ON TRANSFERRING SPATIAL FILTERS IN A BRAIN READING SCENARIO

Jan Hendrik Metzen[1], *Su Kyoung Kim*[1,2], *Timo Duchrow*[2], *Elsa Andrea Kirchner*[1,2], *Frank Kirchner*[1,2]

[1]University Bremen, Robotics Group, Robert-Hooke-Str. 5, 28359 Bremen, Germany
[2]DFKI - Robotics Innovation Center, Robert-Hooke-Str. 5, 28359 Bremen, Germany

ABSTRACT

Machine learning approaches are increasingly used in brain-machine-interfaces to allow the automatic adaptation to user-specific brain patterns. One of the most crucial factors for the practical success of these systems is that this adaptation can be achieved with a minimum amount of training data since training data needs to be recorded during a calibration procedure prior to the actual usage session. To this end, one promising approach is to reuse models based on data recorded in preceding sessions of the same or other users. In this paper, we investigate under which conditions it is favorable to reuse models (more specifically spatial filters) trained on data from historic sessions compared to learning new spatial filters on the current session's calibration data. We present an empirical study in a scenario in which Brain Reading, a particular kind of brain-machine-interface, is used to support robotic telemanipulation.

***Index Terms*—** Brain Reading, spatial filter, model transfer

1. INTRODUCTION

In many scenarios that involve the use of man-machine interfaces, the usability of the system can be improved if the machine is provided with some information about the current state and intent of its user such that the machine can optimize its behavior accordingly. Analyzing the user's electroencephalogram (EEG) is one way to obtain this information since event-related potentials and certain changes in brain wave frequency bands are known to be related to the changes of the user's mental state and intent. Decoding of a user's mental state and intent based on detecting these patterns by external observation of brain activity is denoted as *Brain Reading (BR)*. In contrast to most Brain Computer Interfaces (BCIs, see [1, 2] for a review of works) no active participation of the user is required.

BR systems must be adapted to the current brain patterns of the respective user since these characteristic patterns vary between different subjects and even change over time within the same subject. One option to achieve this kind of adaptability is to use supervised machine learning (ML) to learn user-specific models (see e.g. [3]). In order to apply supervised ML techniques, a training set with labeled examples is needed. A common approach for acquiring such a training set is to perform a calibration procedure at the beginning of each session, in which EEG data is recorded from the user who acts in a controlled and supervised scenario. This means that the user has to perform a time-consuming and potentially exhausting calibration procedure each time he wants to use the system. Thus, it is desirable to keep this calibration procedure as short as possible, i.e. to use a system that requires only a small training set.

One way to reduce the required amount of training data is to reuse data recorded in previous sessions conducted by the same or by other subjects in the same scenario. This approach necessitates dealing with inter-session and, potentially, also with inter-subject variability. Two different approaches can be distinguished: In the first approach, historic models are reused directly (see e.g. [4, 5]) while in the second approach historic data is mixed with data from the calibration procedure to train the adaptable components generating hybrid models (see e.g. [6]).

In this work, we will focus on the first approach, more specifically we investigate the transferability of so-called spatial filters. Spatial filtering denotes a mapping of the original channels (that correspond one-to-one to the electrodes) onto new pseudo-channels that are a mixture of the signals recorded at different electrodes. It allows to reduce the dimensionality of the problem which is important in situations that require fast data processing but only limited computing power is available, such as e.g. in "portable" BR where the data processing has to be performed on an embedded device. The goal of this work is to *investigate empirically under which conditions it is favorable to reuse spatial filters trained on data from historic sessions compared to learning new spatial filters on the current session's calibration data*. Both approaches are compared for different sizes of the training set and different degrees of dimensionality reduction.

