Proceedings of the
7th Graz Brain-Computer Interface Conference 2017 DOI: 10.3217/978-3-85125-533-1-84

SPATIAL FILTERING OF EEG AS A REGRESSION PROBLEM

Martin Spüler[1]

[1]Department of Computer Engineering, Eberhard Karls University Tübingen, Tübingen, Germany

E-mail: spueler@informatik.uni-tuebingen.de

ABSTRACT: In the field of Brain-Computer Interfaces (BCIs), Electroencephalography (EEG) is a widely used, but very noisy method. To improve signal-to-noise ratio (SNR) of the recorded signals, spatial filtering is commonly applied. This paper concentrates on spatial filtering methods to enhance the SNR of evoked- or event-related potentials (ERPs). While methods like Canonical Correlation Analysis (CCA) or xDAWN have been shown to provide good spatial filters, this paper introduces an alternative view on spatial filtering, showing that spatial filtering can be seen as a regression problem. It is shown how regression methods can be used to construct spatial filters and their use is evaluated on an EEG dataset containing error-related potentials (ErrPs), showing that classification accuracy is significantly improved using regression-based spatial filtering. As arbitrary regression methods can be used for construction of spatial filters, non-linear spatial filters can be constructed and new approaches, like deep learning, can be used for spatial filtering.

INTRODUCTION

A Brain-Computer Interface (BCI) allows a person to control a computer by using only his brain activity, without the need for muscle control [1]. While its main goal is to enable communication in paralyzed patients [2], it is also used in other fields like rehabilitation of stroke patients or the detection of mental states. As Electroencephalography (EEG) is a relatively cheap and non-invasive method, it is commonly used to measure the brain activity for the use with BCI. However, EEG is a rather noisy technique, which makes it difficult to correctly interpret the recorded brain signals.

One commonly used method to improve the signal-to-noise-ratio (SNR) of EEG, is the use of spatial filters. Spatial filters can be seens as mathematical operation, which mixes the signal from the EEG electrodes in a way that the signal of interest is enhanced, while noise or artifactual components are reduced. This can be implemented by a linear transformation matrix W_s that transforms the raw input signal X_r into the spatially filtered signal X_s.

$$X_s = W_s \cdot X_r \quad (1)$$

The general question is, how to find an optimal W_s that enhances the signal while reducing the noise.

There are basic spatial filters like common average referencing (CAR) or Laplacian spatial filters [3], which can be applied for any type of EEG signal and whithout any training process. There are also more sophisticated, data-driven methods for the creation of spatial filters like common spatial patterns (CSP) [4], whitening [5], xDAWN [6] or canonical correlation analysis (CCA) [7], which are optimized on a specific dataset and therefore need data to be trained. Depending on the type of BCI, different spatial filtering methods can be applied. For BCIs in which classification is done in the frequency domain, e.g., motor-imagery BCIs, CSP can be used to improve the SNR of selected oscillations. If classification is done in the time domain, to detect evoked- or event related potentials like in the popular P300 speller [8], methods like whitening, xDAWN or CCA can be used. It should be noted that CCA is also often used in SSVEP and c-VEP BCIs, where it is used as a method for combined spatial filtering and classification [9, 10] or used solely for spatial filtering in combination with a different method for classification [11].

In the course of this paper, only spatial filter for time-domain classification will be considered. Unsurprisingly, data-driven spatial filter work better than basic spatial filters [7], but a clear comparison of the three methods is missing. In [7] whitening and CCA were compared on five different datasets with CCA yielding the better results on average, although whitening performed exactly the same on some datasets. Roy and colleagues [12] found that CCA performed slightly better than xDAWN in a test on workload EEG data, but the difference was not significant. Iwane and colleagues [13] compared CCA and xDAWN of data containing error-related potentials, and also showed CCA to have better results, but again, the difference was not significant.

As an alternative to the previously mentioned methods, this paper describes how spatial filtering can be seen as a regression problem and how arbitrary regression methods can be used to construct spatial filters. As all previously used spatial filtering methods create linear filters, it is of special interest that the use of regression methods also allows the construction of non-linear spatial filters.

