Ensemble Learning in BCI-SSVEP Systems
for Short Window Lengths[*]

Henrique L. V. Giuliani[1]**, Patrick O. de Paula**[1]**,**
Diogo C. Soriano[2]**, Ricardo Suyama**[2]**, Denis G. Fantinato** [1]

[1]Centro de Matemática, Computação e Cognição
Universidade Federal do ABC (UFABC) – Santo André – SP – Brazil

[2]Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas
Universidade Federal do ABC

{henrique.voni, ricardo.suyama,
diogo.soriano, denis.fantinato}@ufabc.edu.br
patrick.oliveira@aluno.ufabc.edu.br

***Abstract.** Different approaches have been investigated to implement effective Brain-Computer Interfaces (BCI), translating brain activation patterns into commands to external devices. BCI exploring Steady-State Visually Evoked Potentials usually achieve relatively high accuracy, when considering 2-3 second sample windows, but the performance degrades for smaller windows. So, we investigate the use of an ensemble method, the Adaboost algorithm, combining two different structures, the Logistic Regressor and the Multilayer Perceptron, whose diversity shall bring relevant information for more accurate classification. Simulation results indicate that the proposed method can improve performance for smaller windows, being a promising alternative to reduce model variance.*

1. Introduction

A Brain-Computer Interface (BCI) system is an apparatus capable of translating brain activation patterns into commands to external device, which varies from assistive technologies to entertainment [Nam et al. 2018]. Generally, brain activity is measured by means of electroencephalography (EEG) recorded, which is a non-invasive and relatively low cost technique. However, EEG signals present low spatial resolution, noise, and artifacts. Besides, brain activation patterns tends to change significantly along time for a single subject (BCI user), making EEG signals processing and classification a very challenging problem [Nam et al. 2018].

After a pre-processing step, the EEG signal is usually classified by a Machine Learning (ML) technique, in which we highlight the Artificial Neural Networks (ANNs), such as the Multilayer Perceptron (MLP) [Bishop 2006]. Indeed, in the context of Steady State Visually Evoked Potential (SSVEP), the achieved performance in terms of accuracy is relatively high, about 90% for sample windows (with period) of 3 and 2 seconds [Nam et al. 2018]. However, for smaller sample windows, such as 1 or even 0.5 seconds, the achieved performance still requires improvement and is a crucial issue in BCI applications that demand high accuracies in the smallest response time.

In that sense, in this work, we investigate the use of an ensemble method, the Adaboost algorithm [Bishop 2006], as an attempt for dealing with high variability in smaller sample windows. The proposed ensemble will combine two different structures, the logistic regressor and the MLP, whose diversity shall bring relevant information for more accurate classification.

2. Brain-Computer Interfaces and Ensemble Learning

Among many approaches, BCI approaches make use of signals encoded in EEG activity [Nam et al. 2018]. Once collected, signal processing techniques attempt to extract relevant information from EEG data that will be used by machine learning algorithms to map the original intent to a specific command. Researchers have exploited a broad set of brain response types for building BCIs. One particular case and main focus of this work are BCIs systems based on SSVEP, which is the response provoked in neurons of the visual cortex when the subject is fixating on a visual stimulus. Visual stimuli are associated with different commands and oscilate in different frequencies presented to the user by a visual device. Spectral-based SSVEP decoders usually evaluate peak intensity on the visual stimuli (and their harmonics) aiming to define the features for a classification structure. However, the smaller the fraction of window length used as an input, the harder it is to classify it [Nam et al. 2018]. In light of this tradeoff, more robust models seem to be a feasible solution to handle complex scenarios, and ensemble learning can provide a path to explore hard classification tasks involving BCI applications.