2. LABYRINTH ODDBALL SCENARIO

The empirical evaluation was conducted on an EEG dataset recorded in the Labyrinth Oddball scenario (see Figure 1). The Labyrinth Oddball scenario is a testbed for the use of man-machine interfaces in robotic telemanipulation and is well suited for evaluation of BR (for more details we refer to Kirchner et al. [7]*). In this scenario, the task for the BR system is to discriminate between the EEG patterns associated with the successful cognitive processing of two specific kinds of visual stimuli presented to the user, called 'standard' and 'target' stimulus. While 'standard' stimuli are frequent ($n = 720$ presentations per run) but irrelevant for the user, 'target' stimuli are rare ($n = 120$) but require a reaction (pressing a buzzer) by the user. Such a scenario is called "oddball discrimination paradigm" and the successful cognitive processing of the rare 'target' stimuli is known to elicit a special kind of event-related potential, called P300 [8]. The visual presentation (shape and color) of standard and target stimuli was very similar in order to avoid differences in early visual processing and to make sure that differences in the EEG recorded after the presentation of both stimuli types are actually due to higher cognitive processing. In contrast to many BCIs (e.g. [2]), the classification has to be done based on the individual epoch (instance) and not on an average over several repetitions of the same condition. Note that no feedback sessions for subject training were performed.

The dataset used in this paper consists of the labeled EEG data

This work was supported by the Federal Republic of Germany through a grant of the Federal Ministry of Education and Research (BMBF, FKZ 01IW07003) and a grant of the Federal Ministry of Economics and Technology (BMWi, FKZ 50 RA 1011).

* Note: Since the electronic proceedings of [7] are not yet available online, the paper may be obtained temporary (for reviewing purposes only) from `http://dl.dropbox.com/u/8318678/reference7.zip` (password: ssp2011).

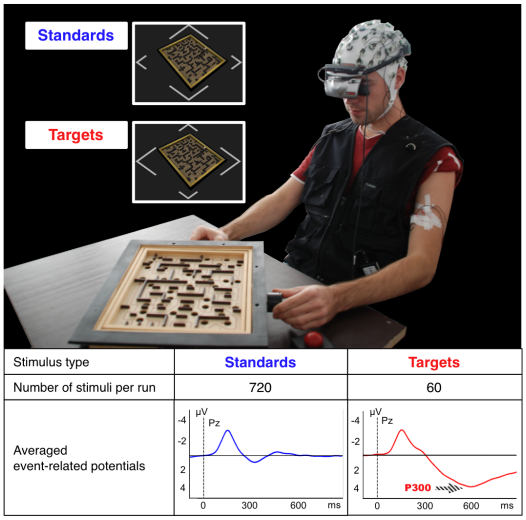

Fig. 1. Labyrinth Oddball scenario: Subject playing a physical simulation of the BRIO® labyrinth and responding to rare "target" stimuli by pressing a buzzer. Target stimuli were mixed with more frequent "standard" stimuli that differed only slightly in presentation.

recorded in ten sessions[1] from five (male) subjects; each subject performed two sessions and each session consisted of five repetitions (called "runs") of the Labyrinth Oddball paradigm. EEGs were recorded continuously from 64 electrodes (extended 10-20 system with reference at electrode FCz), using an actiCap system (Brain Products GmbH, Munich, Germany). Two of the 64 channels were used to record EMG (electromyography) signals of muscles of the upper arm. EEG and EMG signals were amplified by two 32 channel BrainAmp DC amplifiers (Brain Products GmbH, Munich, Germany) and sampled at $1000\,\mathrm{Hz}$. The impedance was kept below $5\mathrm{k}\Omega$ in order to minimize artifacts that are due to high impedance.

3. METHOD

3.1. Single-trial data processing

As a first step in the single-trial data processing system used for discrimination of the 'standard' and the 'target' condition, the EEG signal recorded during the experiment was split into distinct time windows. For each presented standard and target stimulus, one time window was extracted beginning $0\,\mathrm{ms}$ and ending $1000\,\mathrm{ms}$ after the stimulus presentation. Each of the resulting time windows was normalized so that the mean value of each channel within the window was 0. After this, the signal was low-pass filtered and downsampled from $1000\,\mathrm{Hz}$ to $25\,\mathrm{Hz}$ and band-pass filtered with a pass-band of $(0.0, 4.0)\,\mathrm{Hz}$ in order to focus on slow potentials like the P300.

