Learning Dictionaries of Spatial and Temporal EEG Primitives for Brain-Computer Interfaces

Benjamin Hamner   BENJAMIN.HAMNER@EPFL.CH
Ricardo Chavarriaga   RICARDO.CHAVARRIAGA@EPFL.CH
Jose del R. Millán   JOSE.MILLAN@EPFL.CH
École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

Abstract

Sparse methods are widely used in image and audio processing for denoising and classification, but there have been few previous applications to neural signals for brain-computer interfaces (BCIs). We used the dictionary-learning algorithm K-SVD, coupled with Orthogonal Matching Pursuit, to learn dictionaries of spatial and temporal EEG primitives. We applied these to P300 and ErrP data to denoise the EEG and better estimate the underlying P300 and ErrP signals. This methodology improved single-trial classification performance across 13 of 14 subjects, indicating that some of the background noise in EEG signals, presumably from neural or muscular sources, is highly structured. Furthermore, this structure can be captured via dictionary learning and sparse coding algorithms, and exploited to improve BCIs.

1. Introduction

Sparse methods, including unsupervised dictionary learning algorithms and sparse coding algorithms, have been widely used in audio and visual processing for denoising and classification. Dictionary learning algorithms take in unlabeled visual data (pixel intensity values) or audio data (spectrograms) and then construct dictionaries of elements that commonly occur in this data. Sparse coding algorithms approximate points in the dataset as the linear combination of a small number of dictionary elements. This nonlinear transformation creates a new representation of the data, which may be more useful for denoising and supervised machine learning applications than the raw dataset. (Coates et al., 2010; Klein et al., 2003)

In this paper, we apply sparse methods to electroencephalogram (EEG) signals for applications in non-invasive Brain-Computer Interfaces (BCIs). Many different EEG-based brain-computer interfaces BCIs have recently been developed. In these, features are extracted from EEG signals to restore communication to patients with locked-in syndrome or to aid patients with muscular deficiencies. EEG signals useful for BCI applications can be split into two broad categories: event-related potentials (ERPs) and spontaneous rhythms. ERPs occur as a result of external stimulation, and include the P300 response and Error Potential (ErrP). Spontaneous rhythms, such as the sensorimotor rhythm (SMR) can be modulated spontaneously by the subject, without any external stimuli (Millán et al., 2010). For this study, we focus on the P300 and ErrP signals.

The P300 signal occurs during oddball paradigm, where subjects observe randomly occurring stimuli and attend one of them. A positive deflection in central locations on the scalp EEG occurs approximately 300 ms after the attended stimuli. This is termed the P300 and has been used to make BCI applications for communication with severely paralyzed patients (Donchin et al., 2000). The ErrP signal is observed when subjects perceive erroneous actions, and is characterized by activity in fronto-central areas approximately 200–500 ms after the error is observed (Chavarriaga & Millán, 2010). P300 BCIs discriminate between attended and unattended stimuli, and BCIs using ErrPs discriminate between correct and erroneous actions.

While sparse techniques have been applied to blind source separation in EEG (Studer et al., 2006), these have involved predefined dictionaries and have not been geared towards BCI applications. Other approaches include using dynamic Bayesian networks to

model the non-stationary temporal structure of EEG SMRs (Song et al., 2009). In this study, we apply sparse coding algorithms to learn dictionaries of spatial and temporal EEG primitives across subjects in an unsupervised manner. We then use these primitives as models of background neural activity and other sources of noise to denoise single-trial estimates of P300 and ErrP signals.

2. Methodology

2.1. Experimental Protocol

For the P300 data, eight subjects were recorded with the BCI2000 P300 Speller application. Each subject was recorded for a training session and testing session, with testing sessions 5–8 days after the training sessions. The sessions consisted of 2–6 runs, where subjects were given a word to spell. In each run, a 6-by-6 grid of alphanumeric characters was displayed on the screen. At the beginning of each trial, the next letter in the word was displayed on the top of the screen. Subjects were instructed to count the number of times this letter flashed. Then, rows and columns individually flashed at a rate of 4.5 Hz in pseudorandom order, with each row and column going through 15 intensifications. EEG signals were recorded at a 250 Hz sampling rate.

