Brain Computer Interfacing: State of the Art, Probabilistic Advances and Future Perspectives

P. Sykacek

Department of Engineering Science,
University of Oxford
Oxford, UK
psyk@robots.ox.ac.uk

October 6, 2003

Abstract

Brain computer interface (BCI) research, carried out in our group, has focused on two aspects. In terms of signal processing, we focused on improving BCI performance by using advanced probabilistic machine learning approaches. On the other hand even the most advance signal processing can only recover information which is actually contained in the data. Hence a second central research direction of the project was concerned with the statistical evaluation of different neuro-cognitive experimental settings that were proposed by our clinical partners[1].

1 Introduction

Research on Brain computer interface (BCI) technology has been performed for more than a decade. The idea of BCI is to enable people with severe neurological disabilities to operate computers by thought rather than by physical means. A number of groups around the world are developing BCI systems,

[1]We would at this stage like to acknowledge the work of Prof. Stokes, director of the Research Department of the Royal Hospital of Neuro-Disability, Putney, London, for providing clinical expertise and data used in these evaluations

in which surface electroencephalography (EEG) is used to record the brain signals that control cursor movements on a computer screen ([1], [2], [3] and [4]). A good average case performance of such a system will provide bit rates in the range of 5-25 bit per minute ([5]). This is sufficient for non time critical communication tasks ([1]), but clearly not enough for high bandwidth real time applications. According to our expertise in signal processing and probabilistic modelling, the project thus focused on investigation of the signal processing part of BCI's and on statistical evaluations of neuro-cognitive design considerations suggested by our expert colleagues. In particular we put emphasis on:

- Careful methodological design considerations to evaluate the suitability of different neuro-cognitive settings for "driving" BCI's [6]. This study was at large an empirical evaluation of general methodological considerations, like optimal model choice to represent EEG and to model the probabilities of different cognitive states. The most suitable model was then used to evaluate different neuro-cognitive settings.

- Research towards *optimal and feasible* probabilistic models for BCI's. This aspect of BCI's proved to provide a translation algorithm which, evaluated on a large set of subjects, results on average in increased bit rates.

- Research towards a *truly adaptive* BCI. There are several accounts from findings of other BCI-groups, why this is necessary. Current BCI architectures used so far and currently developed by *all!* other research groups, rely on a *stationarity assumption* about how the EEG is generated during cognitive tasks. This assumption *must* be wrong for several reasons:

  - There are technical problems with the electrolyte fluids used for electrode placement. It simply dries out and thus changes impedance. Hence both signal amplitudes and dynamics are subject to temporal variations during a BCI session.
  - Both learning (habitation) effects and fatigue (Note at this point that sleep staging is entirely based on variations of EEG dynamics.) change the dynamics during and between BCI sessions.

  We thus derived and evaluated a fully adaptive approach for the translation algorithm that, even in short time use of a BCI, has proven to result in higher communication bandwidth than conventional static BCI's.

Table 1: Comparison of different tasks.

comparison	accuracy (1)	bit/s (1)	accuracy (2)	bit/s (2)	P_{null}
(a) vs. (b)	74 %	0.173	69 %	0.107	$\ll 0.01$
(a) vs. (c)	74 %	0.173	71 %	0.131	< 0.01
(a) vs. (d)	74 %	0.173	71 %	0.131	0.01
(b) vs. (c)	69 %	0.107	71 %	0.131	0.02
(b) vs. (d)	69 %	0.107	71 %	0.131	0.03
(c) vs. (d)	71 %	0.131	71 %	0.131	0.40

The advantage of probabilistic methods over other approaches to data driven research, like development of a BCI is the simplicity how they can be applied [7]. If implemented appropriately the inference results do not depend critically on any parameter settings. This is in contrast to classical approaches which require to fine tune parameters by trial and error.