Subsequently, the signal was spatially filtered (see Section 3.2) and the values of the resulting pseudo-channels were used directly as features, i.e. the 26 samples of each of the m retained channels that fell into the time window from 0 to $1000\,\mathrm{ms}$ were used as features, resulting in $26m$ features. Thereupon, each feature dimension was normalized such that its 2.5th percentile on the training data was mapped onto 0 and the 97.5th percentile was mapped onto 1. The resulting feature vectors were classified using a Support Vector Machine with linear kernel. The implementation of the data processing system is mainly based on the "Modular toolkit for Data Processing" [9], see also Kirchner et al. [7].

3.2. Spatial filtering

Spatial filtering denotes a mapping of the original n channels that correspond one-to-one to the n electrodes onto new pseudo-channels that are a mixture of the signals recorded at different electrodes. Spatial filtering aims at cumulating class-correlated EEG patterns that are superimposed by other, potentially stronger non-relevant components and spatially distributed over several electrodes into a single pseudo-channel. This allows to reduce the number of channels by retaining only the subset of the $m < n$ most promising pseudo-channels. This has two advantages: *first, dimensionality reduction reduces computing time for subsequent processing stage and secondly, classifier overfitting is less likely in lower dimensional spaces*.

Data-driven approaches based on unsupervised or supervised ML allow to learn subject- (and session-) specific spatial filters. While learning subject/session-specific filters is desirable since spatial localization of patterns might vary between subjects and also between different sessions of the same user (for example due to slightly different EEG cap placements), it also requires a sufficient amount of calibration data recorded in the current session. An alternative is to reuse spatial filters trained on historic sessions of the same or different users. While these filters have not been specifically adapted for the current session, selecting an adequate filter from a large ensemble of historic filters might still yield superior performance when only a small calibration data set is available.

In this work, we use the supervised common spatial patterns (CSP) algorithm (see for example Blankertz et al. [10]) for learning spatial filters. CSP maps the data onto axes such that the variance for instances of the first class is maximized and the variance for the second class is minimized (or vice versa). This is achieved by a simultaneous diagonalization of the two empirical intra-class covariance matrices $\Sigma_1 = n_1^{-1}\sum_{i=0}^{n_1}(x_i^{(1)} - \mu^{(1)})(x_i^{(1)} - \mu^{(1)})^T$ and $\Sigma_2 = n_2^{-1}\sum_{i=0}^{n_2}(x_i^{(2)} - \mu^{(2)})(x_i^{(2)} - \mu^{(2)})^T$, i.e. by solving $\Sigma_1 W = \Lambda \Sigma_2 W$ where Λ is the vector of generalized eigenvalues and W is the matrix of generalized eigenvectors corresponding to the learned projections.

4. EVALUATION

4.1. Setup

We compared 4 different settings for spatial filtering ($f \in \{$'NoFilter', 'FromScratch', 'LeaveOneSessionOut', 'InterSubject'$\}$): not using any spatial filtering ('NoFilter'), learning CSP filters based solely on the session's calibration data ('FromScratch'), selecting a CSP filter from a historic set including sessions from the same user ('LeaveOneSessionOut') and a historic set without any sessions of the current user ('InterSubject'). All four settings have been tested for different numbers of retained pseudo-channels ($m \in \{2, 4, 8, 16, 32, 62\}$) and different sizes of the calibration data set ($t \in \{42, 84, 168, 252, 420, 840\}$, where t is the number of instances and $t = 840$ corresponds to a calibration run of approx. 16 minutes). For the 'NoFilter' condition, domain experts have selected fixed electrode sets for the different values of m[2].

[1] Sessions have been recorded on different days and thus the positioning of the EEG cap can vary between sessions of the same user.