METHODS

In this section, it is explained first, how Canonical Correlation Analysis (CCA) can be used for spatial filtering.

 DOI: 10.3217/978-3-85125-533-1-84

Based on this method, it is shown how spatial filtering can be seen as regression problem and how regression methods can be used to design spatial filters. At last, different spatial filtering methods are evaluated on an EEG dataset containing error-related potentials (ErrPs).

CCA for spatial filtering

CCA is a multivariate statistical method developed by H. Hotelling [14]. When having two datasets, which may have some underlying correlations, CCA can be used to find linear transformations for these two datasets, which maximize the correlation between the transformed datasets. Assuming there are two multidimensional datasets X and Y and their transformed datasets $x = W_x^T X$ and $y = W_y^T Y$, CCA can be used to find the two transformations W_x and W_y, which maximize the correlation between x and y by solving

$$\max_{W_x, W_y} \rho(x, y) = \frac{W_x^T X Y^T W_y}{\sqrt{W_x^T X X^T W_x \cdot W_y^T Y Y^T W_y}} \quad (2)$$

The process of using CCA for spatial filtering was previously described in [7]. To use CCA for spatial filtering, one needs to make a distinction between one-class problems and two-class problems, because the process of creating a spatial filter is slightly different in both cases.

For one-class problems (e.g. c-VEPs or SSVEPs), the classification is based on properties of the potential, like the time delay (c-VEP) or the frequency (SSVEP). For two-class problems (e.g. P300 or ErrP), the presence of such a potential is classified, if such a potential is found or not.

As signal-to-noise ratio (SNR) of single-trial EEG data is usually low, a common method to improve SNR is to average over multiple trials. The idea behind using CCA for spatial filtering is to find a linear transformation that maximizes the correlation between the recorded signal and the average evoked response, thereby improving the SNR of the transformed signal on a single-trial basis.

For the application of CCA, X is the raw EEG data and Y is the waveform of the average evoked response. CCA is then applied to find W_x and W_y, with W_x being used as spatial filter.

In the case of a one-class problem, we have k trials with EEG data, each consisting of a $n \times m$ matrix with n being the number of channels and m being the number of samples. For the application of CCA, all trials are concatenated to a new matrix X with new dimensions $n \times (k \cdot m)$. To obtain Y, first the average waveform of the evoked potential R is generated by averaging over all k trials, then R is replicated k times, to obtain a $n \times (k \cdot m)$ matrix $Y = [RR \ldots R]$. Since R does not necessarily has to contain all n channels, also a subset of $n_s \leq n$ channels can be used, so that Y has dimensions $n_s \times (k \cdot m)$. Regardless of the channelsubset used in R and Y, respectively, all n channels should be used in X, since this achieved better performance in previous, unpublished offline experiments.

For two-class problems, CCA is used similarly. Assume we have the EEG data X_1 containing all trials without the evoked potential and X_2 containing all trials with the evoked potential. For X_1 and X_2, Y_1 and Y_2 are obtained in the same way as for a one-class problem. Then X and Y are generated by concatenating $X = [X_1 X_2]$ and $Y = [Y_1 Y_2]$ and CCA is applied on X and Y to find W_X, which can be used as a spatial filter.

Regression for spatial filtering

A regression tries to predict a variable y_i based on a vector x_i, with x_i having n dimensions. In the case of a least-squares regression, the squared difference between the actual variable y_i and the prediction $\hat{y}_i$ is minimized

$$\min_w \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \quad (3)$$

For an optimal prediction, the goal is to find a set of weights w which minimize the above equation.

$$\hat{y}_i = \sum_{j=1}^{n} x_{ij} w_j \quad (4)$$

Regarding the use of regression for spatial filtering, it should be noted that the raw EEG signal consists of the ERP signal plus a lot of noise. A good spatial filter transforms the raw EEG signal in a way that the noise is reduced while keeping the ERP signal. As the averaged EEG signal contains the (nearly) noise-free ERP signal, we want to find a transformation, so that the transformed signal is very similar to the noise-free ERP signal. Using the notation of the regression described above, we want to find a set of weights w, which minimizes the difference between the noise-free ERP signal y and the spatially filtered EEG signal $\hat{y}$.

