

Classification of Executed and Imagined Motor Movement EEG Signals

Jason Sleight, Preeti Pillai, Shiwali Mohan

{jsleight, preetijp, shiwali}@umich.edu

Computer Science and Engineering, University of Michigan, Ann Arbor

December 17, 2009

Abstract

Electroencephalography (EEG), which contains cortical potentials during various mental processes, can be used to provide neural input signals to activate a brain machine interface (BMI). The effectiveness of such an EEG-based prosthetic system would rely on correct classification of executed motor signals from imagined motor movement signals; an executed motor signal should initiate movement in the artificial limb while signals from motor imagery should be filtered out. This work evaluates the performance of features based on average EEG signal power contained in different frequency bands in order to distinguish between the executed and imagined EEG signals. We also investigate Independent Component Analysis (ICA) as a scheme to remove irrelevant artifacts from the EEG signals. Results demonstrate that using EEG for classification can be performed effectively; however results vary significantly from patient to patient, suggesting that BMI must highly specialized for an individual patient.

1 Introduction

Brain-machine interface (BMI) has recently emerged as an important and interesting topic of research with widespread applications. It has evolved as a novel technique in assisting people to communicate with their environment or activate certain devices by brain signals. Currently there are two main methods for accomplishing BMI: use electroencephalography (EEG) signals, or Electrocorticography (ECoG), which measures the electrical activity of the brain taken from beneath the skull. However, the latter method is accompanied by a host of risks and complications; therefore most BMI research is currently focused on EEG signals. In this method, electrodes are placed on the scalp, and the brain's electrical activity is monitored and recorded. By using the EEG signals, classification algorithms can be developed to detect distinct brain functions.

Within the task of detecting brain functions, a special set of problems concerning motor movement has proven to be one of the most active research areas (largely due to the immediate applications in prosthetics). Correct classification of executed motor signals and neural activity arising due to imagining movement is important in such applications because individuals may use imagined motor movement for certain cognitive spatial reasoning tasks like assessing if they can reach a cup placed on the nearby table. The effectiveness of BMI applications like prosthetics counts on two factors - extracting distinguishable neural patterns from the EEG signal and building efficient classifiers. This study is aimed at:

1. Reliable feature extractions from an EEG signal using Independent Component Analysis (ICA) and Channel Selection and,
2. Classification of features using Support Vector Machines to determine whether classification can be performed globally across a population, or must be done individually on a per patient basis.

The rest of the paper is organised as follows - Section 2 discusses previous work done with motor imagery datasets, some terms and machine learning methods popular in the BMI community are introduced in Section 3. Section 4 describes our experimental setup, the data, and the features and the classifiers explored. We present and discuss results from our experiments in Section 5. We conclude with a discussion on our observations and the future scope of this work in Section 6.

2 Related Work

Classification of motor movements is an important problem in the field of BMI due to the immediate applications in prosthetics. As such, a significant amount of research has gone into classification methods for distinguishing between various motor movements (i.e. between moving your left or right hand) [1] by using executed motor movement EEG signals. Additional research has also been performed in using imagined motor movement in EEG signals to classify between various motor movements [2,3]. These experiments differ from our work in that we focus on distinguishing between the executed motor movement and the imagined motor movement EEG signals irrespective of the task being performed, while previous work treated motor and imagined tasks identically. We feel this is an important distinction, since as BMI becomes more detailed and complex, imagining a movement and performing a movement should yield two distinct actions (i.e. imagining movement should not produce movement).

Other work [4], has shown that executed and imagined motor movements have a similar, but distinct neural activity. This motivates the necessity of a sophisticated learning mechanism to accurately classify between executed and imagined motor movement.

3 Background

3.1 Independent Component Analysis

ICA [5] is very closely related to blind source separation (BSS) or blind signal separation problems. We assume that we observe n linear mixtures x_1, \dots, x_n of n independent components

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n, j = 1, n$$

It is assumed that each mixture x_j as well as each independent component s_i are random variables. It is also assumed that both the mixture variables and the independent components have zero mean [6]. If not, subtracting the sample mean can always center the observable variables .

Let x be the random vectors whose elements are the mixtures and let s be the random vector with the components s_1, \dots, s_n . Let A be the matrix containing the elements a_{ij} . The model can now be written:

$$x = As$$

or

$$x = \sum_{i=1}^n a_i s_i$$

The statistical model above is called independent component analysis or ICA. The ICA model is a generative model - it describes how the observed data are generated by a process of mixing the components s_i . The independent components are latent variables - they cannot be directly observed. The mixing matrix is also assumed to be unknown. ICA estimates the mixing matrix A and independent components s , given an observation x . ICA assumes that the components s_i are statistically mutually independent and have non-Gaussian distribution. The unmixing matrix W is an inverse of the matrix A and independent components can be obtained by

$$s = Wx$$

The Infomax algorithm is one way to find the unmixing matrix W by means of maximizing the entropy of the outputs, as proposed by Bell and Sejnowski [6].

