

Estimation of Task Workload from EEG Data: New and Current Tools and Perspectives

Christian A. Kothe and Scott Makeig

Abstract— We report, as part of the EMBC meeting Cognitive State Assessment (CSA) competition 2011, an empirical comparison using robust cross-validation of the performance of eleven computational approaches to real-time electroencephalography (EEG) based mental workload monitoring on Multi-Attribute Task Battery data from eight subjects. We propose a new approach, Overcomplete Spectral Regression, that combines several potentially advantageous attributes and empirically demonstrate its superior performance on these data compared to the ten other CSA methods tested. We discuss results from computational, neuroscience and experimentation points of view.

I. INTRODUCTION

RECENT sensor technology and analysis advances in signal processing and machine learning make it possible to noninvasively monitor brain signals and derive from them useful aspects of a person's cognitive state in near real time [1,2]. It is now becoming feasible to integrate this technology into real-world, real-time systems to enhance human-machine interaction across a wide range of application domains including clinical, industrial, military and gaming [3,4,5,6].

However, progress in cognitive monitoring requires parallel development of new recording and analysis methods, experimental research, and empirical studies of experimental data recorded under quasi-realistic yet well-controlled operating conditions in representative subject populations. Such data allows often-neglected aspects of brain-computer interface (BCI) problems, including inter-individual differences and day-to-day variability, to be addressed. The present Cognitive State Assessment (CSA) competition 2011, its associated experimental task and accumulated data therefore provide much-needed steps toward the development of robust methods and applications.

Workload measurement technology has been incrementally improved and tuned to the point where claims of near-perfect accuracy are not unheard of, despite relatively high recording noise levels, tremendous complexity of the brain, and current incomplete understanding of the underlying brain EEG signals [7,8]. There is an increasing need to compare and evaluate cognitive state estimation methods on equal footing, in particular because of the great variety of experimental tasks that have been proposed to assess different aspects of

workload [8,9,10]. The ever-present risk of circular analysis in complex pattern recognition problems [11] also demands reliable and agreed-upon evaluation procedures for measuring estimation performance.

A watershed separating the current state of the art of CSA from demonstrated robust performance in real-world settings is likely less the details of the estimation method applied, and more the amount, type and expense of the training data that is required to calibrate a predictive model capable of robust performance on later in-use data. Several factors make learning robust cognitive state estimation models difficult. First, every person has unique anatomic and functional brain geometry – both contributing to the observed inter-individual differences in the measured scalp signals. Second, because of non-reproducible sensor positioning and varying electrical conductivities at the electrode-skin interface every EEG recording session involves a sensor montage with a slightly different geometry and placement with respect to the underlying brain EEG source signals. EEG brain activity is itself highly non-stationary at all time scales (seconds to years). Thus, the further any two measurements are separated in time, the stronger the expected differences in the observed brain dynamics. To develop cognitive estimation methods that (like some recent voice recognition systems) do not require lengthy, repeated calibration or individualization, there is a need for data sets that span multiple sessions from a large number of individuals.

In this initial phase of the CSA competition, the main goal is to establish a performance baseline for current state-of-the-art methods for real-time monitoring of workload. The results we present here are restricted to predictive estimation – performing a two-class discrimination task between “high” and “low” workload levels – of the performance of CSA systems that are trained on relatively short (7.5 minutes x 2 conditions), low-density (19 EEG + 2 EOG channel) data recorded from the same person on the same day and using the same montage as the data on which they are to be tested. We present preliminary performance comparisons of several state-of-the-art CSA methods including a new computational approach, introduced here, that leverages recent advances in convex optimization and statistical modeling of brain sources via a recent extension of Independent Component Analysis (ICA). Other results to be presented at EMBC 2011 will also evaluate performance of methods trained across days and montages.