When applying a regression to find a spatial filter matrix W, the first step is the same step as for CCA, where X is created as a concatenation of the single-trial EEG data and Y is the concatenation of the (noise-free) averaged potentials, with Y_c being a vector containing a concatenation of the averaged potential at EEG channel c and X_c being the concatenation of the raw signal at channel c.

After that, a regression method is used for each channel c to find a transformation w_c that minimizes the distance between the spatially filtered signal $\hat{Y}_c$ and the average potential Y_c.

$$\min_{w_c} ||Y_c - \hat{Y}_c|| \quad (5)$$

$$\min_{w_c} ||Y_c - w_c X|| \quad (6)$$

By concatenating the w_c of all channels, a quadratic filter matrix W can be obtained, which can be multiplied with the raw EEG signal to obtain a spatially filtered signal. Essentially, arbitrary regression methods can be used to find the spatial filter weights w_c for each channel c. As the above formulation only considers linear regression methods, it is important to note, that also non-linear

 DOI: 10.3217/978-3-85125-533-1-84

methods can be used in which w_c is not a vector, but a function that is optimized.

$$\min_{w_c} ||Y_c - w_c(X)|| \quad (7)$$

Thereby also kernel methods or deep learning methods could be applied to find an optimal spatial filter function.

Evaluation on EEG dataset

To test the spatial filtering methods, we used data collected in a previous study [15], which contained error-related potentials (ErrPs). The subjects had to use a P300 speller [8] and if the BCI detected the wrong letter, the user should recognize the error and an ErrP should be elicited by the erroneous feedback. By detecting the ErrP, the wrong letter could be deleted and thereby the detection of ErrPs serves as an error correction system. EEG was recorded from electrodes F3, Fz, F4, T7, C3, Cz, C4,T8, CP3, CP4, P3, Pz, P4, PO7, PO8, Oz with a g.USBamp amplifier (an internal 0.5-30 Hz order eight Chebyschev bandpass filter was active) and digitised at 256 Hz. Ground and reference electrodes were placed at the left and right mastoid, respectively. We kept the impedance of all electrodes below 10 kΩ , in most cases below 5 kΩ. Impedance was measured before and after every session.

The 23 participating subjects were split into 3 different groups. H1 was drawn from the student population (N = 9, four female, mean age = 24.6 (SD ± 2.3), range $20-28$), all right-handed). H2 comprised a second group of elderly subjects age-matched to the group of participants with motor impairment(N = 8, two female, mean age = 45 (SD ± 5.2), range $39-52$). Group A2 (N = 6, one female, mean age = 51.2, SD ± 10.2, range $36-63$) includes 5 individuals diagnosed with ALS and one individual with Duchenne muscular dystrophy (participant A2u).

To evaluate the benefit of the different spatial filtering methods, we used the ErrP data from the above mentioned study, which consisted of 2 sessions per subject. To simulated the online case, we used the same data for training and testing the classifier as was used online. The training data consisted on average of 294 trials per subject (SD ± 45), while the test data consisted on average of 217 trials per subject (SD ± 78). After the display of the letter (at t=0 ms), the interval t=100-800ms was used as input for classification. After spatial filtering the raw EEG data, the data was bandpass filtered in the range of 0.5-16 Hz (by fast Fourier transform (FFT), removal of unwanted frequency bands, followed by inverse FFT). Subsequently the data was downsampled to 32 Hz. Thereafter, linear trends were removed from the EEG data and the data was scaled by centering and mapping the absolute maximum value to ± 1. All 16 channels were used as input for classification. As classifier we used a Support Vector Machine (SVM) with the LibSVM [16] implementation (RBF-Kernel with default parameters $\gamma = 1/(2\sigma)$ and $C = 1$). Due to the imbalanced classes (more correct trials than erroneus ones), we used a weighted SVM [17] with $w_{-1} = 0.3$.

