HAL open science

Learning from M/EEG data with variable brain activation delays

Wojciech Zaremba, M. Pawan Kumar, Alexandre Gramfort, Matthew Blaschko

To cite this version:

Wojciech Zaremba, M. Pawan Kumar, Alexandre Gramfort, Matthew Blaschko. Learning from M/EEG data with variable brain activation delays. International Conference on Information Processing in Medical Imaging, Jun 2013, Asilomar, United States. hal-00803981v2

HAL Id: hal-00803981

https://inria.hal.science/hal-00803981v2

Submitted on 25 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning from M/EEG data with variable brain activation delays

Wojciech Zaremba[1,2], M. Pawan Kumar[1,2,3], Alexandre Gramfort[4,5], and Matthew B. Blaschko[1,2,3]

[1] Center for Visual Computing, École Centrale Paris, Châtenay-Malabry France
[2] INRIA Saclay–Île-de-France
[3] Université Paris-Est, LIGM (UMR CNRS), École des Ponts ParisTech, France
[4] Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, Paris, France
[5] CEA/Neurospin bât 145, 91191 Gif-Sur-Yvette

Abstract. Magneto- and electroencephalography (M/EEG) measure the electromagnetic signals produced by brain activity. In order to address the issue of limited signal-to-noise ratio (SNR) with raw data, acquisitions consist of multiple repetitions of the same experiment. An important challenge arising from such data is the variability of brain activations over the repetitions. It hinders statistical analysis such as prediction performance in a supervised learning setup. One such confounding variability is the time offset of the peak of the activation, which varies across repetitions. We propose to address this misalignment issue by explicitly modeling time shifts of different brain responses in a classification setup. To this end, we use the latent support vector machine (LSVM) formulation, where the latent shifts are inferred while learning the classifier parameters. The inferred shifts are further used to improve the SNR of the M/EEG data, and to infer the chronometry and the sequence of activations across the brain regions that are involved in the experimental task. Results are validated on a long term memory retrieval task, showing significant improvement using the proposed latent discriminative method.

Keywords: magnetoencephalography (MEG), electroencephalograpy (EEG), Latent SVM, classification, independant component analysis (ICA), functional connectivity, single-trial variability

1 Introduction

Magnetoencephalography and electroencephalography (M/EEG) measure the electromagnetic fields induced by brain activity. Typically, when collecting M/EEG data in neurosciences, the same task is repeated several times, resulting in hundreds of trials. Given such data, a classical way to distinguish between two tasks (also called experimental conditions) is to average all the trials for each condition and compare the difference between the averages. The main issue with such an approach is that the latency and amplitudes of the responses of each individual activated brain region can vary across the trials. For example

the measured P300 wave, often used for brain computer interface (BCI) systems, is a mix of P3a and P3b waves which are almost concomitant with the P2 wave [16]. Each wave can suffer from different variabilities. The reasons for such variabilities are many fold: fatigue, habituation or changes in attention to name but a few. This makes the process of averaging prone to modeling errors. An alternative approach is to cast the statistical test of distinguishing two tasks as a classification problem. This is similar to a BCI system, which predicts a behavioral variable from raw M/EEG recordings. When using such a supervised learning approach, repeated trials increase the amount of training data, which could in theory lead to better prediction. However, even in this setting, the prediction accuracy is inevitably affected by the variability between trials.

The above argument suggests that it is important to explicitly model the variabilities in brain responses in order to improve classification accuracy. To this end, we propose to use a supervised learning algorithm with latent variables. Specifically, we introduce latent variables for each trial, which represent its variability. This allows us to learn a classifier using the latent support vector machine (LSVM) framework, which iteratively estimates the value of the latent variables such that the training error is reduced. Our experiments show that this approach can provide a significant improvement in the prediction accuracy over a baseline method that does not explicitly model the sample transformation (Section 5.1). Moreover, the imputed latent variables allows us to improve the quality of the brain sources visualization (Section 5.2). Finally, as explained in Section 5.3, the latencies of the brain source responses offers the possibility to investigate the chronometry in functional networks at a millisecond time scale. Code of this implementation is available online. [6]