[2] $m = 2$: Cz, Pz; $m = 4$: Fz, Oz additionally; $m = 8$: T7, T8, Fp1, Fp2 additionally; $m = 16$: F3, F4, C3, C4, P3, P4, O1, O2 additionally; $m = 32$: standard 32 electrodes cap setting

Historic spatial filters were trained using the CSP algorithm on the EEG data of a whole session, resulting in $10 - 1 = 9$ CSP filter matrices for the 'LeaveOneSessionOut' setting and $10 - 2 = 8$ for the 'InterSubject' setting. Among these filter matrices, the most adequate one for the current session and the selected m was determined by five-fold cross-validation on the available calibration data[3]. In the same cross-validation loop, the complexity parameter $C \in \{0.01, 0.1\}$ of the SVM and the relative weight $w_t \in \{5, 10\}$ of the target-class to the standard-class were determined. Overweighting the target-class was necessary since the ratio of standard and target class instances in the data set was highly unbalanced (roughly $6 : 1$).

A subset of the data recorded during the first run of a session was used as calibration data while the remaining four runs were used as test data (intra-session setting). Note that the EMG channels were discarded prior to the data processing. The F_1-measure on the target class was used as performance metric to emphasize the role of the minority class. For the given ratio of $6 : 1$ between standard and target class, the trivial classifiers "always standard", "always target", and "uniform random guess" would obtain an F_1-measure on the target class of $F_1 = 0.0$, $F_1 \approx 0.25$, and $F_1 \approx 0.222$ respectively.

4.2. Results

Figure 2 summarizes the results of the study. The results were analyzed using repeated measures ANOVA with three within-subjects factors: transfer of spatial filter (f), training size (t), and number of retained channels ("dimensionality" m). If needed, the Greenhouse-Geisser correction was applied. For pairwise comparisons, Bonferroni correction was applied. Additionally, linear regression analyses were applied separately for each spatial setting.

In general, the more training data became available the better the performance of the system got [main effect of t: $F(5, 195) = 453.24$, $p < 0.001$]. The effect of dimensionality [main effect of m: $F(5, 195) = 76,617$, $p < 0.001$] was different depending on the spatial filter setting. In particular, the "NoFilter" condition was strongly affected by the dimensionality, namely by an improving performance with increasing m. This positive correlation was also present for small t [$r = 0.273$, $p < 0.001$ for $t = 42$]. In contrast, for the other settings a negative correlation (decreasing performance with increasing m) was observed for small t ["LeaveOneSessionOut": $r = 0.448$, $p < 0.001$, "FromScratch": $r = 0.204$, $p < 0.001$ (for $t = 42$)]. Accordingly, the highest performance for small t and large m was achieved by the "NoFilter" setting.

The session-specific filter ("FromScratch") was strongly influenced by t, but the effect of m was less pronounced compared to "NoFilter". For example, for small t, the "FromScratch" was the worst among the four settings regardless of the value of m [$p < 0.001$ for $t = 42$]. In contrast, for large values of t and medium values of m, the performance of the "FromScratch" setting was significantly the best among the four settings [$p < 0.003$ for $t = 840$ combined with $m = 8, 16$].

The two settings that are based on reusing historic spatial filters ("LeaveOneSessionOut" and "InterSubject") were least affected by both m and t [main effect of f: $F(3, 117) = 32,846$, $p < 0.001$]. Thus, for small m, these two "reusing" settings outperformed "NoFilter" and "FromScratch" for all values of t [$p < 0.026$ for $m = 2$]. However, for larger values of m, the performance of the "reusing" settings did not benefit as strong as the other two settings from increasing t. Accordingly, for $m = 4$ the "reusing" settings outperformed the other settings only for small t [$p < 0.017$ for $m < 4$ combined with $t < 252$] but not for larger values of t.

[3]Note that cross-validation only affects the runtime of the method during training but not during prediction and thus does not counteract the reduction of runtime achieved by dimensionality reduction.