Table 1: Classification accuracies on the ErrP dataset using different methods for spatial filtering: no spatial filtering (none), canonical correlation analysis (CCA), ridge regression (RR), linear support vector regression (lSVR) and support vector regression with an RBF-kernel (rSVR)

Subj.	none	CCA	RR	lSVR	rSVR
H1a	79.2 %	84.0 %	82.2 %	79.5 %	78.9 %
H1b	81.8 %	89.5 %	87.7 %	86.8 %	81.8 %
H1c	82.8 %	93.6 %	91.7 %	88.7 %	84.4 %
H1d	64.3 %	75.9 %	73.8 %	69.7 %	67.0 %
H1e	66.7 %	79.3 %	78.4 %	78.7 %	65.8 %
H1f	77.6 %	87.1 %	79.7 %	87.0 %	74.6 %
H1g	77.1 %	87.0 %	82.3 %	86.7 %	74.0 %
H1h	65.0 %	80.8 %	80.3 %	73.4 %	61.1 %
H1i	62.4 %	81.0 %	77.0 %	73.0 %	71.2 %
mean	**73.0 %**	**84.2 %**	**81.5 %**	**80.5 %**	**73.2 %**
H2j	76.4 %	100 %	100 %	100 %	100 %
H2k	60.4 %	68.9 %	59.5 %	60.4 %	59.5 %
H2l	93.4 %	93.4 %	92.9 %	82.5 %	81.5 %
H2m	75.6 %	100 %	100 %	100 %	100 %
H2n	81.6 %	84.0 %	84.4 %	86.0 %	80.8 %
H2o	78.0 %	87.0 %	85.3 %	79.1 %	72.9 %
H2p	62.1 %	80.3 %	76.8 %	74.2 %	59.6 %
H2q	79.6 %	84.1 %	76.6 %	76.6 %	76.6 %
mean	**75.9 %**	**87.2 %**	**84.4 %**	**82.3 %**	**78.9 %**
A2s	63.8 %	81.5 %	82.3 %	66.9 %	61.5 %
A2t	80.0 %	92.0 %	91.5 %	93.0 %	84.5 %
A2u	76.6 %	87.3 %	78.5 %	78.5 %	78.5 %
A2v	75.0 %	78.6 %	79.7 %	79.7 %	79.7 %
A2w	82.4 %	80.7 %	77.3 %	79.8 %	74.0 %
A2x	63.7 %	78.3 %	78.3 %	73.3 %	72.0 %
mean	**73.6 %**	**83.1 %**	**81.3 %**	**78.5 %**	**75.0 %**
mean	**74.2 %**	**85.0 %**	**82.5 %**	**80.6 %**	**75.6 %**

For the different spatial filter methods, we evaluated classification accuracy without any spatial filter, when using CCA for spatial filtering and when using three different regression methods. We usedthe MATLAB implementation of a ridge regression with a regularization parameter of $\lambda = 0.0001$ and a support vector regression with default paramters. To also test a non-linear regression, we evaluated the support vector regression with an RBF kernel using the LibSVM [16] implementation with default parameters.

RESULTS

The detailed results for the classification accuracy on the ErrP dataset with different spatial filtering methods can be seen in Table 1. While the average accuracy without spatial filtering is 74.2 %, it could be improved to 85.0 % by using CCA for spatial filtering, which is significantly better ($p < 0.001$, Wilcoxon ranksum test). Using ridge regression for the creation of a spatial filter resulted in an average accuracy of 82.5 %, which is not significantly lower than CCA ($p > 0.05$). Using support vector regression for spatial filter creation results in an average accu-

 DOI: 10.3217/978-3-85125-533-1-84

racy of 80.6 % when using a linear kernel and 75.6 % with an RBF kernel. Results with linear kernel are not signficantly different to CCA($p > 0.05$), but results with RBF kernel are significantly worse ($p < 0.005$).