2 Methodological and experimental achievements

2.1 Experimental issues

In order to assess the effects on the bit rates of BCI's, we compare in [8] the communication bandwidth we may achieve with different cognitive tasks [6]. We base comparisons on generalization accuracies obtained for independent test data. Differences are assessed for *statistical significance* using Mc. Nemars test, a test for analyzing paired results that can be found in [9]. In order to allow comparisons with other BCI systems, we also report bit rates as is suggested in [5]. The BCI experiments of this study were done by 10 young, healthy and untrained subjects. They are based on 3 cognitive tasks: an auditory imagination, an imagined spatial navigation task and an imagined right motor task. Each experiment consists of 10 repetitions of alternating pairs of these tasks each of which have been done for seven seconds. EEG recordings are obtained from two electrode sites: T4, P4 (right tempero-parietal for spatial and auditory tasks), C3' , C3" (left motor area for right motor imagination). The ground electrode is placed just lateral to the left mastoid process. An investigation of different classification paradigms reveals that on this data the BCI classifiers perform significantly better, when allowing for a nonlinear decision boundary. The method applied in this comparison uses autoregressive features (AR) extracted from successive segments of EEG. We use a *generative classifier* that predicts probabilities of cognitive states. Table 1 summarizes the results of this comparison. Task pairing (a) refers to the combination navigation - auditory, task pairing (b) refers to the combination navigation - right motor, task pairing (c) refers to the combination auditory - right motor and task pairing (d) refers to the combination left motor - right motor, which we include in order to allow for a comparison with these classical tasks. Our results allow to conclude that (a) vs. (b) result in slightly better correct classification rates as the classical imagined motor task. However, since we can extract information about the cognitive state in all cases, the main conclusion is that we might significantly increase the bit rate of BCI systems by using more than two cognitive tasks. For more details on this study we refer to [8].

2.2 Probabilistic models for improved static BCI systems

Probabilistic models can be used to describe many architectures that have been applied to static and adaptive BCI systems. Examples are Hidden Markov models, that have been successfully applied to BCI in ([10]). Probabilistic models have also been quite popular tools in the machine learning and statistics community. Recently these communities have investigated efficient algorithms that allow inference of very complex models. These findings are of interest for the BCI community since they allow us to go beyond classical time series models and by that improve different aspects of BCI systems. We have recently evaluated two such generalizations in the context of BCI systems. Coupled HMM's are generalizations of ordinary HMM's, where two hidden state sequences are probabilistically coupled using arbitrary lags. In ([11]) these models have been applied to movement planning and shown to outperform classical HMM's.

Another modification of HMM's was proposed in [12], where we follow probabilistic principles and suggest that classifications based on feature extraction (like the use of spectral representations or AR models as used in our BCI) have to regard the features as *latent variables.* Hence inference and predictions need to marginalize over this latent space. The practical advantage of the proposed architecture is that both the feature and the model uncertainty (the latter means the uncertainty about model order) are automatically taken into consideration. This effects model estimation and prediction and results in *automatic artefact moderation* and thus in *intelligent sensor fusion.* The idea exploits a property found by marginalisation (i.e. in-

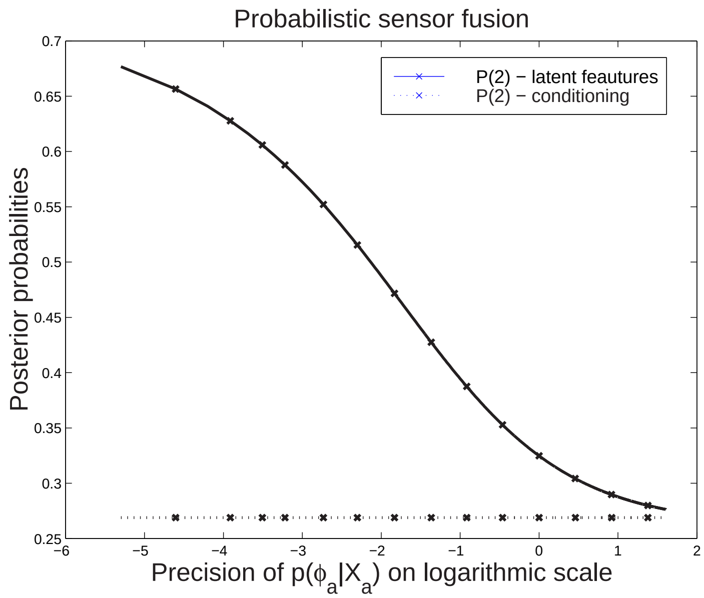

Figure 1: This Figure illustrates the dependency of the probabilities $P(y|\mathcal{X}_a, \mathcal{X}_b)$ on the precision of the latent space distribution $p(\boldsymbol{\varphi}_a|\mathcal{X}_a)$. In the case of probabilistic sensor fusion, the optimal decision about y depends, even without doubt option, on the precision of $p(\boldsymbol{\varphi}_a|\mathcal{X}_a)$.

tegration over the distribution) over (at least two) uncertain latent variables estimated from two sensor signals (e.g. EEG recorded at different electrode sites). Depending on the variance of the distribution, we will obtain different posterior probabilities. Figure 1 illustrates the effect using two sensor signals $\mathcal{X}_a$ and $\mathcal{X}_b$ and the state of interest (e.g. the cognitive state we want to predict) y. We see how the posterior probability depends on the variance. The effect may even result in a different assessment w.r.t. the predicted state.

