Pushing the limits of BCI accuracy: Winning solution of the Grasp & Lift EEG challenge.

Conference Paper · May 2016

Alexandre Barachant
Burke Neurological Institute

Pushing the limits of BCI accuracy: Winning solution of the Grasp & Lift EEG challenge.

A. Barachant[1*], R. Cycon[2]

[1]Burke Medical Research Institute, White plains, NY, USA. [2]FORNAX sp. z o.o., Krakow, Poland

*E-mail: alexandre.barachant@gmail.com

Introduction: To better understand the relationship between EEG signals and hand movements the WAY Consortium has organized the Grasp-and-Lift EEG Detection challenge. It was held in 2015 from 29th June to 31th August on Kaggle, a platform for competitive predictive modeling, and attracted 379 contesting teams. The goal of the challenge was to detect 6 different events related to hand movement during a task of grasping and lifting an object, using only EEG signal. The 6 events were representing different stages of a sequence of hand movements (hand starts moving, starts lifting the object, etc.). True labels were extracted from EMG signal, and provided as a +/-150ms frame centered on the occurrence of the event. Contestants were asked to provide probabilities of detection for the 6 events and for every time sample. The evaluation metric for this challenge was the Area Under the ROC Curve (AUC) averaged over the 6 event types. Finally, the model must be causal, i.e. only the data from the past can be used to predict the events. This abstract presents the winning solution of this challenge.

Material, Methods and Results: EEG data that was provided was recorded with 64 electrodes on 12 different subjects while they were performing around 300 tasks of grasping and lifting an object. In many ways, the formulation of the problem differed from a typical motor imagery BCI problem: 1- The 6 events were representing different stages of a sequence of hand movements and therefore the temporal structure of the sequence had to be taken into account. In addition, some events were overlapping, and some others were mutually exclusive. 2- the events to detect were short timed (300ms) and positives predictions have to be provided for the entire frame. The sharpness of the prediction was critical for achieving optimal accuracy. 3- the predictions have to be provided for every time sample (3 million in total), which represents a considerable amount of data.

As a consequence, most of the current approaches used in motor imagery failed to produce accurate results. For this challenge, we used three different types of features: time domain signal low pass filtered by a bank of filters, covariance matrices estimated on different time window and frequency band, and a special form covariance dedicated to asynchronous detection of evoked potential. In the above context, we employed stacking to build a 3-level classification pipeline described in Figure 1.

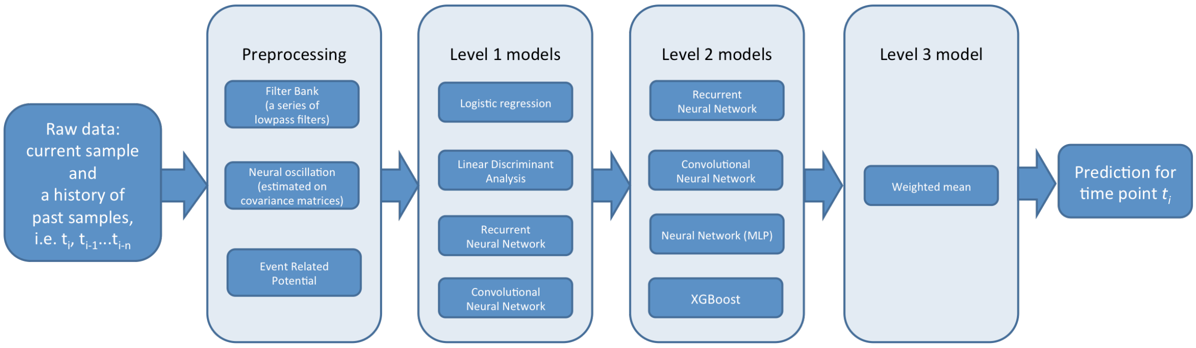

Figure 1. *Overview of the 3-level classification pipeline.*

Level1 models provided support and diversity for level2 models by embedding subject and events specificities using different types of features. A total of 51 level1 models were developed, the best was a convolutional neural network and achieved 0.95 AUC. Level2 models are global models (i.e. not subject-specific) that are trained on level1 predictions. Their main goal is to take into account the temporal structure and relationship between events. 32 level2 models were used, the best being a recurrent neural network (0.98 AUC). Level3 models ensemble level2 predictions via an algorithm that optimizes level2 models' weights to maximize AUC. This step improves the sharpness of predictions while reducing overfitting. The final model scored 0.981 AUC and allowed our team to take the first position of this challenge.

Discussion: Robust features, advanced machine learning methods, and algorithms able to model highly nonlinear relationships were crucial to achieving top performance. The score was gradually boosted with each stacking level, at the costs of increased solution complexity and computation time, without risk of overfitting due to the amount of data and the stability across time. Significant improvements of the score was achieved with the addition of level2 recurrent neural networks to the ensemble, which were able to model relationships between events and the temporal structure of the sequence of hand movements more accurately than other algorithms.