Manuscript received April 15, 2011. This work was supported in part by a gift of The Swartz Foundation (Old Field NY) and a basic research grant from the Office of Naval Research. C. A. Kothe and S. Makeig are with the Swartz Center for Computational Neuroscience, Institute for Neural Computation, UCSD, La Jolla CA 92093-0559 USA (corresponding author phone: +1-858-822-7538; e-mail: {christian,scott}@scn.ucsd.edu).

II. MATERIALS AND METHODS

A. Overview

Most approaches for real-time mental state estimation from the EEG (or other physiological measures) of a person can be understood as signal processing systems that translate biosignals into a time-varying (and often binary) output [12], in this case a workload index [8,10]. In recent years, these and other cognitive monitoring systems have adopted a (supervised) machine learning approach to adaptively learn person-specific and session-specific parameters from calibration data [5,12,13]. This gives rise to a two-stage task. First, given a calibration recording with known ‘workload-level’ labels, learn a predictive model. Second, given some period of previously unseen signal, predict or estimate its respective workload index using the previously trained model.

The workload estimators reviewed and compared below operate on spectral features of the signal, thereby capturing features of the oscillatory dynamics of cortical networks we assume are linked to a latent factor indexing mental workload. Existing approaches differ primarily in what spectral dynamics are extracted, under which assumptions, how the spectral estimates are computed, and how they are jointly mapped to produce the final CS estimate. The CSA methods selected below have been chosen to investigate the effects of a variety of these design parameters using common data.

All analyses have been performed using our open-source BCILAB toolbox compatible with the EEGLAB signal processing environment for Matlab (The Mathworks, Inc.) [14]. The BCILAB toolbox is freely available for download and non-commercial use at <ftp://sccn.ucsd.edu/pub/bcilib>. The scripts used to reproduce the analyses are also available online at <ftp://sccn.ucsd.edu/pub/csac/>.

B. Data

The CSA competition data were recorded from 8 participants each of whom performed multiple sessions of the Multi-Attribute Task Battery [15] over 5 days (three runs per day) spread out over a month. From these data, 32 data sets were generated by the competition organizer to serve as training data sets, each containing an unspecified subset of concatenated blocks recorded from one participant on one day. Each data set contains an equal amount of data collected under the "high workload" condition and the "low workload" condition. The measurements include EEG recorded at 19 scalp sites according to the international 10-20 system, as well as vertical and horizontal electrooculographic (EOG) data measured using two bipolar electrode pairs. Each training set is accompanied by several short unlabeled test sets that form the basis for competition results comparison. For this purpose, we have submitted predicted labels for the test set for the new approach we propose (Section II D).

C. Comparisons

We contrast methods that extract spectral features directly from sensor signals [8] with methods that extract spectral features from estimated (brain) source signals [16,17]. Since EEG sensor signals are a linear mixture of source signals conveyed to the electrodes by volume conduction, this amounts to a choice of a linear spatial filter and a method to determine the filter's parameters [18,19]. Here, we compare *a priori* filters based on the surface Laplacian [20], the popular Common Spatial Patterns supervised learning approach [16] and an unsupervised "bottom-up" learning approach based on Independent Component Analysis [18].

The second method comparison, largely orthogonal to the first, selects the spectral features of the derived signals to use for CSA. Here, we compare use of fixed *a priori* frequency bands (in particular, the common delta, theta, alpha, beta, and high beta bands, defined here as 0.5-3, 4-7, 8-12, 13-30, and 31-42 Hz, respectively) [21,22] versus PCA-based power spectral dimensionality reduction [23] or raw (high-dimensional) Fourier amplitude estimates. A third comparison here is the choice of spectral estimator: Welch-type windowed FFT [24] or multi-taper spectral estimation [25]. We use higher-quality estimators (in particular multi-taper spectral estimation, here using 15 tapers) where applicable. An aspect we omitted here because of time and computational constraints is the use of measures of linear delayed dependencies between (source or sensor) signals such as coherence [26]. Methods taking into account such coupling will be discussed at the Competition session.