The ErrP experimental protocol and data described in (Chavarriaga & Millán, 2010) was used for this study. Six subjects were recorded on two different days using a Biosemi ActiveTwo system with a sampling rate of 512 Hz. Subjects observed a green square ("cursor") on a horizontal line of 20 squares on a computer screen. The target square was highlighted, and the cursor moved towards the target with probability 0.8 and away from it with probability 0.2. Data from the first recording day was used for training, and data from the second was used for testing. The time between the training and testing sessions ranged from 50–643 days.

2.2. EEG Preprocessing

All EEG signals were re-referenced with the common average reference and band-pass filtered with a 0.5–10 Hz 4th-order Butterworth filter. No trials were removed from either experimental dataset.

2.3. Dictionary Learning

K-SVD was used to learn dictionaries of spatial and temporal primitives across the P300 training data and ErrP training data. K-SVD was investigated as a generalization of the N-Microstate algorithm[1] (Pascual-Marqui et al., 1995) that permits multiple simultaneous component activations. K-SVD alternates between a sparse coding stage and a codebook update stage (Aharon et al., 2006). Orthogonal Matching Pursuit (OMP), which starts with an empty set of dictionary atoms and then greedily adds atoms in order to minimize reconstruction error, was used in the sparse coding stage (Pati et al., 1993). K-SVD is parameterized by the number of dictionary elements and OMP is parameterized by the maximum number of atoms used to reconstruct the signal. A dictionary of 100 primitives was learned for each feature set, with two components simultaneously active.

2.4. Feature Extraction

Epochs in the P300 dataset were taken from 0–732 ms post-stimulus and downsampled by a factor of three, giving a 61 time-sample by 61 channel matrix per epoch. Epochs in the ErrP dataset were taken from 0–750 ms post-stimulus and downsampled by a factor of 6, giving a 64-by-64 matrix per epoch.

For each P300 subject, the grand average was computed across the P300 epochs to estimate the underlying neural signal. The same was done for the ErrP subjects. This provided a set of spatial filters (rows of the grand average matrices) and temporal filters (columns) for feature extraction. Two types of features were calculated in each domain: raw and denoised. To calculate the raw spatial features, we took the dot products between the signal at each time point in the individual epochs and the corresponding time point in the grand average. Raw temporal features were computed in a corresponding manner.

To calculate the denoised features, we used the dictionaries of spatial or temporal elements as a model of the noise in the EEG signal (which may be from background neural activity, muscular artifacts, or electrical noise) and applied a modified version of OMP. Instead of initializing OMP with an empty set, we added the spatial or temporal component of the grand average for the estimated feature to the corresponding dictionary, and automatically included this component in the sparse approximation. In this manner, we estimated the current component of EEG activity that could be due to an underlying P300 or ErrP, after "explaining away" a portion of the noise with the learned dictionary. A maximum of five dictionary elements were used in the sparse approximation of the signal.

---

[1]This algorithm was designed to segment EEG data into topographic maps, with only one map active at any point in time.

Sample EEG Spatial Primitives
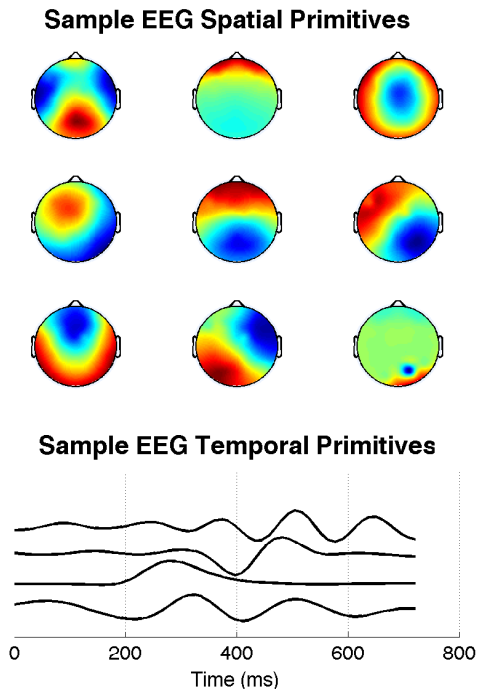
Sample EEG Temporal Primitives

