Proceedings of the 1st International IEEE EMBS
Conference on Neural Engineering
Capri Island, Italy • March 20-22, 2003

Automated EEG Feature Selection for Brain Computer Interfaces

Michael Schröder, Martin Bogdan
and Wolfgang Rosenstiel
Wilhelm Schickard Institut für Informatik
Universität Tübingen, Germany
Email: schroedm@informatik.uni-tuebingen.de
bogdan@informatik.uni-tuebingen.de
rosenstiel@informatik.uni-tuebingen.de

Thilo Hinterberger and Niels Birbaumer
Institut für medizinische Psychologie
und Verhaltensneurobiologie
Universität Tübingen, Germany
Email: thilo.hinterberger@uni-tuebingen.de
niels.birbaumer@uni-tuebingen.de

Abstract— **A Brain Computer Interface (BCI) utilizes signals derived from electroencephalography (EEG) to establish a connection between a person's state of mind and a computer based signal procession system that interprets the EEG signals. The choice of suitable features of the available EEG signals is crucial for good BCI communication. The optimal set of features is strongly dependent on the subjects and on the used experimental paradigm. Based upon EEG data of an existing BCI system we present a wrapper method for the automated selection of features. The proposed method combines a genetic algorithm (GA) for the selection of feature with a support vector machine (SVM) for their evaluation. Applying this GA-SVM method to data of several subjects and two different experimental paradigms, we show that our approach leads to enhanced or even optimal classification accuracy.**

Keywords - **BCI, feature selection, EEG classification, genetic algorithm, support vector machine**

I. Introduction

Signals derived by electroencephalographic (EEG) recordings of the scull surface reflect the electric activity of large neuron ensembles of the cortex. Most current research projects on the field of Brain Computer Interfaces (BCI) utilize these signals to detect distinguishable brain states of human subjects.

A. Feedback Training and BCI Systems for Communication

The Thought Translation Device (TTD) [1] is a realtime BCI system developed for severely paralysed patients who are not in control of any voluntary motor actions, e.g. patients suffering from amyotrophic lateral sclerosis (ALS). In an operant conditioning paradigm using visual, auditive or haptic feedback, those Locked-In patients are trained with the TTD in order to gain voluntary control over certain characteristics of their EEG-signals. The underlying feedback training principles are well established and have been used since the 1960s for the self regulation of muscle tone and blood pressure, for the treatment of migraine, chronic or phantom pain, the attention deficit hyperactivity disorder and for the prediction and avoidance of epileptic seizures [2]. Using the TTD system, trained Locked-In patients can voluntarily generate EEG signals which are interpreted by the TTD as bits of communication. A built-in spelling device enables the patients to communicate with their social environment or to control external devices.

Studies [1] have shown, that the self control of EEG signals can be learned solely based on feedback training. Nevertheless, introducing learning capabilities also for signal processing and classification offers advantages: The often time consuming feedback training of the patient can undergo a significant initial speedup when machine learning algorithms are applied.

The discrimination of brain states is performed in general by the classification of two or more types of EEG signals. This is a difficult task for a BCI system as single trials have to be classified which may contain various artifacts and intensive background activity.

B. Feature Selection for BCI Classification

In a BCI experiment every single trial delivers time series from several EEG channels that are sampled with at least 100Hz. In order to avoid dealing with this very high dimensional and noisy data, certain features can be selected or calculated from the all-channel time series before the classification starts. Ideally, those features meaningful for classification are identified and chosen, while others (including outliers and artifacts) are omitted. In existing BCI systems, various features are used. A common way of selecting them is based on physiological expert knowledge or a priori expectations that strongly depend on the nature of the imagination tasks the subject performs. In a feedback paradigm the pure or band pass filtered time series [1] of slow cortical potentials (SCP) can be used without asking the subject to perform a specific imagination task. The band power coefficients of (motor) rhythms [3] are used for the discrimination of motor imagination tasks. Coefficients derived from AR or AAR models of the time series or principal components are used for various mental imagination tasks [4].

The choice of a subset of all available EEG channels also affects the classification performance. This choice usually precedes the calculation of any features, but in a broader sense the choice of channels itself can be considered a feature selection. Usually the assortment of appropriate channels is

0-7803-7579-3/03/$17.00 © 2003 IEEE

either physiologically justified or it is calculated based on the statistics of example data, e.g. using the method of common spatial patterns [5]. For different applications, variants of independent component analysis were successfully applied for this spatial filtering of EEG data.

Once the feature vectors are calculated, the actual classification problem can be solved by linear or kernel discriminant analysis [6], artificial neural networks [7], decision trees and various other methods.

In this paper we propose a feature selection approach based on a genetic algorithm (GA) to pick most promising channels of EEG signals for the classification via support vector machines (SVM). The next section will describe the data used for our experiments before the GA-SVM method is explained in detail. The results of this method are then compared to physiologically motivated feature selection methods and - where applicable - also to the brute force choice of channels.