The application of such a marginalisation idea to BCI is illustrated in table 2. We illustrate results obtained with this paradigm using two task pairings of the study reported in section 2.1. We compare the generalization accuracies of the fully probabilistic model (full Bayes) with those of a classical approach that does feature extraction separately. Table 2 also shows in brackets the probabilities of the null hypothesis that the result of one method are equal to the method in the previous column. We may thus conclude that a fully Bayesian approach significantly outperforms classifications obtained when conditioning on feature estimates. Despite having found that a fully Bayesian approach improves BCI performance, the proposed method has the disadvantage of not being directly applicable to *online* BCI. The computational complexity simply does not allow that. Hence we investigated an

Table 2: Generalization accuracies and bit rates for fully Bayesian method

	classical model		full Bayes	
task pair	gen. acc.	bit rate	gen. acc.	bit rate
(d)	75.9%	0.20	$81.4\%_{(0.04)}$	0.31
(a)	76.2%	0.21	$84.5\%_{(\ll 0.01)}$	0.38

approximation which can be used in real time and nevertheless achieve the desired effects [13].

2.3 Probabilistic models for improved adaptive BCI

Probabilistic models can also be of advantage in describing algorithms for adaptive BCI systems. A graph structure that illustrates such an approach is shown in figure 2. We assume a model that predicts the probabilities of cognitive states (y_n) and regard the parameters of the classifier are regarded as latent variables in a first order Markov process. The solution in a linear Gaussian case are the well known Kalman filter equations. In our case the nonlinearity introduced by predicting probabilities requires us to use an approximation. We suggest for that purpose a so called *variational* technique and thus obtain variational Kalman filtering as an inference method ([14]). Variational methods (e.g. [15] and [16]) are attractive for BCI systems because compared with Laplace approximations (as e.g. used for classification problems in [17]), they allow for more flexibility and as opposed to particle filters they still provide a parametric form of the posterior. Having a parametric posterior is important since it allows efficient implementations. Results applying the variational Kalman filter classifier to the BCI data (described in section 2.1) are summarized in table 3. These results suggest that a truly adaptive BCI (column vkf) significantly outperforms the equivalent static method (column vsi). A detailed description of our adaptive translation algorithm and a thorough evaluation can be found in [18].

3 Dissemination of Results

The publication effort for this project has been two fold. On one hand the problem of designing an optimal BCI translation algorithm required two methodological innovations. In order to make sure that we obtain feedback from the corresponding community, both methods were published in a high

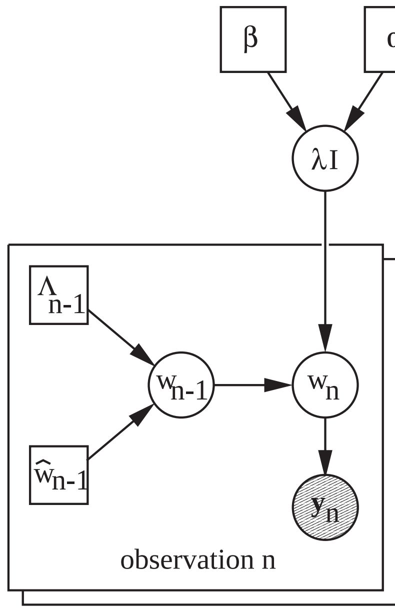

Figure 2: This figure illustrates adaptive inference as a directed acyclic graph. The coefficients of the classifier, $\boldsymbol{w}_n$, are assumed to be Gaussian, following a first order Markov process. The posterior at time $n-1$, $p(\boldsymbol{w}_{n-1}|\hat{\boldsymbol{w}}_{n-1}, \Lambda_{n-1})$, together with the process noise specified by $\lambda \boldsymbol{I}$ determine the prior for $\boldsymbol{w}_n$. The hyper parameter λ is given a Gamma prior specified by parameters α and β. In order to make inference of λ not too sensitive on the prior settings, we use a reasonable flat prior and assume a constant adaptation rate within a window of appropriate size.

profile conference ([12] and [14]), with a follow up published as a book chapter [13] and corresponding journal versions still being under review.

To get feedback from the BCI community, we published our suggestions for improved BCI translation algorithms in relevant journals [7, 8, 13]. Peter Sykacek gave a talk on "Probabilistic methods for BCI research" at the BCI workshop in Albany, NY, June 12-17, 2002. He presented there a poster on "Adaptive BCI by variational Kalman filtering" which obtained a best engineering award. He also gave a talk on the BUPA BCI project "Towards adaptive BCI" at the NCAF meeting on human computer interaction, 3-4 September 2003, Cambridge, UK.