Finally, we compare the application of a variety of machine learning methods on the resulting spectral estimates. Linear classifiers [27,28,29] have been shown to be good contenders, and are widely appreciated for their speed, simplicity, and interpretability. Nevertheless, non-linear classifiers can potentially capture richer relationships between signal features and cognitive state, as has been demonstrated empirically in some scenarios [8,30]. As our data is high-dimensional, we use a feature-selecting non-linear classifier. While there exist greedy methods for selecting features with non-linear methods, here we test the use of Hierarchical Kernel Learning (HKL) [31], a recently proposed globally optimal method for non-linear feature selection that generalizes the concept of multiple-kernel learning by means of structured sparsity [32].

D. Proposed Approach

For the given task, we propose a family of methods that combines several of these potential advantageous elements, a method we here call Overcomplete Spectral Regression (OSR). The method first expands the original multivariate sensor signal by mapping it onto an overcomplete source signal representation (further subdivided into a spectral representation), and selects, in a second step, a sparse set of features using a strong linear or non-linear sparse learning method. The first step is accomplished by Adaptive Mixture ICA (AMICA), a recent generalization of independent component analysis (ICA) into a temporal mixture of independent component decompositions, which is efficiently solvable in an EM framework using the Newton method [19]

The AMICA decomposition, including automatic artifact rejection, was performed without expert intervention using the BCILAB defaults. This step gave a set of linear unmixing matrices and models (here we chose to use 3 models), that we concatenated into a joint, overcomplete component unmixing matrix. Applying this unmixing to the data, a set of 63 (less some duplicates) largely statistically independent signal components were obtained. From the resulting signals, spectral properties can be extracted in a variety of ways; here we extracted log power spectral density using multi-taper spectral estimation averaged over consecutive windows.

Other factors that determine performance are the window length used for spectral estimation. We used here a moderate Hann window of 12 seconds and did not average outputs. To allow more effective later use of sparse learners, we decorrelated the windowed spectrum of each identified component process across training data segments by shrinkage PCA [33]. We then pruned the spectrum to the leading 10 principal components per component process (in part to save computation time). Finally, a supervised sparse linear or non-linear learning function was applied to the data to select the relevant basis function subset. In the linear case, the Least Angle Regression approach [28] with nested block-wise cross-validation to select the regularization parameter is a good choice and is also easily an order of magnitude faster than comparable methods. The effect of using sparse non-linear learners such as HKL on these data will be discussed in a subsequent report, in particular with respect to overcomplete source coupling parameters.

E. Evaluation

Empirical evaluation is a crucial step in machine learning studies, particularly for problems in which ground truth results are not available and the data generation process is poorly understood. In the present competition, in which each labeled data set is only 15 minutes in length and has an unknown internal block structure and the workload variable of interest is constrained to be slow-changing, assessing the predictive performance of a machine learning method without "double dipping" into the data [11] or over-fitting is a difficult challenge. We list some relevant caveats and approaches to tackle these problems in the following.

The standard evaluation approach in machine learning is to separate the data into trials followed by cross-validation (either, e.g., k -fold randomized or leave-one-out) [34]. Because here, the trials on which the evaluation is performed come from a time series with multi-scale quasi-stationarity, the basic assumption of testing on independently and identically distributed trials does not apply. For this reason, cross-validation needs to be performed at the block granularity (i.e., leave-one-block-out cross-validation), with the block size ideally matching up with underlying natural boundaries in the data (as e.g. is natural in dealing with data from a block design experiment). Also, data within a few seconds from testing data should be excluded from training.

Because some pre-processing steps compute statistics across trials (in particular, ICA, PCA and, to a lesser extent, various artifact rejection methods) these steps also have to

be computed within the respective training data only (and be repeated independently for each fold of the cross-validation). This crucial step is easily overlooked, in some cases yielding drastically optimistic evaluations. Here, these considerations were automatically accounted for by the BCILAB toolbox, which partitions the data before these operations.

