Bayesian Approaches for adaptive Brain-Reading

Dennis Becker
Thesis advisor: Sirko Straube
Jan Hendrik Metzen

July 19, 2012

Abstract

Classification of event related potentials with a brain computer interface is a difficult task. The data are high dimensional, the signals differ in strength, time domain, from subject to subject, and even among different training sessions. Often, time-consuming training of the classifier and training sessions with the subject for the acquisition of a training set are necessary before good classification can be achieved and the actual task can be started. Conducted research indicates that a Bayesian classifier, trained on other subjects data, is able to perform classification without previously captured data of new subjects. This thesis proposes the implementation and evaluation of existing Bayesian approaches, and the development of a new one. Therefore it is planed to use an existing algorithm and extend it with generalisation to out-of-the-box classification and further adaptation to novel data.

1 Introduction

Human computer interfaces are in discussion for a long time and are constantly improved. One promising interface are the brain computer interfaces (BCI). With electrodes placed on the scalp, brain activity can be measured. Different neurophysiological signals are known and can be used to build a BCI. Some of them require training of the user while others do not. Brain-Reading (BR) can be considered as a subarea of BCI. With a BCI, the user gets a response for his actions. The brain is then likely to adapt to the feedback and as a result, the brainwaves may change. With a BR system, the user gets no feedback, therefore the brain cannot adjust. To detect the potentials correlated with a certain stimulus is quite difficult, so machine learning algorithms are applied. Here, new classifiers for each subject and session are trained. The final classifier is highly specific, because potentials differ slightly from subject to subject and even among different sessions. This makes it a difficult task. Furthermore, many machine learning algorithms, cross-validation is used to estimate the best parameters of the algorithm. Cross-validation is computationally expensive because the algorithm has to be trained many times with different parameters to find the ones that perform best. In Bayesian learning no time-consuming parameter estimation is needed. The methods regularization parameters are also estimated from the training. After estimating the posterior from the data, this is used to calculate the regularization parameters which fit best. All parameters become updated regularly and so no validation set for the estimation regularization parameter is needed. Another advantage is the probabilistic output of the algorithm together with the class label. Non-probabilistic algorithms solely output the class label without providing a probability. Hoffman showed in his thesis [3] that his algorithm only needed 15 minutes for training on a competition data set. In contrast the winner of this BCI competition data set needed up to two hours with comparable results in classification rate. Another possible benefit is the adaptivity of the learned model. It was shown by Alamgir et al.[1] that a classifier that was learned from a group of users can perform out-of-the-box classification for novel subjects and then adapt to new data. Out-of-the-box classification means that a classifier that was trained from task-related data of other subjects, can perform classification, considerably above chance level, for a novel unseen subject.

2 Prior Work

In the following, the chronologically listed work is the basis for this thesis.

2.1 Bayes' theorem

Bayes' theorem ($P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$) was named after Thomas Bayes. This work in the field of probability theory was published postmortem [2]. Laplace reproduced and extended Bayes' work without the knowledge of Bayes work. Bayesian statistic was used in geophysics in the 1960s and introduced to the machine learning community by MacKay in 1992 [6]. Bayes' theorem allows to calculate from the observation B back to the cause A. Therefore, it uses our prior believe in the cause $P(A)$ and the likelihood $P(B \mid A)$ that explains our observations best. The multiplication of both results in the posterior $P(A \mid B)$, but it has to be rescaled to a probability distribution with the term P(B) which is prior probability for the event B .

2.2 Application of the Evidence Framework to Brain-Computer Interfaces (Hoffmann at al., 2004)

Hoffmann et al. [4] presented a modified version of the Linear discriminant analysis (LDA). Since LDA is a powerful algorithm in the area of BCI, they extend LDA with the Bayesian framework. In 2007, Hoffmann published his PhD Thesis [3] about this work. He named the algorithm Bayesian Discriminant Analysis. The advantage of Bayesian model selection is that the algorithm has to be fitted only once, in contrast to algorithms that use cross-validation for parameter estimation. The probabilistic class output of the algorithm can be another advantage.

2.3 Adaptive Multitask Learning (Alamgir et al., 2010)

Alamgir et al. demonstrated in a paper[1] how to use a prior that was learned from the training sets from different subjects to classify P3 evoked potentials out of the box for new subjects with satisfactory classification performance. To achieve this, two steps are performed. In the first step a shared prior was learned from the data. For this a regression model was used where the parameters are estimated with cross-validation. Afterwards, the resulting shared prior is used for classification and adaptation to novel subjects. The adaptation can be done either in batch processing of the new data or in online mode one-by-one. The previous learned Bayesian prior reduces the time required for the training session, which is desirable since training sessions are time-intensive and can be exhausting for the subjects. Figure 1 illustrates the basic idea behind the adaptive Bayesian approach.