2 Related work

This work explores use of latent support vector machines (LSVM) in M/EEG studies to improve prediction and discover brain functional connectivity. The problem of prediction using M/EEG signals has been extensively studied in the context of mind reading [6]. Recent works in this field mostly use classifiers like SVM or LDA, which cannot explicitly model the variability over trials. To overcome this deficiency, we employ the latent support vector machine (LSVM) classifier, whose ability to handle latent variables has been successfully exploited in other fields of research such as bioinformatics [20] and computer vision [9]. Note that, in contrast to previous latent models used in brain imaging that are purely unsupervised [11], LSVM is a supervised learning approach (that is, it makes use of the knowledge of the experimental conditions for various trials while estimating the latent variables).

The topic of brain functional connectivity has also received considerable attention in the literature [5]. Recent works in this field mostly use non-stationary time-frequency transforms for feature selection [4]. In contrast to previous work,

[6] https://github.com/wojzaremba/active-delays

our features are based on activation peak misalignment (where the misalignments can be estimated using any set of features using our LSVM formulation). While the latency in brain signals has been studied since at least the late 1960s [19], to the best of our knowledge, it has not been previously considered in the context of brain functional connectivity.

3 Latent SVM for M/EEG data

In this section, we will describe how the latent support vector machine (LSVM) [20] framework can be used to classify M/EEG data in the presence of significant variability among trials. Furthermore, we will describe how LSVM can be used to improve the visualization of brain sources and to estimate the brain functional connectivity. We begin with a brief description of the general LSVM framework.

LSVM is an extension of the well-known support vector machine (SVM) [8] classifier, which allows for missing information in the training samples. Formally, let $x \in \mathcal{X}$ denote an input that needs to be assigned a classification label $y \in \mathcal{Y} \equiv \{-1, +1\}$. In the present case corresponds to one or multiple time series. The latent variable $h \in \mathcal{H}$ represents any missing information that can aid the classification process. Note that, by definition, the value of the latent variable is unknown while the domain, $\mathcal{H}$, is a modeling choice. We represent the joint feature vector of the input x and the latent variable h by $\phi(x, h)$. Given a training dataset $D = \{(x_i, y_i), i = 1, \cdots, n\}$, the parameters w of the LSVM are learned by solving the following optimization problem:

$$\min_{w \in \mathbb{R}^d, \xi \in \mathbb{R}^n} \quad \frac{1}{2}\|w\|^2 + C \sum_i \xi_i \tag{1}$$

$$\text{s.t.} \quad \max_{h \in \mathcal{H}} y_i w^\top \phi(x_i, h) - \max_{\hat{h} \in \mathcal{H}} (-y_i) w^\top \phi(x_i, \hat{h}) \geq 1 - \xi_i, \tag{2}$$

$$\xi_i \geq 0, \forall i. \tag{3}$$

The regularization term $\|w\|^2$ in the objective function helps to avoid over-fitting. In addition, the objective function also minimizes the sum of the slack variables ξ_i, one for each sample (x_i, y_i). A small value of the slack variable results in the correct classification of a training sample. The constraints in problem (1) encourage the best latent variable for the correct output to have a score that is greater than all other possible latent variable assignments for the incorrect output. The number of constraints (2) is large. They consist of all possible assignments of $\hat{h}$ for every sample (precisely $|H| \times |X|$ constraints). However, the cutting plane algorithm [13] enables this optimization procedure in an efficient way regardless of the number of constraints.

In other words, the values of the latent variables are estimated such that the classification performance is maximized over the training set. Note that, for simplicity, we have restricted our description to a binary LSVM. However, we note that more general structured output LSVMs have also been proposed in the literature [20].

Algorithm 1 The CCCP method for learning the parameters of LSVM.