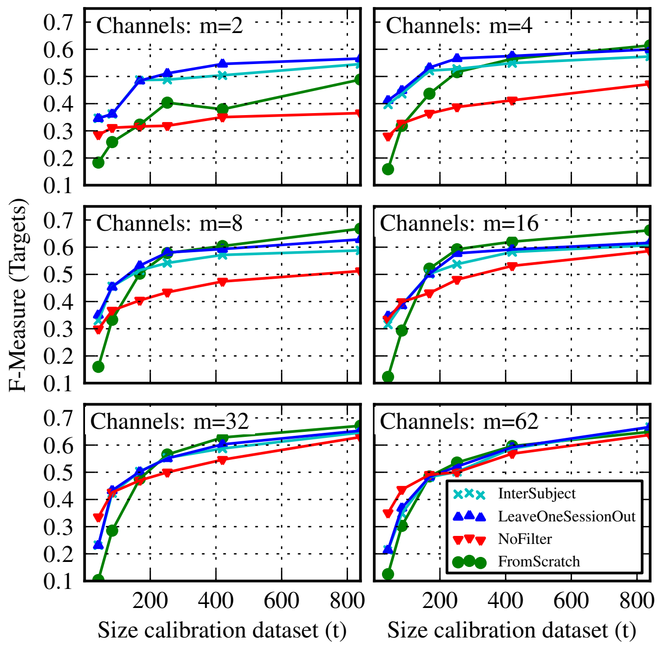

Fig. 2. Average F_1-measure on target class for the four spatial filtering settings for different sizes of the training set and different numbers of retained channels after spatial filtering ("dimensionality").

4.3. Discussion

As expected, making available more calibration data improves the performance of the system. In all settings, a better adapted classifier is one explanation for this increased performance. In the "FromScratch" setting, better adaptation of the spatial filter to the specific subject/session can also contribute to the increased performance, while in the "reusing" settings, increased performance might be attributed to a more robust selection of the historic spatial filter that fits optimally to the current session. The significantly better performance of the "reusing" settings compared to the "FromScratch" setting for small values of m and t shows that it is easier to select an appropriate historic filter than to learn from scratch a new filter that cumulates the relevant information into the most prominent pseudochannels.

Figure 3 summarizes in which situations which kind of approach can be recommended: In situations when the calibration run must be short and the dimensionality has to be reduced ($m \leq 8$) for computational reasons, reusing historic filters is recommended. In particular, if only $m = 2$ channels can be retained, reusing spatial filters is significantly the best approach regardless of t. If the calibration run must be short but a larger number of channels can be retained ($m \geq 32$), spatial filtering should be discarded altogether. A possible explanation for the harmfulness of any kind of spatial filtering under this condition is that most of the pseudochannels do not contain any relevant information and the number of "noise" features is thus very large. If the training set is very small, this may result in severe overfitting. When a longer calibration run is possible, learning a new spatial filter tends to be the best approach even though reusing

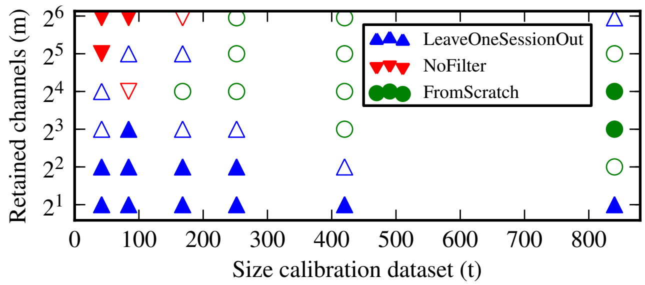

Fig. 3. The figure shows which method achieves the highest F_1-measure for a given combination of retained channels and size of calibration dataset. Filled symbols denote settings where the "winner" method is significantly better than each of the other two methods. The "InterSubject" setting is omitted since it is mostly on par with the "LeaveOneSessionOut" setting.

historic filters is usually only slightly and not significantly worse.

Under a few conditions, the "LeaveOneSessionOut" setting is slightly superior compared to the "InterSubject" setting but largely, the two different types of spatial filter transfer are on par. This shows that spatial filter transfer is also feasible for novel users (corresponding to the "InterSubject" setting) while experienced users might have a small advantage under certain conditions.

We would like to point out that the results for the "NoFilter" condition are specific for the respective electrode selections and other electrode selections might perform better for a given m. However, this only indicates that choosing a good electrode set is a non-trivial problem. In contrast, spatial filtering allows for an automatic selection of pseudo-channels based on their corresponding eigenvalues.