The analyses below were obtained using five-fold blockwise cross-validation (i.e., leaving out 20% of the data for testing in each fold), with 15-sec boundary epochs ignored bordering test data blocks. Trials have been extracted from the raw data as 12-second windows overlapping by 11 seconds and tapered using the Hann function. Because the periods of high and low workload were sent to us without their original temporal relationships, the training / test sections were drawn separately from the respective two data sets. All hyper-parameters (including regularization parameters) were also optimized using five-fold nested blockwise cross-validation [35].

III. RESULTS

Results of comparative evaluations across the first 10 of the 32 data sets are presented in Table 1. A subset of the computationally less demanding methods, applied to all 32 data sets, gave very similar results (results not shown). The proposed Overcomplete Spectral Regression method outperformed all other methods (both channel-based and spatial filter-based methods) by a wide margin ($p < 0.01$). Note also that adaptive learning of relevant frequency bands via PCA coupled with feature selection (FFTDC-LARS; cf. Table 1 for the spelled-out method names) significantly outperformed analogous methods that used fixed bands, including MBLP-LARS, FBCSP-VBARD, and FBCSP-LDA ($p < 0.05$). Further, as expected, the standard single-band BCI methods, such as Common Spatial Patterns, performed significantly worse than their multi-band analogs (Filter Bank CSP). Among the unexpected outcomes, the recently-proposed and elsewhere highly successful Dual-Augmented Lagrangian (DAL) method [36] did not perform significantly better than the other multi-band approaches, and supervised methods for learning spatial filters (CSP and DAL) did not improve significantly over their channel-wise analogs. Finally, the best channel-wise method (MTDC-LARS) outperformed a variety of (single-band) spatially filtered methods, including CSP ($p < 0.05$).

IV. DISCUSSION

Our evaluation was set up to be able to capture at least some of the block-to-block transitions between training and test sets, which is presumably the main reason for the performance estimates from other methods here being somewhat lower than those of comparable methods reported in the literature. Other factors that determine performance include the averaging of multiple successive classifier outputs or longer spectral estimation windows. In agreement with the literature, the analyses revealed that multiple frequency bands are clearly necessary for EEG-based workload classification. Moreover, the results suggest that adaptive band selection appears to be generally beneficial.

A key result of this study, however, is the remarkably good performance of the OSR method introduced here; it returned just over one third of the number of errors compared to the next best-performing method. In difficult estimation problems, very good results are often an indicator of invalid analysis methods (inadvertent ‘cheating’); here, however, during careful double-checking of the results we have discovered no such problems. To possibly increase confidence in their validity and encourage additional comparisons, we are making our Matlab scripts available online for wider testing (and/or retesting).

The main differences between OSR and other methods we tested is that it extracts spectral properties of maximally *independent source signals* of which many are compatible with an origin in anatomically and functionally distinct cortical source areas [37], as opposed to channel-based estimation methods that attempt to estimate brain (workload) state based on highly correlated channel *mixtures* of the underlying source signals that project to each sensor.

While CSP (or any EEG-based algorithm involving linear filtering) can be said to operate *indirectly* on source signals, typically only a small fraction of the extracted signals are biologically interpretable; for CSP, for example, almost all are orthogonally constrained by the second-order (variance ratio) objective of the method. In contrast, OSR first retrieves far more signals than sensors (in our case over 60), a large fraction of which can be interpreted as signals produced by cortical patches (or equivalent dipoles) in the brain. Given this very large pool of data projections, the subsequent learning procedure must determine the set of relevant source signals and signal frequencies. When then combined with a convex linear or non-linear feature selector (such as l_1 -regularized logistic regression or hierarchical kernel learning), the method extracts the affected oscillatory properties of a parsimonious set of source processes, many of which are likely biologically and functionally distinct.