DISCUSSION AND CONCLUSION

In this paper, it was described how spatial filtering of EEG can be seen as regression problem and how arbitrary regression methods can be used for the construction of spatial filters. Three different regression methods were tested and compared to CCA on an EEG dataset containing error-related potentials. Classification accuracy was highest when using CCA for the construction of spatial filters, but performance with linear regression methods was not significantly worse. Using a non-linear support vector regression with an RBF-kernel resulted in significantly lower performance.

Based on the presented results it should be discussed what the benefits of using a regression method for spatial filtering are, or if there are any at all. Although performance difference to CCA was not significant, the results give a hint that when in doubt, better use CCA. Also from a theoretical standpoint, CCA seems to be better suited. As CCA uses two transformation matrices W_x and W_y, W_x is used as spatial filter and W_y transforms the averaged potential to a subspace containing different ERP components. With this last step, CCA bears similarity to principal component analysis (PCA). The spatial filter generated by CCA thereby does not try to increase the SNR on EEG sensor level, but separates the average ERP into (uncorrelated) components and improve the SNR for those components. On the other hand, regression tries to increase SNR on EEG sensor level. As neighboring sensors are correlated, regression-based spatial filters deliver some redundant information and thereby the spatial filter created by CCA might be better for classification as components are uncorrelated and thereby contain less redundant information.

The most interesting thing about using regression methods for spatial filtering is the possibility to use non-linear methods. So far, all spatial filtering methods used in EEG signal processing are linear methods. Being able to use arbitrary regression methods for spatial filtering means that also kernel methods or artificial neural networks and deep learning can be used for the creation of spatial filters. But why should non-linear spatial filters be superior to linear filters, as the results in this paper rather point in the other direction? The signal recorded at the EEG sensors is generally considered to be a linear mixture of electrical sources in the brain and artefactual/noise sources [18]. As spatial filters are trying to eliminate noise sources, it is basically a reversal of this mixture process and if the mixture is a linear process, a linear spatial filter should be able to yield optimal results. However, this is only true under certain assumptions: that all sources are stationary and that there are equal or less sources than we have channels. If a source is moving, the influence of the source on the sensors depends non-linearly on its position and therefore non-linear filters might be better to remove those sources. If there are more sources than sensors (and assuming some independence between the sources) the sources can not be perfectly reconstructed and hence, non-linear methods might achieve better results in reconstructing and removing these sources. So, it depends on the assumptions one makes about EEG if non-linear spatial filter can provide better results than linear filters.

A further argument that questions the use of non-linear spatial filters (or spatial filtering in general) is that classifiers can also integrate spatial filtering. Assuming an optimal spatial filter function $s(x)$, the raw EEG data x_r and a classification method that always finds an optimal classifier. If this method is trained on the spatially filtered data $x_s = s(x_r)$ it would return a function $g(x)$, so that $g(x_s)$ is the optimal classification result. But as the classification method always finds the optimal classifier, it would return the function $f(x) = g(s(x))$ if it is trained on the raw EEG data. Thereby, if one has a classification method that always gives the optimal classifier, spatial filtering is obsolete and a non-linear classifier would be able to also learn a non-linear spatial filtering. However, this is a rather theoretical remark. As this and previous papers [5, 6, 7, 12, 13] have shown, for classifiers commonly used in BCI applications spatial filtering always improves results. It should also be noted that an optimal classifier is only able to learn spatial filtering when trained on the raw EEG data, i.e. time-domain features. If there is a feature extraction step, like power spectrum estimation, an optimal classifier can not learn the spatial filtering anymore. While SSVEP is a good example where evoked potentials are often classified in the frequency domain, a classification of event-related potentials in the frequency domain can also be used if there is no clear stimulus onset, as it was shown for such asynchronous classification that ErrPs [19] and P300s [20] can be reliably detected based on power spectral features. In these cases a spatial filter could be trained on ERP data and then applied before power spectral estimation.