Input: Training set $\mathcal{D} = \{(x_i, y_i), i = 1, \cdots, n\}$, initial parameters w_0, tolerance ϵ.
Initialize $w = w_0$. Set $t = 0$.
repeat
Estimate the latent variables as $h_i \leftarrow \arg\max_{h \in \mathcal{H}} y_i w_t^\top \phi(x_i, h)$, for all i.
Update the parameters by solving the following convex optimization problem:

$$w_{t+1} = \underset{w \in \mathbb{R}^d, \xi \in \mathbb{R}^n}{\arg\min} \quad \frac{1}{2}\|w\|^2 + C \sum_i \xi_i \tag{4}$$
$$\text{s.t.} \quad y_i w^\top \phi(x_i, h_i) - \max_{\hat{h} \in \mathcal{H}} (-y_i) w^\top \phi(x_i, \hat{h}) \geq 1 - \xi_i, \xi_i \geq 0, \forall i.$$

Set $t \leftarrow t + 1$.
until The decrease in the objective function of problem (1) is below tolerance ϵ.

While problem (1) is not convex, it was shown to have the special form of a difference-of-convex program [20]. This observation leads to an approximate algorithm based on the concave-convex procedure (CCCP) [21] as outlined in Algorithm 1. The CCCP method iterates over two main steps: (i) the latent variable values are imputed using the current set of parameters; and (ii) the parameters are updated while keeping the imputed latent variables fixed, which is equivalent to optimizing the convex problem (4). In our work, we used the 1-slack reformulation based cutting plane algorithm [13] to solve problem (4). Each iteration of CCCP decreases the objective function of problem (1) until we reach a local minimum or saddle point solution [17].

3.1 Classification of M/EEG data

We now describe how the above LSVM framework can be adapted for the classification of M/EEG data. The input x corresponds to an M/EEG recording where data was collected from a single subject. The output y denotes the outcome. The unknown latent variables model the variation of a sample. The latent variable represents the possible transformations that M/EEG data may undergo. Such transformations are a result of the variability of the brain responses over trials. The latent space $\mathcal{H}$ can vary from a simple translation of the signal to multiple translations for different signal components (as determined by ICA).

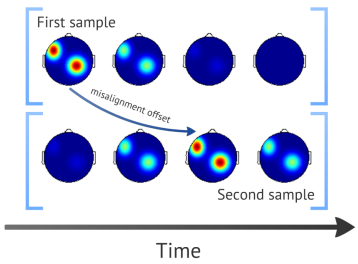

Fig. 1: In this work, we primarily consider variations in the data due to offsets in the time domain. By appropriately modeling such offsets, we are able to register the data samples with respect to each other and to enable the use of non-shift-invariant function classes. Additionally, the values of the latent variables are informative to quantify variations in brain responses.

We consider a simple distortion model where samples are shifted with respect to each other. We model this distortion using latent variables that represent the

putative offset of the misalignment as shown in Figure 1. The latent space $\mathcal{H}$ contains a finite set of translations.

In more detail, we consider an input x that consists of c channels and multiple samples collected from the single subject. Each channel consists of m observed values:

$$x = x^{(1)}, \ldots, x^{(c)} \tag{5}$$

$$x^k = (a_1^k, a_2^k, \cdots, a_m^k)^\top \tag{6}$$

In the absence of misalignment between the trials, we use the elements in the range $(s, s+l)$ for each of the c channels to perform classification. However, as mentioned earlier, the prediction performance can be considerably improved by explicitly modeling the misalignment using a variable $h \in \mathcal{H}$. In such a setting, we define the joint feature vector $\phi(x, h)$ as follows:

$$\phi(x, h) = (\phi(x^{(1)}, h)^\top, \ldots \phi(x^{(c)}, h)^\top)^\top \tag{7}$$

$$\phi(x^{(k)}, h) = (a_{s+h}^{(k)}, a_{s+1+h}^{(k)}, \ldots, a_{s+l+h}^{(k)})^\top, \tag{8}$$

$$1 \leq s + h \leq s + l + h \leq m. \tag{9}$$

The joint feature vector consists of elements in the range $(s+h, s+l+h)$ for all the c channels. Note that when the latent space $\mathcal{H} = \{\gamma\}$ for any constant γ, the resulting LSVM simplifies to the standard SVM formulation.