II. Data

A. Data Sets Recorded with the TTD

This data was recorded during feedback training sessions of four healthy subjects (VP03, VP16, VP17 and VP18) using the TTD system. The subjects were asked to move a feedback cursor up and down on a computer screen, while their cortical potentials and EOG were recorded using a PsyLab EEG8 amplifier. The data was sampled by a Computer Boards PCIM-DAS1602/16 bit A/D-converter with 256 Hz sampling frequency. Previous work [1] has shown that healthy subjects as well as Locked-In patients are in general able to gain voluntary control of their slow cortical potentials (SCP) by feedback training. During the recording, the subjects received visual feedback of their SCP taken from the *Cz-Mastoids*: cortical positivity lead to a downward movement of the feedback cursor, whereas cortical negativity lead to an upward movement. The following channels of EEG were recorded (denotation follows the 10/20 system):

- Ch1: A1(left mastoid)-Cz
- Ch2: A2(right mastoid)-Cz
- Ch3: (2 cm frontal of C3)-Cz
- Ch4: (2 cm parietal of C3)-Cz
- Ch5: (2 cm frontal of C4)-Cz
- Ch6: (2 cm parietal of C4)-Cz

After EOG correction, the position of the feedback cursor was calculated based on the average potential of Ch1 and Ch2. Channels 3 to 6 were not used for feedback. The subjects were trained during at least five sessions (on different days). Every session covered several runs of 50 trials each. Positive and negative trials were presented in random order. One trial lasted 6s. From second 0.5 until the end of the trial, the associated task was visually presented by a highlighted goal at either the top or the bottom of the screen in order to indicate either negativity or positivity. The visual feedback was presented from second 2 to second 5.5.

The GA-SVM feature selection analysis described in this paper was performed offline. Only the feedback interval of 3.5s duration of every trial was considered for our experiments. Thus the data of every trial resulted in 6 channels containing 896 samples each. Sessions that showed strong EOG activity were omitted. No further artifact removal was performed. The binary classification of positive and negative trials was performed separately for every patient and every session. Table I shows the composition of the session data.

TABLE I
OVERVIEW OF THE TTD DATA SETS

Subject	used session	total trials	pos. trials	neg. trials
VP03	S1	350	176	174
	S2	350	176	174
	S5	150	75	75
VP16	S1	300	151	149
	S3	250	126	124
VP17	S1	150	73	77
	S5	250	124	126
VP18	S3	200	101	99
	S5	300	151	149

B. NIPS 2001 Data Set

This data set was submitted by Blanckertz et al. [8] as part of the classification contest of the Neural Information Processing Systems (NIPS) Conference in 2001. The 27-channel EEG data was recorded in a free running left-right finger tapping paradigm of a healthy subject and sampled with 100Hz. Every trial contains 27 time series of 1.51s duration each. The data set comprehends 413 trials with 194 trials belonging to class 1 and 219 to class -1.

III. Algorithms

A. SVM Classification

The binary classification of any data set was carried out by the mySVM implementation of the support vector machine (SVM) which is based on the SVMlight algorithm [9]. The SVM as a classifier was chosen for several reasons. First, the fine tuning of learning parameters turned out to be quite simple for the TTD and NIPS data. Second, the SVM can deal with high dimensional data. Third, the SVM is a very reliable classifier that - given a certain function class - finds the best class separating function, i.e. the one that will have the lowest expected classification error on further data of the same kind. For an introduction into support vector learning see [6]. All SVM results presented are averaged values based on cross validation and the dot product kernel. Further mySVM parameters have been optimised beforehand but were kept constant for all results.

For our purposes, the channels of the EEG data are considered interesting features. The feature selection determines for each of the channels if its time series is used for classification or not. A possible outcome of such a choice of features can be coded in a binary string. It contains as many elements as channels are available. For the 6-channel TTD data described above, the example string '100011' would code for the selection of channels 1, 5 and 6 whereas the time series of channels 2, 3 and 4 are not used for classification. Likewise a binary string for the NIPS data contains 27 elements.

B. GA Feature Selection

For evolutionary optimization with a genetic algorithm (GA) [10], such a binary string can be considered the genetic information of an individual. Initially a whole population consisting of several individuals are randomly initialised. For simplification, the characteristics of such an individual is completely determined by its string. A so called fitness value is assigned to every individual by generating the corresponding data set and classifying it. Therefore, the time series of selected channels (defined by ones in the string) are concatenated, the '0'-channels are omitted. The resulting data set is then used for SVM training. The classification accuracy is considered the fitness of the individual.