Ensemble learning is a technique that combines several base learners to generate an improved predictive model. We can highlight the *Boosting* approach, where base learners are sequentially trained in a way that the samples presented to the current base learner depend on the incorrect outputs of its predecessor. One of the most popular strategies of boosting is the *Adaboost* algorithm [Freund et al. 1999], which aims to perform a weighted voting among all base learners to generate the ensemble output. Originally proposed for binary classification tasks, Adaboost was improved to properly handle multiclass datasets through SAMME (Stagewise Additive Modeling using a Multi-class Exponential loss function) [Hastie et al. 2009]. SAMME applies subtle differences over the original structure to handle a multiclass scenario in a scenario where the importance (*alpha*) factor of each base learner is always positive and better than random guessing. In order to investigate the data space from different perspectives, we combined two different models into an ensemble: linear (*Logistic Regression - LR*) [Bishop 2006] and nonlinear (ANN) [Bishop 2006] models.

3. Methodology and Results

The dataset used in this work for running experiments is a contribution of previous researches authored by the Laboratory of Computing Methods for Bioengineering (Federal University of ABC) [da Silva Jr. 2017]. Data was collected[1] using a 16-channel EEG cap, with electrodes positioned according to the 10-20 system. 15 healthy subjects participated in 12 6-second sessions for each of the 4 frequencies of interest (10, 11, 12 and 13Hz – frequency spectrum examples are illustrated in Figure 1) – adding up to 48 sessions per subject. The sample rate used for data collection was 256Hz.

---

[1]Approved by the Ethics Committee of the Federal University of ABC (CAAE:51005615.2.0000.5594)
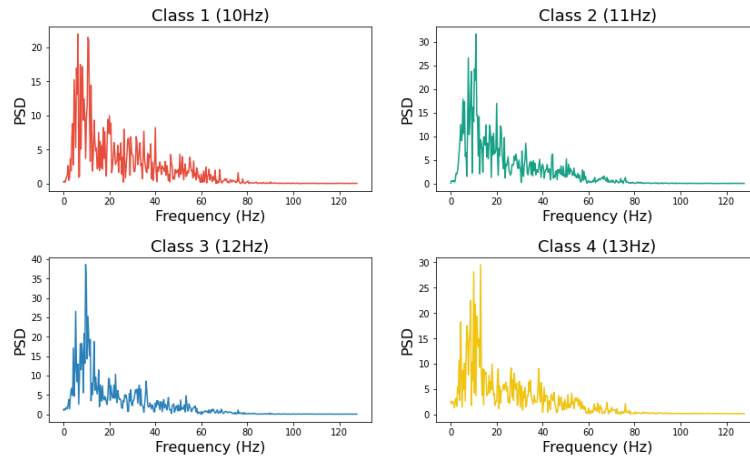
Figure 1. Example of power spectral density analysis of samples of each class.

3.1. Dataset description

Data acquisition sessions were perfomed using a monitor presenting different light sources in a checkered pattern, which oscillated at different frequencies (each of which is associated with a particular class). To carry out the experiments, EEG datasets were partitioned into time windows of 2 and 1 seconds with a 0.5-second overlap (or 128 points) and a 0.5-second time window without overlap; totalizing respectively: 240, 288 and 576 samples per individual. It is worth mentioning that the use of windows or acquisition periods equal to 3 and 2 seconds have been previously used in the literature [da Silva Jr. 2017]. However, cases with 1 and 0.5 s windows, as it has lower spectral resolution and is consequently more challenging, still need more attention.

The SSVEP-BCI system is composed of (*i*) Common Average Reference (CAR) filter as a preprocessing step; (*ii*) Fast Fourier Transform algorithm, in which the used features were the first 3 harmonics of the target frequencies (a normalization step was previously added to standardize data with mean 0 and unit variance to keep all features at an equivalent scale); and (*iii*) a classifier, which in this case were a linear (Logistic Regression), a nonlinear (Multilayer Perceptron - MLP [Bishop 2006]) or an ensemble mixing up both types of classifiers (Adaboost - ADA). Then, we can compare the robustness of the classifiers in different scenarios while searching for a BCI system with smaller window lengths.

For all experiments, MLP networks were implemented to have 1 hidden layer with 15 neurons, using a sigmoidal activation function and a maximum number of 2000 epochs. The LR classifier was configured with a maximum number of 2000 iterations as well. Finally, Adaboost was implemented with 15 base learners, being 7 LR classifiers and 8 MLPs. Regarding experiment settings, data was split into 50% for training and 50% for testing adopting accuracy as an evaluation metric.