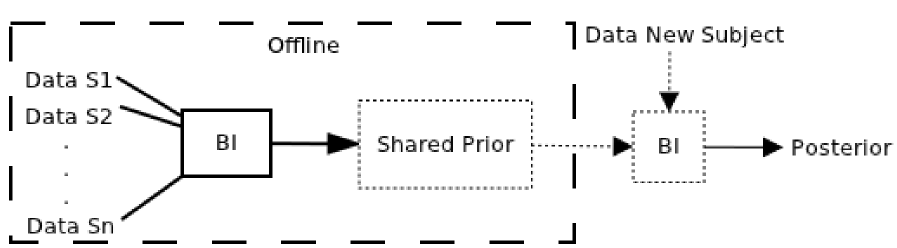

Figure 1: Schematic Bayesian Multitask learning

The data from different subjects are used to estimate a Bayesian prior. First, the Bayesian framework considers all models possible. In the presence of data it then reduces the probability of all models that do not explain the observed data. The left part of the figure (Fig. 1) can be viewed as the offline part. In this part the Bayesian algorithms becomes trained with all data sets from different subjects. Different algorithms and models for the posterior estimation can be used. The remaining part of the algorithm is responsible for classification and adaptation to novel subjects. It can be seen as the online part. In this part it is beneficial to perform a quick update procedure for the classifier. Alamgir presented a novel task update that adapted to the subject and abandoned computationally expensive operations.

2.4 Other Work

Two different algorithm comparison papers[7, 5] confirm the efficiency of Bayesian approaches in the field of BCI, and that for BCI classification, linear classifiers are sufficient.

3 Thesis

The aBRI-group (adaptive Brain-Reading Interface) conducts research on Brain-Reading Interfaces. Presently, for parameter estimation cross-validation is used. This thesis should answer the question if a Bayesian algorithm can be advantageous to the currently used support vector machine in the aBRI-group and facilitate the possibility of out-of-the-box classification. It is intended to avoid cross-validation during the training phase as well as computationally expensive operation during subject adaptation. The use and limitation of different priors will be investigated as well as their computational feasibility.

WP 1) Literature research / State of the Art

Literature research is necessary to review the currently used classifiers in the field of BCI. The Bayesian approaches in this field will be emphasized, and it will also be pointed to other research fields where Bayesian approaches are used with success. In addition, it will be useful to give an introduction into this section of machine learning.

WP 2) Implementation of Bayesian algorithms

Existing Bayesian approaches should be implemented and evaluated as baseline for further research. This basis is needed to compare the performance in terms of calculation speed and classification rate of new methods. The Bayesian Discriminant Analysis (BDA) will then also be used for further modification, to achieve out-of-the-box classification and adaptation.

Algorithms to be implemented:

- Naive Bayes is very simple classifier, where all attributes are considered to be independent given the class.
- Bayesian Discriminant Analysis. There is a freely available implementation in Matlab and also the used data sets are on the internet. This algorithm has to be ported to python and extended for the ability of out-of-the-box classification and adaptation to novel subjects.

WP 3) Out of the box classification and online adaptation

The main objective is to gain out-of-the-box classification for unseen subjects and further adaptation of the classifier to the subject. The algorithm proposed by Alamgir [1] seems to be a good basis. But the offline part is defined for a regression problem that relies on cross validation for parameter estimation. In this thesis, it will be adapted to classification based on Bayesian LDA. This should combine the advantages of faster learning in training, out-of-the-box classification and further adaptation to novel subjects.

WP 4) Evaluation with the aBRI framework

To make a statement about the performance, the algorithm will be compared in the aBRI-framework against the currently used classifiers. Data sets from the aBRI-Group are already available. It is to evaluate the classification rate among the different classifiers. The classification rate with further adaptation, is to compare to the rate without further adaptation.

It is to examine, if the out-of-the-box classification provides already a satisfying classification rate, and if not how many data samples are necessary to achieve this classification rate. Also, these Bayesian classifier, with and without the possibility for adaptation, will be compared to more traditional algorithms like LDA and SVM without the probability of out-of-the-box classification and adaptation.

WP 5) Discussion of applicability

After the implementation and evaluation of the new algorithm, possible benefits and drawbacks will be discussed. The difference between the new Bayesian classifier and the currently used SVM will be presented in order to discuss their limitations and applicability. Further, it will be exposed, if a probability for a class label is superior to a fixed class label output, and the number of classes the algorithm can cope with.

4 Costs

No costs are incurred.

References

[1] M. Alamgir, M. Grosse-Wentrup, and Y. Altun. Multitask learning for brain-computer interfaces. In M. T. Teh, Y.W., editor, *JMLR Workshop and Conference Proceedings Volume 9: AISTATS 2010*, pages 17–24, Cambridge, MA, USA, 5 2010. Max-Planck-Gesellschaft, JMLR.

[2] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. of the Royal Soc. of London*, 53:370–418, 1763.

[3] U. Hoffmann. *Bayesian machine learning applied in a brain-computer interface for disabled users*. PhD thesis, Lausanne, 2007.

[4] U. Hoffmann, G. Garcia, J. Vesin, and T. Ebrahimi. Application of the Evidence Framework to Brain-Computer Interfaces. In *Proceedings of the 26th IEEE Engineering in Medicine and Biology Conference*, Motion analysis and image sequence processing. IEEE, 2004.

[5] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw. A comparison of classification techniques for the P300 Speller. *Journal of Neural Engineering*, 3(4):299–305, Dec. 2006.

[6] D. J. MacKay. *Bayesian Methods for Adaptive Models*. PhD thesis, California Institute of Technology Pasadena, California, 1992.

[7] N. V. Manyakov, N. Chumerin, A. Combaz, and M. M. V. Hulle. Comparison of classification methods for p300 brain-computer interface on disabled subjects. *Computational Intelligence and Neuroscience*, 2011.