After the fitness values of every individual of a population has been calculated, the next generation of individuals (i.e. the next population) can be created. This next population is derived from the previous one by applying the evolutionary operators *selection, recombination* (with cross over) and *mutation*. They operate directly on the binary strings and imitate the sexual reproduction. The chance of reproduction for an individual of the parent generation is correlated to its fitness value. Thus the genetic strings of very fit individuals are likely to influence the strings contained in the next generation whereas the strings of individuals with lower fitness are not likely to do so. For the genetic algorithm used on the TTD data sets, we chose populations of 13 individuals each and evolved them over 40 generations. Due to the higher number of channels in the NIPS data set a population size of 29 individuals was chosen and evolved for 80 generations. The selection scheme of both settings allowed elitism, i.e. the fittest individual was taken over directly into the next generation.

C. Expert Feature Selection

For the TTD training paradigm there exists a clear expert choice of channels that is used for the feedback training. This choice comprehends the first two channels omitting channels 3 to 6. This choice is physiologically motivated and has proven to be very successful in the long run when subjects or patients are trained over many sessions. For the special situation of the first few training sessions, this choice must not necessarily be optimal as the subjects usually have not yet acquired good control of their SCP signals. For the NIPS data set with 27 EEG channels, no such physiologically motivated expert rating was available.

IV. Experiments

The GA optimises the feature selection as it evolves good combinations of EEG channels. Since it is wrapped around the SVM which classifies and thus calculates the fitness values of the individuals, this GA-SVM combination is called a wrapper method for feature selection. In order to make this wrapper method practical, all the fitness values of individuals that belong to one generation are evaluated in parallel on a linux cluster. The fitness of every individual is determined by applying 10-fold cross validation during SVM training.

The GA-SVM terminates with a selection of good individuals, e.g. channel combinations that produce high classification scores in SVM training. The classification accuracy of the best GA individual (from now on referred to as *G*) is investigated in more detail: After the channel information was reduced to the channels specified by the according string, the trials contained in the data set were permuted ten times. For each permutation, a 11-fold and a 13-fold cross validation SVM training was performed. Thus for every G-individual twenty cross validations were calculated. This corresponds to 240 single SVM evaluations. In section V we show the average accuracy values of these cross validations. The classification performance of the physiologically motivated 2-channel expert choice for the TTD data (from now on referred to as *P*) was investigated in the same manner.

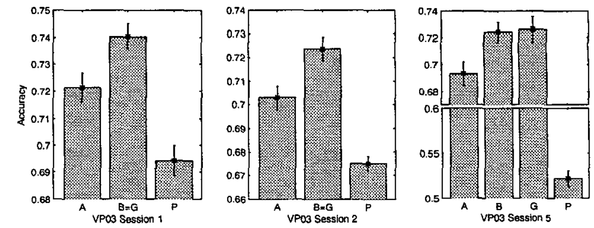

Fig. 1. Classification accuracy for 3 sessions of VP03. Boxes show average accuracy, lines show confidence intervals for $p < 0.05$. Comparison of channel selection modes *all channels* (A), best channels found by *brute force search* (B), good channels *determined by GA* (G) and *physiologically determined* channel selection (P). If B results were identical to G results then only one box (B=G) is shown. Note the different scalings of the Y-axis (accuracy).

For the case of the TTD data, which provides only 6 channels, the total number of possible channel combinations is $2^6 - 1 = 63$. As this search space is rather small, we additionally performed a brute force test. For the evaluation of every possible channel combination (including the all-channel combination *A*, a 10-fold cross validation was performed. Based on these results, the best brute force individual (from now on referred to as *B*) was chosen. The accuracy values for the A and B channel choices were also consolidated by ten 11-fold and ten 13-fold cross validations.

V. Results

A. TTD data sets

Over all subjects and sessions (see figures 1, 2 and 3) the classification accuracy of the all-channel training (A) is significantly better than for the physiologically motivated P-choice. On average, an improvement of 8.11% absolute accuracy was gained. As this result is derived from data that was recorded during the first five sessions only, no conclusions about the subsequent feedback training sessions can be drawn, although the initial training with the TTD system might be enhanced by the all-channel choice.

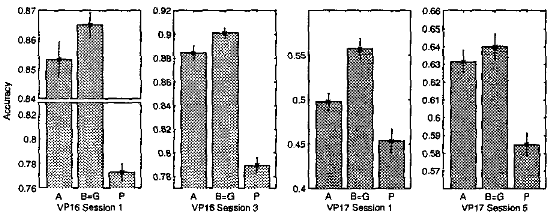

Fig. 2. Classification accuracy for 2 sessions of VP16 and VP17. Boxes show average accuracy, lines show confidence intervals for p <0.05. Note the different scalings of the Y-axis (accuracy).