Despite their relatively good performance on this restricted training set, it is hard to predict from it the across-session or day-to-day behavior of any of the presented methods; i.e., it is entirely possible that the best method for these data sets has inferior performance for other data. To assess and address such effects, it is necessary to measure performance systematically across different sessions, days, and, ideally, subjects. The CSA competition data set provides a unique opportunity for this investigation, as the data recording schedule was spread out over a month. Given recordings across two or three days will allow for learning models that specifically account for the day-to-day variability. Going beyond methods that assume a homogeneous (unstructured) pool of training data, a setting that exposes this grouping structure (e.g., the session and subject labels) will enable the use of hierarchically structured statistical models including mixed-effects models [38], hierarchical Bayes [39], and a variety of other approaches that were recently proposed in the multi-task and transfer learning literature [40]. Being linked to the cortical

source space, methods such as OSR should be positioned well for such extensions, as both high-quality data co-registration tools, as well as anatomical (and to some extent functional) priors are applicable. We believe that adoption of these techniques in the cognitive monitoring domain, combined with wider data availability has great potential to overcome some long-standing barriers in the BCI/CSA field.

TABLE I
EVALUATION RESULTS

Method	Cross-validation error	Details
OSR	6.1 ± 5.5	Overcomplete Spectral Regression (error level decrease rel. to other methods is significant at $p < 0.01$)
<i>MBLP-LARS</i>	19.9 ± 11.8	Multi-band log-spectral power, using sparse linear Least-Angle Regression (fixed bands)
<i>MBLP-LARS-SL</i>	19.8 ± 9.9	Multi-band log-spectral power on Laplacian-derived channels using sparse linear Least-Angle Regression
<i>MTDC-LARS</i>	17.2 ± 9.6	Multi-taper (log-) spectral estimates per channel, decorrelated via Principal Component Analysis, using sparse linear Least-Angle Regression
<i>MT-LARS</i>	23.1 ± 11.9	Sparse linear Least-Angle Regression on raw multi-taper (log-)spectral estimates per channel
<i>FBCSP-LDA</i>	19.8 ± 14.7	Filter-Bank CSP using shrinkage Linear Discriminant Analysis
FBCSP-VBARD	19.6 ± 11.6	Filter-Bank CSP using sparse variational Bayes logistic regression
<i>FBCSP-HKL</i>	20.0 ± 14.5	Filter-Bank CSP using Hierarchical Kernel Learning
<i>FB-DAL</i>	23.4 ± 11.6	Filter-Bank Dual-Augmented Lagrangian
WB-CSP	23.2 ± 13.0	Wide-band filtered CSP (7-30 Hz)
WB-Spec-CSP	22.5 ± 13.7	Wide-band filtered Spectrally weighted CSP (7-30 Hz)

Cross-validation errors given as percentage of trials misclassified. Here, lower is better and chance level is 50%.

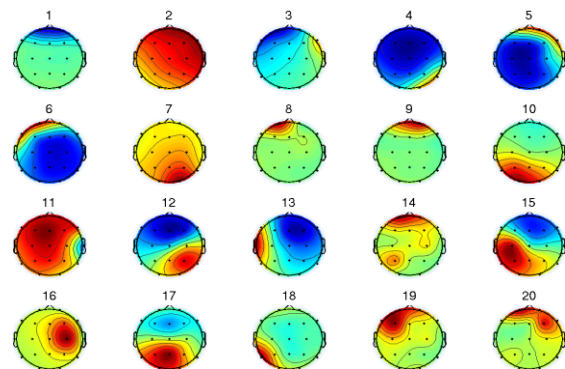


Fig. 1. Topographic forward projections (filter inverses) to the scalp of the first 20 independent components used by OSR in set 9 (median performance). Note the dipolar scalp projection patterns recovered for some components, compatible with a source in a single cortical patch.