Figure 1. Samples of spatial and temporal EEG primitives learned from the P300 subjects' training data

2.5. Classification

Fischer's Linear Discriminant (FLD) was used to classify each of the feature sets. Area under the receiver-operator characteristic (AUC) was used to evaluate the performance of the different feature sets after classification. For each subject and feature set, forty bagged training and testing sets were constructed and evaluated.

3. Results and Discussion

Figure 1 shows examples of spatial and temporal primitives learned through K-SVD and OMP on the preprocessed EEG data. The spatial primitive with a strong frontal activation (top-middle) results from EEG artifact contamination due to eye blinks, the one on the bottom-right comes from noise and artifacts in a single occipital electrode, and the rest likely have neural origins. Some of the temporal primitives show temporary deflections in the EEG signal, whereas others show sustained oscillatory behavior. While only a few elements of the spatial and temporal dictionaries learned for the P300 training data are depicted, these are qualitatively similar to the remaining dictionary elements and to the dictionaries learned for ErrP data.

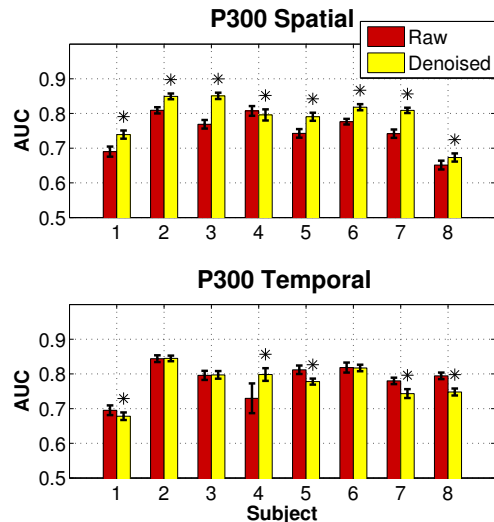Figure 2 shows the classification performance of dif-

Figure 2. The classification performance of spatial and temporal features on P300 test data. * designates denoised features that show a significant difference in performance from the corresponding raw features ($p < 0.01$).
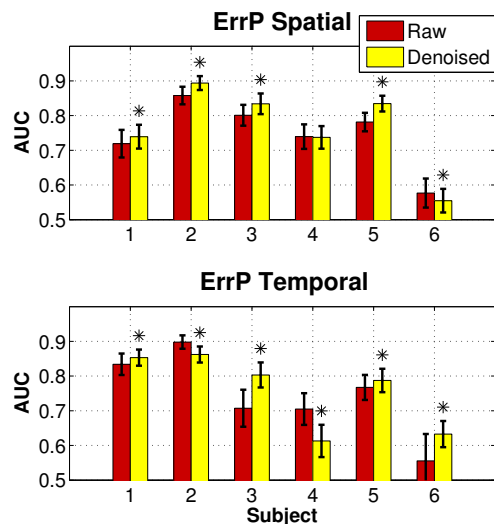

Figure 3. The classification performance on the ErrP data.

ferent features for the P300 dataset. For 7 out of the 8 subjects, the denoised spatial features show a significant improvement compared to the raw spatial features ($p < 0.01$). These performance gains, however, do not hold with the temporal features. Only one subject (the subject that showed a decrease in performance with the spatial features) showed significant performance improvements with the denoised temporal features, three saw no significant change, and four saw a significant decrease.