Table 3: Generalization accuracies and bit rates of adaptive BCI

	Generalization results				
	vkf		vsi		
Cognitive task	acc.	bit/s	acc.	bit/s	P_{null}
navigation/auditory	0.86	0.42	0.83	0.34	0.02
navigation/movement	0.80	0.28	0.80	0.28	0.31
auditory/movement	0.78	0.24	0.76	0.21	0.00

4 Conclusion and future perspectives

The decision to investigate an adaptive translation algorithm proofed very fertile. We are convinced that such algorithms will play a central role in future BCI research. Encouraged by the results obtained with adaptive methods and the fully Bayesian methodology, our current research direction aims to combine the aspects of both methods. The idea is to obtain real time adaptive methods that achieve optimal bit rates and thus are useful as a robust communication device. The long term goal for this communication device is to use it along the lines of [1] as an aid for patients with neurological disabilities.

Acknowledgements

Peter Sykacek's research project on BCI was funded by grant Nr. F46/399 kindly provided by the BUPA foundation.

References

[1] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kubler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398:297–298, 1999.

[2] J. R. Wolpaw, D. J. McFarland, D. J. Neat, and C. A. Forneris. An EEG-based Brain-Computer Interface for Cursor Control. *Electroencephalography and Clinical Neurophysiology*, 78:252–259, 1991.

[3] G. Pfurtscheller, D. Flotzinger, and J. Kalcher. Brain-Computer Interface – a new communication device for handicapped people. *Journal of Microcomputer Applications*, pages 293–299, 1993.

[4] W. Penny, S. Roberts, E. Curran, and M. Stokes. EEG-based communication: a pattern recognition approach. *IEEE Trans. Rehab. Eng.*, pages 214–216, 2000.

[5] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan. Brain-Computer Interface technology: a review of the first international meeting. *IEEE Trans. Rehab. Eng.*, pages 164–173, 2000.

[6] E. Curran and M. Stokes. Learning to control brain activity: A review of the productuon and control of EEG components for driving brain-computer interface (BCI) systems. *Brain and Cognition*, 51:326–335, 2003.

[7] P. Sykacek, S. J. Roberts, M. Stokes, E. Curran, M. Gibbs, and L. Pickup. Probabilistic methods in BCI research. *IEEE Trans. Neural Syst. Rehab. Eng.*, pages 192–195, 2003.

[8] E. Curran, P. Sykacek, M. Stokes, S. Roberts, W. Penny, I. Johnsrude, and A. Owen. Cognitive tasks for driving a Brain Computer Interfacing system. *IEEE Trans. Neural Syst. Rehab. Eng.*, page to appear, 2003.

[9] B. D. Ripley. *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge, 1996.

[10] B. Obermeier, C. Guger, C. Neuper, and G. Pfurtscheller. Hidden Markov models for online classification of single trial EEG. *Pattern Recognition Letters*, pages 1299–1309, 2001.

[11] I. Rezek, M. Gibbs, and S. Roberts. Maximum a posteriori estimation of coupled hidden markov models. *Advances in Neural Networks for Signal Processing*, page to appear, 2002.

[12] P. Sykacek and S. Roberts. Bayesian time series classification. In T.G. Dietterich, S. Becker, and Z. Gharamani, editors, *Advances in Neural Processing Systems 14*, pages 937–944. MIT Press, 2002.

[13] P. Sykacek, I. Rezek, and S. J. Roberts. Bayes Consistent Classification of EEG Data by Approximate Marginalisation. In S. J. Roberts and R. Dybowski, editors, *Applications of Probabilistic Models for Medical Informatics and Bioinformatics*, page to appear. Springer Verlag, Berlin, 2003.

[14] P. Sykacek and S. Roberts. Adaptive classification by variational Kalman filtering. In S. Thrun, S. Becker, and K. Obermayer, editors, *Advances in Neural Processing Systems 15*, page to appear. MIT Press, 2003.

[15] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. In M. I. Jordan, editor, *Learning in Graphical Models*. MIT Press, Cambridge, MA, 1999.

[16] H. Attias. Inferring parameters and structure of latent variable models by variational Bayes. In *Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI–99)*, pages 21–30, San Francisco, CA, 1999. Morgan Kaufmann Publishers.

[17] W. Penny and S. J. Roberts. Non stationary logistic regression. In *Proceedings of the IJCNN 1999*, 1999.

[18] P. Sykacek, S. J. Roberts, and M. Stokes. Adaptive BCI based on variational Bayesian Kalman filtering: an empirical evaluation. *IEEE Trans. Biomed. Eng.*, page to appear, 2003.