5. RELATED WORKS

In this section, we classify the proposed method into the related work. Related work has been done mainly in the field of BCIs; thus a direct comparison between our and related work in terms of achieved performance and duration of calibration procedures is not possible. Instead, we focus on a delineation of concepts and preconditions of the individual methods.

A method targeted at long-term BCI users was proposed by Krauledat et al. [4]: for each historic session conducted by the user and each class, the three top CSP filters were stored. A clustering in the CSP space was used to select six prototypical CSP filters from this historic CSP set. Furthermore, the data from the historic sessions was pooled and six CSP filters were learned based on this data. The main difference to our work is that single CSP filters were transferred and not whole filter matrices and that these filters were selected using unsupervised learning. Furthermore, the evaluation of the approach was restricted to intra-subject transfer and thus only long-term BCI users were addressed.

A method that is also feasible for novel subjects was proposed by Lotte and Guan [6]: In order to reduce the calibration time for novel subjects, EEG data recorded from known subjects was reused for the adaptation to a novel subject. The BCI system discussed by the authors was based on CSP and linear discriminant analysis. Both methods require to estimate covariance matrices; Lotte and Guan proposed to compute the covariance matrices as a mixture of the estimates obtained from data recorded during the calibration procedure and of the data recorded from other subjects. They gave a procedure for selecting a subset of the subjects whose data can be well transferred onto the current subject and a procedure for determining the mixture coefficients. The main differences to our work are that not models but data was transferred and that the effect of varying the number of retained CSP channels was not investigated.

6. CONCLUSION

We have presented an empirical study in a Brain Reading scenario that indicates that spatial filtering is recommendable for single-trial discrimination of classes in an oddball-like scenario. Furthermore, we have shown that reusing spatial filters trained on historic sessions achieves superior results compared to learning a session-specific filter anew in situations when the training set is small or when the number of channels must be reduced. This holds true also for novel users for which spatial filters of other users need to be reused.

7. REFERENCES

[1] N. Birbaumer, "Breaking the silence: brain-computer interfaces (BCI) for communication and motor control," *Psychophysiology*, vol. 43, no. 6, pp. 517–532, Nov. 2006.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, June 2002.

[3] B. Blankertz, G. Dornhege, S. Lemm, M. Krauledat, G. Curio, and K.-R. Müller, "The Berlin Brain-Computer Interface: Machine learning based detection of user specific brain states," *Journal of Universal Computer Sciene*, vol. 12, no. 6, pp. 581–607, 2006.

[4] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards zero training for Brain-Computer interfacing," *PLoS ONE*, vol. 3, no. 8, pp. e2967, 2008.

[5] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K.-R. Müller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural Networks*, vol. 22, no. 9, pp. 1305–1312, Nov. 2009.

[6] F. Lotte and C. Guan, "Learning from other subjects helps reducing Brain-Computer interface calibration time," in *Proceedings of the International Conference on Audio Speech and Signal Processing (ICASSP)*, 2010.

[7] E. A. Kirchner, H. Wöhrle, C. Bergatt, S. K. Kim, J. H. Metzen, D. Feess, and F. Kirchner, "Towards operator monitoring via brain reading - an EEG-based approach for space applications," in *Proceedings of the 10th International Symposium on Artificial Intelligence, Robotics and Automation in Space (iSAIRAS)*, Sept. 2010.

[8] N. K. Squires, K. C. Squires, and S. A. Hillyard, "Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli," *Electroencephalography and Clinical Neurophysiology*, vol. 38, no. 4, pp. 387–401, April 1975.

[9] T. Zito, N. Wilbert, L. Wiskott, and P. Berkes, "Modular toolkit for data processing (MDP): a python data processing framework," *Frontiers in Neuroinformatics*, vol. 2, pp. 8, 2008.

[10] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG Single-Trial analysis," *Signal Processing Magazine, IEEE*, vol. 25, no. 1, pp. 41–56, 2008.