.8	58.4	39.9
D1+D2/D3	71.2	57.8	42.6

Table 9 – Accuracy for LVQ.

TRN/TST	S1	S2	S3
D1/D2	57.6	48.3	53
D2/D3	64.6	49.4	35.1
D1+D2/D3	54.1	61.3	39.6

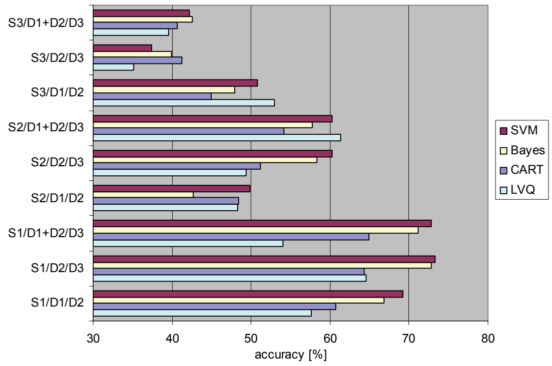

Figure 1 – Classifiers comparison.

Table 3 shows that the accuracies are on the same order as for CART, Bayes or SVM, which shows that our classification method is valid.

Classifier used

Given the results obtained for labelled test sets, we choose a SVM classifier because it gives the best results in most of the training conditions. If we use datasets D1+D2+D3 as training set, the classification accuracy should be approximately the same as accuracies from Table 7, so the mean accuracy would be about 58.4% on the unlabelled test set.

Training data used

Given the good results generally obtained when using more than one dataset for training, we will generally use D1+D2+D3 as a training dataset and D4 as test dataset. For Subject3, D2 is discarded from the training set, because the classification accuracy obtained when D2 was in the training set was not good for the 3 classifiers. Some training instances (3-5%) are discarded using an instance selection method derived from stratified random sampling, which should improve the accuracy by 3-4%.

Results

The classified labels are stored in separate Matlab6 ".mat" files. The submitted results correspond to (subject/training/test):

S1/D1+D2+D3/D4
S2/D1+D2+D3/D4
S3/D1+D3/D4

References

Millán, J. del R.. On the need for on-line learning in brain-computer interfaces *Proc. Int. Joint Conf. on Neural Networks.*, 2004.