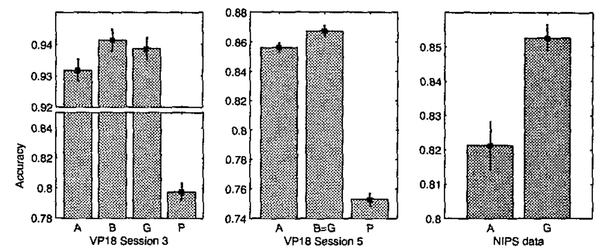

Fig. 3. Classification accuracy for 2 sessions of VP18 and for the NIPS data set. Boxes show average accuracy, lines show confidence intervals for p <0.05. For VP18 Session 3 the GA could not detected the best possible channel combination. Note the different scalings of the Y-axis (accuracy).

On average, the best individuals generated by the GA-SVM for the nine data sets use 3.67 channels while the nine best individuals generated by the brute force search use 3.78 channels. None of these individuals uses all 6 channels. Only 2 individuals contain (besides other channels) both P-channels 1 and 2. Except for two cases, the channel combinations calculated by the GA-SVM are identical with the best channel combinations derived from brute force search under all possible combinations. The best GA-SVM individual of VP03 Session 5 is not identical with the best brute force individual. Although not statistically significant, its average accuracy is even higher than that of the best brute force individual. This difference is explained by the fact that the final average accuracy values plotted in figure 1 are based on 20 cross validations whereas the brute force algorithm and the GA search internally only used one cross validation. The data of VP18 session 3 in figure 3 is the only example where the GA-SVM is outperformed by the brute force algorithm by 0.27% (although this difference is also not statistically significant). Compared to the all-channel choice, the GA-SVM choice on average improves the classification accuracy over all nine TTD data sets about 2.08%, and the comparison of the GA-SVM with the physiologically motivated channel choice thus reveals an average improvement of 10.19%. Please note that the best GA individuals are significantly better than the all-channel choice for seven of the nine data sets.

B. NIPS data set

The best individual that resulted from the GA-SVM method applied to the NIPS data set only used 13 out of 27 channels. Compared to the all-channel selection, the average classification accuracy was significantly improved by the GA-SVM. The wrapper method raised the overall classification accuracy for 3.15% to 0.8527. A comparison of the corresponding confidence intervals for $p<0.05$ is shown in figure 3.

VI. Conclusion

The presented results base on the time series of EEG recordings that contained various artifacts as well as redundant information. Although no hand optimised data preprocessing or feature definition steps were undertaken, the presented GA-SVM method allowed for good improvement of classification accuracy compared to the all-channel feature choice and - even better - compared to the physiologically motivated feature choice. As the TTD data sets are simple enough to allow the application of a brute force search, we could show that the GA-SVM feature selection is optimal for most of the studied sessions. Also for the NIPS data set that possessed more channels the application of GA-SVM achieved a significant improvement although the overall classification performance level for the all-channel selection was already quite high.

In future experiments, the GA-SVM method will be tested with more high-level features than time series and compared to other algorithms for feature selection. For the TTD paradigm it is also an interesting question if the optimal choice of features derived for one session stays a reliable and stable choice for following training sessions. Our further research work will investigate on this question.

References

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

[2] B. Kotchoubey, U. Strehl, C. Uhlmann, S. Holzapfel, M. König, W. Fröscher, V. Blankenhorn, and N. Birbaumer, "Modification of slow cortical potentials in patients with refractory epilepsy: A controlled outcome study," *Epilepsia*, vol. 42, no. 3, pp. 406–416, 2001.

[3] J. del R. Millan, M. Franze, J. Mourino, F. Cinotti, and F. Babiloni, "Relevant eeg features for the classification of spontaneous motor-related tasks," *Biological Cybernetics*, vol. 86, pp. 89–95, 2002.

[4] G. Pfurtscheller, C. Neuper, A. Schlögl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Transactions on Rehabilitation Engineering*, vol. 6, pp. 316–325, 1998.

[5] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time eeg analysis with subject-specific spacial patterns for a brain-computer interface (bci)," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 447–456, December 2000.

[6] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–202, March 2001.

[7] M. Schröder, M. Bogdan, W. Rosenstiel, T. Hinterberger, U. Strehl, and N. Birbaumer, "Online classification of eeg signals using artificial neural networks for biofeedback training of patients with epilepsy," in *9th International Workshop on Systems, Signals and Image Processing.* World Scientific Press, 2002, pp. 438–446.

[8] B. Blankertz, G. Curio, and K.-L. Müller, "Classifying single trial eeg: Towards brain computer interfacing," in *Advances in Neural Information Processing Systems (NIPS01)*, T. Diettrich, S. Becker, and Z. Ghahramani, Eds., vol. 14. MIT Press, 2002.

[9] T. Joachims, *Learning to Classify Text Using Support Vector Machines.* Kluwer, 2002.

[10] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, 1989.