Coming back to the question if non-linear spatial filtering can improve results compared to linear spatial filtering, the results presented in this paper should be seen merely as a proof-of-concept to demonstrate that non-linear spatial filtering is possible. In future work, different non-linear methods like neural networks should be tested to evaluate if non-linear spatial filtering can improve results compared to what linear spatial filters can offer. As linear regression did not provide better results than CCA, CCA is still being recommended for the creation of spatial filters as it is easy to use and already implemented in all major frameworks like R, Python or MATLAB.

ACKNOWLEDGMENTS

This work was supported by the *Deutsche Forschungsgemeinschaft* (DFG; grant SP 1533/2-1).

 DOI: 10.3217/978-3-85125-533-1-84

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, and et al. Brain-computer interface technology: A review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, 8:164–173, 2000.

[2] A. Kübler, F. Nijboer, J. Mellinger, T.M. Vaughan, H. Pawelzik, G. Schalk, D.J. McFarland, N. Birbaumer, and J.R. Wolpaw. Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64(10):1775–1777, 2005.

[3] D J McFarland, L M McCane, S V David, and J R Wolpaw. Spatial filter selection for EEG-based communication. *Electroencephalography and clinical neurophysiology*, 103(3):386–94, September 1997.

[4] Y. Wang, S. Gao, and X. Gao. Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 5392–5395. IEEE, 2006.

[5] J. Farquhar and N. J. Hill. Interactions between pre-processing and classification methods for event-related-potential classification. *Neuroinformatics*, 11:175–192, 2013.

[6] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert. xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. *IEEE transactions on bio-medical engineering*, 56(8):2035–43, August 2009.

[7] M. Spüler, A. Walter, W. Rosenstiel, and M. Bogdan. Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in eeg data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(6):1097–1103, 2014.

[8] L. A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol*, 70(6):510–523, December 1988.

[9] G. Bin, X. Gao, Z. Yan, B. Hong, and Sh. Gao. An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. *Journal of Neural Engineering*, 6(4):046002, 2009.

[10] G. Bin, X. Gao, Y. Wang, Y. Li, B. Hong, and S. Gao. A high-speed BCI based on code modulation VEP. *Journal of Neural Engineering*, 8(2):025015, 2011.

[11] M. Spüler, W. Rosenstiel, and M. Bogdan. Online adaptation of a c-VEP brain-computer interface (BCI) based on error-related potentials and unsupervised learning. *PloS one*, 7(12):e51077, 2012.

[12] R. Roy, S. Bonnet, S. Charbonnier, P. Jallon, and A. Campagne. A comparison of erp spatial filtering methods for optimal mental workload estimation. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 7254–7257. IEEE, 2015.

[13] F. Iwane, R. Chavarriaga Lozano, I. Iturrate, and J. Millán. Spatial filters yield stable features for error-related potentials across conditions. In *2016 IEEE International Conference on Systems, Man, and Cybernetics*, number EPFL-CONF-223780, 2016.

[14] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.

[15] M. Spüler, M. Bensch, S. Kleih, W. Rosenstiel, M. Bogdan, and A. Kübler. Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI. *Clinical Neurophysiology*, 123(7):1328–1337, 07 2012.

[16] C. Chang and C. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[17] Y. Tang, Y. Zhang, N. Chawla, and S. Krasser. SVMs modeling for highly imbalanced classification. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(1):281–288, 2009.

[18] A. Delorme, J. Palmer, J. Onton, R. Oostenveld, and S. Makeig. Independent eeg sources are dipolar. *PloS one*, 7(2):e30135, 2012.

[19] M. Spüler and C. Niethammer. Error-related potentials during continuous feedback: using eeg to detect errors of different type and severity. *Frontiers in human neuroscience*, 9:155, 2015.

[20] T. Krumpe, C. Walter, W. Rosenstiel, and M. Spüler. Asynchronous p300 classification in a reactive brain-computer interface during an outlier detection task. *Journal of neural engineering*, 13(4):046015, 2016.