In our experiments, x consist of data on the basis of channels as in the experiment described in Section 5.1 (c indexes channels), or on the basis of ICA components as in the experiments described in Sections 5.2 and 5.3 (c indexes components).

3.2 Component quality measure

ICA components are often considered as a proxy to brain sources [7], and in this studies we perform experiments on ICA components. A common approach in visualization studies is to average data coming from multiple samples in order to improve the signal-to-noise ratio. However, the averaging of slightly misaligned time-series often manifests itself in the elimination of high frequency components of the signal (Figure 2). By using the imputed values of the latent variables, we can correct for this loss and greatly improve the quality of M/EEG signals. In order to quantify this improvement, we propose a quality measure. We would like such a measure to favor sharpness (that is, the presence of high frequency components) over smoothness (that is, a lack of high frequency components). To this end, we propose to use the H^1 norm [1], which is defined as follows:

$$\|u\|_{H^1} = \left(\|u(x)\|_{L^2}^2 + \|Du(x)\|_{L^2}^2\right)^{\frac{1}{2}} \tag{10}$$

For a function of two variables, the H^1 norm tends to infinity if a function has a discontinuity over a 1-dimensional curve. The H^1 norm assigns high values to functions of M/EEG recordings that have sharp transitions spatially and

over short time spans. This therefore favors functions that are spatially and temporally well-localized within the brain. In our experiments, we will calculate the H^1 norm of time series that are 2D flattened topographies arising as the difference of means of samples belonging to individual classes.

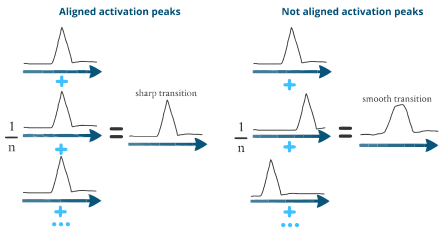

Fig. 2: Averaging over aligned data results in sharp peaks (n is a number of samples). In contrast, averaging over misaligned samples tends to smooth the data. Sharpness and smoothness can be quantified with the H^1 norm that takes a high value for sharp peaks and a low value for smooth time series.

3.3 Inferring brain functional connectivity from the latent variables

The brain is a distributed system with cognitive processes involving multiple brain regions that are recruited sequentially or simultaneously. In order to understand brain processes, one has to find which parts of brain are associated with a particular cognitive task, but also the chronometry of information flow between each of these regions. Here we propose a method to infer statistical dependence and brain functional connectivity. To investigate couplings and interaction between sources we propose to use the estimates of the latent variables that encode the trial-to-trial variability of the response of each source. Intuitively if two sources have similar variability, here delays, it means they have a statistical dependency that could originate from a common node in the brain communication network, or that one of them interacts directly with the another. Delays to a common ancestor might cause a delay in all its descendants (e.g. the visual cortex can process data as soon as it receives a signal from retina, but not before).

Imputed offsets are not easy to compare directly between components. Firstly, as all offsets can be shifted by a constant value, the resulting offsets give the same relative misplacement. Secondly, M/EEG data is noisy and some samples might be aligned incorrectly or only approximately. Even comparing imputed offsets from two perfectly dependent components can be difficult. Rather than compare the resulting offsets from separate LSVM setups (each for different components), we propose to use a single LSVM with a shared latent variable between components. We use obtained offsets to align components and then measure the quality of the result by calculating its H^1 norm as described in Section 3.2. A high value of the H^1 norm indicates correct alignment, as misalignment removes high frequency components of the signal. As each pair of components optimized with a shared latent variable gives us two measures (one for each component), we combine them by multiplication. Multiplication is chosen over addition to ensure a high value of this score only if the resulting functions are sharp for both components, and not just one.