3.2. Results and discussion

Firstly, we investigated a reasonable scenario using a 2-second window and then stressing out models in more challenging scenarios with a sample size of 1 and 0.5 seconds. All results are described in Table 1.

Table 1. Experiment results for 2, 1 and 0.5 s time window samples.

SUBJECT	2 Second			1 Second			0.5 Second		
	LR	MLP	ADA	LR	MLP	ADA	LR	MLP	ADA
1	93.33%	95.83%	94.16%	87.50%	88.19%	88.88%	70.13%	71.87%	72.22%
2	89.16%	89.16%	89.16%	65.27%	65.27%	65.27%	52.43%	52.08%	52.43%
3	99.16%	99.16%	99.16%	87.50%	90.27%	88.19%	71.52%	73.95%	74.65%
4	100.00%	100.00%	100.00%	90.97%	89.58%	90.27%	69.44%	71.52%	71.52%
5	97.50%	95.83%	98.33%	81.25%	84.02%	86.80%	52.43%	54.51%	55.55%
6	100.00%	97.50%	98.33%	83.33%	81.94%	83.33%	70.08%	72.22%	72.91%
7	69.16%	60.83%	64.16%	49.30%	45.13%	45.83%	35.76%	40.27%	39.93%
8	100.00%	97.50%	98.33%	92.36%	94.44%	93.75%	73.95%	72.56%	72.56%
9	82.50%	85.00%	85.83%	51.38%	59.02%	58.33%	39.93%	40.27%	43.75%
10	91.66%	95.00%	96.66%	76.38%	74.30%	74.30%	48.61%	53.12%	53.81%
11	96.66%	99.16%	99.16%	84.72%	87.50%	86.80%	59.72%	60.06%	61.11%
12	98.33%	99.16%	97.50%	86.11%	85.41%	87.50%	52.08%	57.29%	56.94%
13	94.16%	94.16%	95.00%	72.22%	74.30%	75.00%	51.73%	50.34%	50.34%
14	100.00%	100.00%	100.00%	99.30%	100.00%	99.30%	95.48%	95.13%	94.44%
15	98.33%	98.33%	97.50%	90.97%	92.36%	91.66%	58.68%	62.50%	62.15%
MEAN	**93.99%**	**93.77%**	**94.22%**	**79.90%**	**80.78%**	**81.01%**	**60.13%**	**61.85%**	**62.29%**

The ensemble model presents higher mean accuracy in all cases, although the difference is higher for shorter window lengths – a gain of 0.45% for windows of 0.5 second. Although it is not an expressive improvement, results indicates that the ensemble is a promising approach and performs better in more complex scenarios, where the underlying information about input data is scarce. The diversity of base learners also has its contribution, since they explore the data space in different ways. This is confirmed by the results obtained on subjects 9 and 10 in the hardest case, where the difference among Adaboost and the other classifiers reaches 3.82% and 5.20%.

4. Conclusions

Designing a robust BCI system is a challenging task, in which providing correct output in the smallest response time is mandatory. Specifically, it is desired to correctly classify the user's intent using the shorter time window length as possible. In this work, we started performing a classification task in a comfortable scenario and kept investigating how the sample size could be associated with the complexity of models. We conclude that the ensemble outperforms simpler ones as the sample period gets smaller. Also, it is important to highlight the diversity of base learners, which contributed to the ensemble in most cases. For further works, more sophisticated ensemble approaches represent a promising alternative to reduce model variance, such as the Mixture of Experts approach.

References

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

da Silva Jr., J. I. (2017). Comparativo de desempenho de sistemas bci-ssvep off-line e em tempo de execução utilizando técnicas de estimação de espectro e análise de correlação canônica. Master's thesis, UFABC.

Freund, Y., Schapire, R., and Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612.

Hastie, T., Rosset, S., Zhu, J., and Zou, H. (2009). Multi-class adaboost. *Statistics and its Interface*, 2(3):349–360.

Nam, C. S., Nijholt, A., and Lotte, F. (2018). *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*. CRC Press.