Figure 3 shows the classification performance of different features for the ErrP dataset. Both the denoised spatial features and denoised temporal features show significant increases in performance relative to the raw features on four of six subjects, and there was only one subject that did not show a significant increase in performance on at least one of the denoised features.

These results indicate that EEG signals have a sparse spatial and temporal structure, and that this structure may be utilized to improve signal classification for BCIs across multiple subjects and modalities. This merits a more extensive investigation into sparse modeling of the spatiotemporal structure of EEG and comparison with existing techniques, such as Independent Component Analysis. This methodology could also be coupled with inverse modeling techniques to learn primitives directly attributable to specific neural sources.

The dictionary-learning algorithm is computationally intensive, but can be run offline and does not require subject-specific training data. The modified version of OMP used for denoising single trial EEG signals is suitable for online applications.

The methods presented here depend on two underlying assumptions. The first is that each of the P300 and ErrP signals is the result of a corresponding constant underlying neural mechanism. If the neural mechanisms vary according to stimulus or another factor, then using the grand average to approximate the neural response and then comparing single trials to this grand average is a suboptimal learning mechanism. The second is that the ERP is jitter-free. If there is a substantial jitter (variation in the timing of the ERP relative to the stimulus), then the grand average will not represent the underlying neural signal in signal trials. Jitter may be one cause of the relatively lower performance of the denoised temporal P300 features.

4. Conclusions and Future Work

In this study, we have demonstrated that EEG has a sparse spatial and temporal structure that can be used to denoise the EEG signal. This increases the offline single-trial classification performance of EEG signals across two modalities, and we hypothesize that these performance gains will hold in online BCI applications.

This preliminary investigation considered a single unsupervised dictionary-learning algorithm and sparse coding algorithm. In the future, we intend to explore a larger variety of sparse techniques, along with deep learning and Bayesian methods, for their applicability to EEG signals. Also, we considered the spatial and

temporal structure independently. Future work will consider joint spatiotemporal decompositions of EEG signals. Additionally, we plan to investigate whether the performance gains we found for two types of ERPs, the P300 and ErrP, extend to other EEG signals, including slow cortical potentials and SMRs.

5. Acknowledgements

References

Aharon, M., Elad, M., and Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Sig. Proc.*, 54(11):4311–22, 2006.

Chavarriaga, R. and Millán, J.d.R. Learning from EEG error-related potentials in noninvasive brain-computer interfaces. *IEEE Trans. Neural Sys. and Rehabilitation Eng.*, 18(4):381–8, 2010.

Coates, A., Lee, H., and Ng, A. An analysis of single-layer networks in unsupervised feature learning. In *NIPS*, 2010.

Donchin, E., Spencer, K.M., and Wijesinghe, R. The mental prosthesis: assessing the speed of a p300-based brain-computer interface. *IEEE Trans. Rehabilitation Eng.*, 8 (2):174–9, 2000.

Klein, D. J., König, P., and Körding, K. P. Sparse spectrotemporal coding of sounds. *EURASIP J Applied Sig. Proc.*, 7:659–67, 2003.

Millán, J.d.R., Rupp, R., Müller-Putz, G.R., Murray-Smith, R., Giugliemma, C., Tangermann, M., Vidaurre, C., Cincotti, F., Kübler, A., Leeb, R., Neuper, C., Müller, K.-R., and Mattia, D. Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. *Frontiers in Neuroscience*, 4: 161–176, 2010.

Pascual-Marqui, R.D., Michel, C.M., and Lehmann, D. Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Trans. Biomed Eng.*, 42(7):658–65, 1995.

Pati, Y.C., Rezahfar, R., and Krishnaprasad, P. S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. of the 27th Annual Asimolar Conference on Signals Systems and Computers*, 1993.

Song, L., Kolar, Mladen, and Xing, E. P. Time-varying dynamic bayesian networks. In *NIPS*, 2009.

Studer, D., Hoffmann, U., and Koenig, T. From EEG dependency multichannel matching pursuit to sparse topographic EEG decomposition. *J. Neuroscience Meth.*, 153(2):261–75, 2006.