4 Data collection and experimental paradigm

The considered dataset explores the process of long term memory (LTM) retrieval. The goal of the experiment is to elucidate the dynamics of long term memory encoding. The dataset is publicly available. [7] Details of data acquisition can be found in [2]. The task includes visual presentation, and the subject has to determine whether an abstract visual pattern corresponds to a presented natural object. The discriminative task to be solved with the LSVM is a binary classification problem (green color recall vs. red color recall). In these studies we have considered a single subject (number 8).

The long term memory retrieval experiment involves performing a complex, high level task by participants. The outcome of this kind of task is dependent on the subject's mental state, such as the level of concentration, vigilance, or familiarly with the experimental setting. We hypothesize that these factors cause the brain to respond with different temporal delays. While earlier visual processing may additionally have variable delays, high level cognitive functions, which are particularly challenging and interesting to study, are more susceptible to this form of variability due to the longer time frames involved, and the recruitment of multiple brain regions. For this reason, the LTM dataset considered in this study is particularly suited to the form of statistical modeling proposed here.

For further analysis, we have processed the dataset by dropping 10% of trials with the highest variance. 69 trials were removed out of 681 by this process. We reduced the data dimensionality with PCA to 60 dimensions and whitened the data. Finally, we applied InfoMax ICA with full rank [3]. We have applied PCA in conjunction with ICA in line with standard practice [12].

5 Results

To test the efficacy of LSVM, we performed a binary color prediction task on the LTM dataset. We first evaluated the prediction performance quantitatively, and subsequently visualized the ICA components after discriminative alignment with the LSVM. Finally, we used the learned offsets of the ICA components to infer a graph indicating likely functional connectivity between components.

5.1 LSVM using all channels

Here, we used data before the application of ICA, i.e. the dataset has been cleaned by dropping malicious samples and further whitened with PCA, however we have not applied ICA as the learned decision function will linearly transform the data. This experiment does not treat individual components differently, but instead learns a single offset parameter for the entire trial.

We considered a distortion model where the latent space H consists of a finite number of translations. Figure 3 presents the accuracy results for

[7] http://www.biomag2012.org/content/data-analysis-competition

various sizes of the latent space. The dataset was balanced with a chance level at 50%. The point denoted by 10ms in the x-axis denotes the experiment where the putative translation is restricted to lie in the interval $[-10\text{ms}, 10\text{ms}]$. For computational efficiency, we discretized the space of putative translations into 7 equally spaced values, resulting in the latent space $H = \{-10\text{ms}, -6.7\text{ms}, -3.3\text{ms}, 0\text{ms}, 3.3\text{ms}, 6.6\text{ms}, 10\text{ms}\}$ (the data acquisition rate is 300Hz). For the sake of brevity, we refer to the LSVM setup with maximum misalignment N ms as $MisAlign_N$. Particularly, for $N = 0$, $MisAlign_0$ simplifies to a classical SVM setup, where no misalignment is considered.

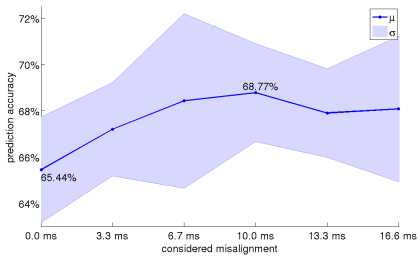

Fig. 3: Results of a LSVM for the long term memory dataset where the latent variable models misalignment. A paired t-test indicates with p-value smaller than 5% that, LSVM for misalignment up to 10ms performs statistically significantly better than a classical SVM.