A. References

- [1] Y. M. Chi, T.-P. Jung, and G. Cauwenberghs, "Dry-Contact and Noncontact Biopotential Electrodes: Methodological Review," *IEEE Reviews in Biomedical Engineering*, 3:106-119, 2010.
- [2] S. F. Liang, C.-T. Lin, R. C. Wu, Y.-C. Chen, T. Y. Huang, and T.-P. Jung, "Monitoring Driver's Alertness based on the Driving Performance Estimation and the EEG Power Spectrum," *Proc of the 27th Int'l Conference of the IEEE Engineering in Medicine and Biology Society*, Shanghai, 2005.
- [3] T. O. Zander, C. Kothe, S. Jatzev, M. Gaertner, "Enhancing Human-Computer Interaction with input from active and passive Brain-Computer Interfaces," *Brain-Computer Interfaces*, Vol. 0, pp. 181-199, 2010
- [4] C.-T. Lin, R.-C. Wu, S.-F. Liang, T.-Y. Huang, W.-H. Chao, Y.-J. Chen, and T.-P. Jung, "EEG-based Drowsiness Estimation for Safety Driving Using Independent Component Analysis," *IEEE Transactions on Circuit and System*, 52(12):2726-38, 2005.
- [5] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, G. Curio, and K.-R. Müller, "The Berlin Brain-Computer Interface: Non-Medical Uses of BCI Technology," *Front Neuroscience*, 4:198, 2010
- [6] A. Lécuyer, F. Lotte, R. Reilly, R. Leeb, M. Hirose, M. Slater, "Brain-Computer Interfaces, Virtual Reality, and Videogames," *IEEE Computer*, vol. 41, no. 10, pp 66-72, 2008
- [7] Y.-F. Tsai, E. Viirre, C. Strychacz, B. Chase, and T.-P. Jung, "Task Performance and Eye Activity Relating to Cognitive Workload," *Aviation, Space, and Environmental Medicine*, 78(5):B176-85, 2007.
- [8] G. F. Wilson, C. A. Russell, J. W. Monnin, J. R. Estep, J. C. Christensen, "How Does Day-to-Day Variability in Psychophysiological Data Affect Classifier Accuracy?," *Human Factors and Ergonomics Society Annual Meeting Proceedings, Augmented Cognition*, pp. 264-268(5), 2010
- [9] B. Cinaz, R. La Marca, B. Arnrich, and G. Tröster, "Towards Continuous Monitoring of Mental Workload," *Computing*, 14-25. Springer, 2010
- [10] J. Kohlmorgen, G. Dornhege, M. Braun, B. Blankertz, K.-R. Müller, G. Curio, K. Hagemann, A. Bruns, M. Schrauf, and W. Kincses, "Improving human performance in a real operating environment through real-time mental workload detection," *Toward Brain-Computer Interfacing*, MIT press, 2007
- [11] N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan, C. I. Baker, "Circular analysis in systems neuroscience – the dangers of double dipping," *Nature Neuroscience* 12(5): 535-40, 2009
- [12] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Trans Biomed Eng.* 51(6):1034-43, 2004
- [13] A. Delorme, C. Kothe, A. Vankov, N. Bigdely-Shamlo, R. Oostenveld, T. O. Zander, and S. Makeig, "MATLAB-based tools for BCI research," *Human-Computer Interaction Series*, 2010, Volume 0, Part 4, 241-259
- [14] A. Delorme, T. Mullen, C. Kothe, Z. A. Acar, N. Bigdely-Shamlo, A. Vankov, and S. Makeig, "EEGLAB, SIFT, NFT, BCILAB, and ERICA: New tools for advanced EEG/MEG processing," *Computational Intelligence and Neuroscience: Special Issue on Academic Software Applications for Electromagnetic Brain Mapping Using MEG and EEG* (in press)
- [15] J. L. Comstock and R. J. Arnegard, "The Multi-attribute Task Battery for human operator workload and strategic behavior research." Technical Report 104174. Hampton, VA: NASA Langley Research Center, 1992
- [16] H. Ramoser, J. Mueller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement." *IEEE Trans. Rehab. Eng* 8 pp. 446, 441, 1998
- [17] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Mueller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process Mag*, 25(1):41-56, January 2008