Based on a preliminary analysis accuracy results obtained for very high C parameter where indistinguishable from results for cross validated C. In this and all further experiments we consider only a hard margin SVM (equivalent to setting C to infinity). The results presented in Figure 3 are averaged over 5-folds. The accuracy obtained for $MisAlign_{10}$ is 3.33% higher compared to the accuracy obtained by a standard SVM. The accuracy peaks when we consider the latent variable to lie in the interval $[-10\text{ms}, 10\text{ms}]$ and it slowly decays for larger values of misalignment. Our experiments indicate that the misalignment of most of the samples is up to 10ms and considering higher values gives too much capacity to the learning algorithm (for the majority of samples, higher values do not correspond to actual data misalignment). A paired t-test indicates that the accuracy of $MisAlign_{10}$ is significantly improved over the accuracy distribution obtained for a standard SVM, and rejects the null hypothesis with p-value equal to 4.36%. Note that we are able to achieve higher classification performance with statistical significance which is a strong evidence that the use of latent variables for discriminative alignment is an appropriate modeling choice for this class of data.

5.2 LSVM on ICA components

Over the course of this experiment, we consider single ICA components computed from 200ms long time slices. We regard a single component as a proxy to a brain source [14]. We visualize components by averaging them over trials. We may consider such averaging with or without first aligning the trials using the offsets learned by the LSVM. We considered 60 components and 3 time intervals (0-200ms, 200ms-400ms, 400ms-600ms). We have used each of these component-interval pairs in separate prediction tasks, and we focus on the four pairs with

highest prediction accuracy. Moreover, we focus only on component-interval pairs that give at least 1% improvement in $MisAlign_{10}$ over a classical SVM. In this way, we examine only components that carry an informative signal (high accuracy), and substantially suffering from misalignment.

To compare the discriminative alignments learned from the LSVM to the previous state of the art, we compare to the continuous profile model (CPM) [15], an unsupervised method. In total, we present three visualizations: (i) the unaligned ICA component, (ii) the ICA component aligned by the application of CPM, and (iii) the ICA component after aligning samples according to the learned offset from LSVM. In order to visualize a single component, first we mapped the data back to the channel space. Next, we took a single time slice and computed the mean for each prediction class separately. Figure 4 presents the absolute value of results. Red indicates that the mean of samples belonging to the one class highly differs from the mean of samples belonging to the opposite class. Figure 4 presents three time slices that demonstrate the difference between methods (additional results do not show a qualitative difference and are omitted due to space restrictions). The visualization obtained after alignment with LSVM is significantly sharper, while the other two visualizations are diffuse and the underlying structure is not visible.

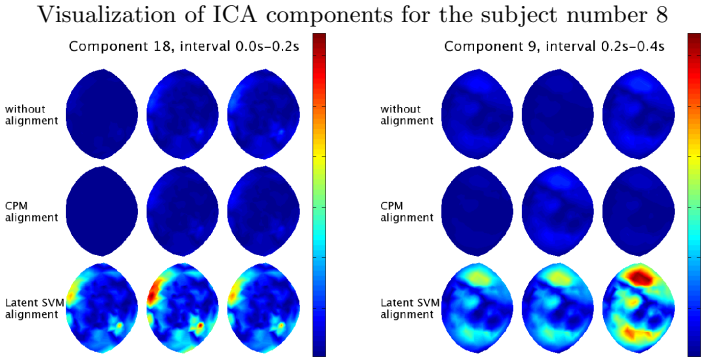

Fig. 4: Visualization of two ICA components with various alignment techniques. The figure presents the absolute value of the difference between the target class means. A difference between samples corresponding to the first outcome of a mental state and the second outcome (in this case mean of samples of green color recall minus mean of samples of red color recall). Red on this figure indicates regions that discriminate between classes. All methods make use of the same color palette to facilitate the comparison between subfigures.

Table 1 presents the H_1 norm for four different components and four different alignment methods. Images are first normalized by setting their mean to zero (centering) and standard deviation to one. The score obtained for data aligned according to the LSVM is much higher than for data without any alignment and for data aligned with the continuous profile model method [15]. Moreover, we evaluated the stability of the H^1 norm over randomly aligned images. We randomly generated alignment offsets and shifted images with respect to them.

		(component; time interval)			
		(18; 0s–0.2ms)	(58; 0s–0.2ms)	(9; 0.2s–0.4s)	(2; 0.4s–0.6s)
method	none	2.44	5.17	2.26	1.61
	random	2.51 ± 0.08	5.07 ± 0.31	2.18 ± 0.05	1.54 ± 0.08
	CPM	3.49	4.71	1.65	1.20
	LSVM	10.28	17.52	7.19	8.02

Table 1: The H^1 norm over normalized difference of means of samples belonging to individual classes. Results are computed for different methods of alignment. For every component LSVM achieves significantly higher values of the H^1 norm. The H^1 norm measures the spatio-temporal sharpness of a time series that are 2d flattened topographies.

For the resulting randomly aligned images we calculated the H^1 norm. The second row of Table 1, results for random alignment, presents the mean and standard deviation achieved over 5-folds. Values in this row are not substantially different from values in the first row, where none of the alignment methods have been applied. The relatively small standard deviation indicates that the H^1 norm is very stable in our setting.

5.3 Inference of brain functional connectivity with LSVM

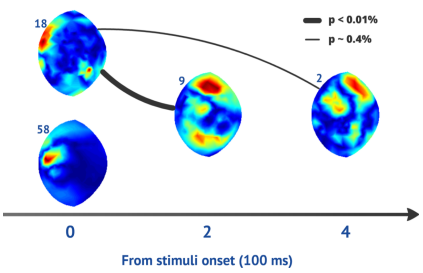

Fig. 5: Components giving similar misalignment offsets. An edge indicates that aligning components according to a common latency results in a high product of H^1 norms (c.f. Section 3.3). Statistical significance was verified with a permutation test. Edges are annotated by their p-values.

As described in Section 5.2, we focus on four components in these experiments. For each of the $\binom{4}{2}$ pairs of components and three subintervals of length 200ms, we have computed the product of their H^1 norms resulting from latent alignments estimated by joint discriminative training. Considering this score as a statistic, we verify if it is significantly larger than chance by computing permutation tests. Under the null hypothesis (H0) that there is no delay dependency between components across trials, we generated permuted data by shuffling the trials for one component leaving the other one unchanged. For each resulting permuted data we computed the same statistic to assemble a histogram generated under H0. The original statistic value is then positioned in the histogram to derive a p-value. Component 18 over interval 0.0s–0.2s and component 9 over interval 0.2s–0.4s achieved significant statistical scores using 10000 permutations (Figure 5). We observe 4 topographies, the 2 on the left exhibit dipolar patterns

located at the back on the helmet above the occipital cortex that contains the visual cortex. Component 9 is statistically related to component 18 over the interval 0.0s–0.2s ($p < 0.01\%$) and shows a relative symmetric pattern that could correspond to a deep subcortical source involved in long term memory. Component 2 reflects an activation over interval 0.4s–0.6s ($p < 0.4\%$ with component 18) on the left side of the helmet over more frontal sensors which could correspond to higher level cognitive processing that naturally appears later in time after the stimulus onset.

6 Discussion

By modeling and estimating parameters of variations on single trial M/EEG data, LSVM has demonstrated a significant improvement with respect to a standard SVM, which has been previously used in neurosciences and for BCI applications [18]. The proper modeling of brain response variabilities via latent variables allowed us to estimate in a supervised way the parameters reflecting the changes in brain activations due for example to fatigue or subject habituation.

Exploiting the ability of ICA to exhibit components that are plausible brain sources according to the physics of the measurement system (high activations spatially localized with spatial smoothness and dipolar field patterns), we then run LSVM on ICA components to investigate the dynamics and chronometry of different brain source configurations. Results from Section 5.3 show the potential of this approach for functional connectivity studies as it offers a way to elucidate delays in brain responses from single trial data. Indeed, from correlated delays between sources one can for example infer if a source activation precedes another one or have a common cause that could be a deep subcortical source.

Future directions for this work is to investigate recovery of brain functionality graph for large number of components. Finally a next step is the localization of the ICA components in the brain by solving the M/EEG inverse problem [10].

Acknowledgements

This work is partially funded by the European Research Council under the Seventh Framework Programme (FP7/2007-2013)/ERC Grant 259112.