- [18] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in D. Touretzky, M. Mozer and M. Hasselmo (Eds). *Advances in Neural Information Processing Systems* 8:145-151, 1996)
- [19] J. Palmer, S. Makeig, K. Kreutz-Delgado, and B. D. Rao, "Newton method for the ICA mixture model," *ICASSP*, 2008
- [20] B. Hjorth, "An on-line transformation of EEG scalp potentials into orthogonal source derivations," *Electroencephalography and Clinical Neurophysiology*, 39 pp. 526-530, 1975
- [21] G. F. Wilson, and C. A. Russell, "Real-time assessment of mental workload using psychophysiological measures and artificial neural networks," *Human Factors* 45(5), 635-643, 2003
- [22] D. Grimes, D. S. Tan, S. Hudson, P. Shenoy, and R. Rao, "Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph," *CHI 2008 Conference on Human Factors in Computing Systems*
- [23] S. Makeig and T.-P. Jung, "Changes in alertness are a principal component of variance in the EEG spectrum." *Neuroreport*.29;7 (1):213-6, 1995
- [24] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," *IEEE ransactions on Audio Electroacoustics*, Volume AU-15, pages 70-73, 1967
- [25] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty – V: The discrete case," *Bell System Technical Journal*, Volume 57, 1371-430, 1978
- [26] L. L. Zheng, Z. Y. Jiang, E. Y. Yu, "Alpha spectral power and coherence in the patients with mild cognitive impairment during a three-level working memory task," *J Zhejiang Univ Sci B*. 8(8):584-92, Aug 2007
- [27] R. Fisher, "The use of multiple measurements in taxonomic problems," *Annals Eugen.* 7, 188, 179, 1936
- [28] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least Angle Regression," *Annals of Statistics* 32(2), 407-499. 2004
- [29] T. S. Jaakkola, and M. I. Jordan, "A variational approach to bayesian logistic regression models and their extensions," in *Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics*, 1997
- [30] J. del R. Millán, F. Renkens, J. Mouriño, W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human EEG," *IEEE Trans Biomed Eng.* 2004 Jun;51(6):1026-33.
- [31] F. Bach, "Exploring Large Feature Spaces with Hierarchical Multiple Kernel Learning," *Advances in Neural Information Processing Systems* (NIPS), 2008
- [32] F. Bach, R. Jenatton, J. Mairal and G. Obozinski, "Convex optimization with sparsity-inducing norms," in S. Sra, S. Nowozin, S. J. Wright., editors, *Optimization for Machine Learning*, MIT Press, 2011. (to appear)
- [33] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices" *J Multivar Anal.* 88(2): 365-411, 2004.
- [34] R. O. Duda, P. E. Hart, D. G. Stork, "Pattern Classification," Wiley Interscience, 2000
- [35] S. Lemm, B. Blankertz, T. Dickhaus, K.-R. Müller, "Introduction to machine learning for brain imaging," *Neuroimage*. Dec 2010 (in press)
- [36] R. Tomioka and K.-R. Mueller, "A regularized discriminative framework for EEG analysis with application to brain-computer interface," *Neuroimage*, 49 (1) pp. 415-432, 2010.
- [37] S. Makeig, M. Westerfield, T.-P. Jung, S. Enghoff, J. Townsend, E. Courchesne, T. J. Sejnowski, "Dynamic brain sources of visual evoked responses," *Science*, 295:690-694, 2002
- [38] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Stat. Soc. B* 68, pp. 49-67 2
- [39] M. Alamgir, M. Grosse-Wentrup, and Y. Altun, "Multitask Learning for Brain-Computer Interfaces," *Thirteenth International Conference on Artificial Intelligence and Statistics* (AISTATS 2010), MIT Press, Cambridge, MA, USA, pp. 17-24, 2010.
- [40] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," in *IEEE Transactions on Knowledge and Data Engineering* 22(10), pp.1345-1359, 2010