References

1. Adams, R., Fournier, J.: Sobolev Spaces: Pure and Applied Mathematics. Pure and Applied Mathematics - Academic Press, Academic Press (2003)
2. Backus, A., Jensen, O., Meeuwissen, E., van Gerven, M., Dumoulin, S.: Investigating the temporal dynamics of long term memory representation retrieval using multivariate pattern analyses on magnetoencephalography data. MSc thesis (2011)
3. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7, 1129–1159 (1995)

4. Bénar, C., Clerc, M., Papadopoulo, T.: Adaptive time-frequency models for single-trial M/EEG analysis. In: Proceedings of the 20th international conference on Information processing in medical imaging. pp. 458–469. IPMI'07
5. Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic resonance in medicine 34(4), 537–541 (Oct 1995)
6. Blankertz, B., Muller, K., Curio, G., Vaughan, T., Schalk, G., Wolpaw, J., Schlogl, A., Neuper, C., Pfurtscheller, G., Hinterberger, T., et al.: The bci competition 2003: progress and perspectives in detection and discrimination of eeg single trials. Biomedical Engineering, IEEE Transactions on 51(6), 1044–1051 (2004)
7. Chen, Y., Akutagawa, M., Katayama, M., Zhang, Q., Kinouchi, Y.: Ica based multiple brain sources localization. Conf Proc IEEE Eng Med Biol Soc 2008, 1879–82 (2008)
8. Cortes, C., Vapnik, V.: Support-vector networks. In: Machine Learning. pp. 273–297 (1995)
9. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9), 1627–1645 (2010)
10. Gramfort, A., Strohmeier, D., Haueisen, J., Hamalainen, M., Kowalski, M.: Functional brain imaging with m/eeg using structured sparsity in time-frequency dictionaries. In: Proceedings of the 22th international conference on Information processing in medical imaging, IPMI'11, vol. 6801, pp. 600–611. Springer-Verlag, Berlin, Heidelberg (2011)
11. Held, K., Kops, E., Krause, B., Wells III, W., Kikinis, R., Muller-Gartner, H.: Markov random field segmentation of brain MR images. Medical Imaging, IEEE Transactions on 16(6), 878–886 (1997)
12. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley-Interscience (2001)
13. Joachims, T., Finley, T., Yu, C.N.J.: Cutting-plane training of structural svms. Machine Learning 77(1), 27–59 (Oct 2009)
14. Jung, T., Makeig, S., Mckeown, M.J., Bell, A.J., won Lee, T., Terrence, Sejnowski, J.: Imaging brain dynamics using independent component analysis. In: Proceedings of the IEEE. vol. 89, pp. 1107 – 1122 (2001)
15. Listgarten, J., Neal, R.M., Roweis, S.T., Emili, A.: Multiple alignment of continuous time series. In: in Advances in Neural Information Processing Systems. pp. 817–824. MIT Press (2005)
16. Polich, J.: Updating p300: An integrative theory of P3a and P3b. Clinical Neurophysiology 118(10), 2128 (2007)
17. Sriperumbudur, B., Lanckriet, G.: On the convergence of the concave-convex procedure. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I., Culotta, A. (eds.) Advances in Neural Information Processing Systems 22, pp. 1759–1767 (2009)
18. Thulasidas, M., Guan, C., Wu, J.: Robust classification of EEG signal for brain-computer interface. Neural Systems and Rehabilitation Engineering, IEEE Transactions on 14(1), 24–29 (2006)
19. Woody, C.: Characterization of an adaptive filter for the analysis of variable latency neuroelectrical signals. Medical and Biological Engineering 5, 539–553 (1967)
20. Yu, C.N.J., Joachims, T.: Learning structural SVMs with latent variables. In: Proceedings of the International Conference on Machine Learning (ICML) (2009)
21. Yuille, A.L., Rangarajan, A.: The concave-convex procedure. Neural Computation 15(4), 915–936 (2003)