The EEG signal obtained from scalp electrodes is composed of electrical potentials arising from several sources. Each source (including separate neural clusters, blink artifacts, or pulse artifact) project a unique topography onto the scalp. The recorded EEG signal is essentially a mixture formed by linear superposition of these electric potentials. ICA attempts to reverse the superposition by decomposing the recorded EEG into mutually independent components. Although little is known about the nature of brain activity, it is accepted in the EEG research community that EEG signals are non-Gaussian [7]. ICA has been successfully used to remove irrelevant artifacts from the EEG signals in EEG classification research [3].

4 Experimental Paradigm

4.1 Data

We have used the EEG Motor Movement/Imagery Dataset recorded using BCI2000 [8] instrumentation system available through Physionet [9]. This data consists of over 1500 one- and two-minute EEG recordings evenly taken from 109 subjects. Subjects performed different executed/imagined motor movement tasks as EEG activity was recorded on 64 channels placed on their scalp as per the standard international 10-10 system. Each subject performed 12 experimental runs: three two-minute runs of each of the four following tasks:

1. A target appears on either the left or the right side of the screen. The subject opens and closes the corresponding fist until the target disappears. Then the subject relaxes.

2. A target appears on either the left or the right side of the screen. The subject imagines opening and closing the corresponding fist until the target disappears. Then the subject relaxes.

3. A target appears on either the top or the bottom of the screen. The subject opens and closes either both fists (if the target is on top) or both feet (if the target is on the bottom) until the target disappears. Then the subject relaxes.

4. A target appears on either the top or the bottom of the screen. The subject imagines opening and closing either both fists (if the target is on top) or both feet (if the target is on the bottom) until the target disappears. Then the subject relaxes.

Annotation information is available on the 65th channel. Each channel is annotated in the following format using three codes:

- T0 corresponding to the resting period
- T1 corresponding to the onset of motion (real or imagined) of
 - the left fist (Tasks 3, 4, 7, 8, 11, and 12)
 - both fists (Tasks 5, 6, 9, 10, 13, and 14)
- T2 corresponds to onset of motion (real or imagined) of
 - the right fist (Tasks 3, 4, 7, 8, 11, and 12)
 - both feet (Tasks 5, 6, 9, 10, 13, and 14)

We used these annotations to partition each experimental run into 14 data segments which contained only the portions of the experimental run consisting of T1 or T2 segments.

Inaccuracies in the Physionet data - From some preliminary results of our experiments, we observed that for some patients the EEG channels were not annotated as specified in the experiment. Hence from the original 109 patients, we ran our experiments on the correct data obtained from 103 patients.

Summarizing our data hierarchy/terminology, we have data from 103 patients, each patient has 12 experimental runs (also called experimental trials), each experimental run has 14 data segments, and each data segment consists of a 64 channel EEG signal. Each data segment is treated as an individual sample for classification. This yields a total of 17,304 samples.

4.2 Preprocessing

4.2.1 Removing artifacts using ICA

Each pre-processed epoch was arranged across m channels ($m = 64$) and n sampled points ($n = 20000$) into a $m \times n$ matrix X . The i th row contains the observed signal from i th EEG channel, and the j th column vector contains the observed samples at the j th time point across all channels. In the present study, all ICA decompositions were performed using the InfoMax ICA algorithm in the EEGLab [10] toolbox available for processing EEG data. In available related literature [3], it is common for EEG classification experts to hand pick the ICA components based on heuristic knowledge. Since

we lack the intuition to select the relevant ICA components based on their structure and nature we had to rely on the algorithm itself. The implementation of InfoMax algorithm that we used orders the components in decreasing order of variance accounted for by their projections onto the scalp. We selected the first 20 ICA components that accounted for more than 99% of the total variance.

We reconstructed the channel information from the selected ICA components by multiplying the ICA components matrix with the mixing matrix (W^{-1}). This reconstructed signal was used for further processing and eventual classification. This procedure, in effect, removes the components with lower variances of their projections on the scalp. Since we assume that noise contributes a lower variance to the original signal, the reconstructed signal now can be thought to have fewer noisy artifacts.

4.2.2 Channel Selection

Initially, we experimented with different combinations of EEG channels to use as input to our classification algorithm, namely: using all 64 channels, using only channels C3, C4, and Cz of the international 10-20 format, and using eigenvectors corresponding to the five highest valued eigenvalues obtained from a PCA analysis. The reasoning behind each of these selections is as follows:

The reasoning behind each of these selections is as follows:

- By using all 64 channels, we provided our classification algorithm with the maximal amount of information. The classification algorithm should theoretically be able to sort out the relevant channels (those which contain information important for discriminating between executed and imagined motor movements) from the irrelevant channels (those which do not contain information important for discriminating between executed and imagined motor movements). Thus the benefit of this approach is that we do not make any assumptions about which channels are relevant.
- We noticed that many of the channels appeared to contain redundant information. In an attempt to reduce the irrelevant and redundant information, we explored using PCA. After examining the eigenvalues of the PCA, we observed that the first five eigenvalues contributed $> 99\%$ of the information. The benefit of this approach is that the classifier does not receive redundant (potentially distracting) information.
- Previous work [11,12] performed by neurologists has shown that the neural activity which is related to executed and imagined motor movements is almost exclusively contained within channels C3, C4, and Cz of the EEG recordings. The benefit of this approach is that it is well grounded in previous studies.

After performing some initial experimentation – using methods identical to those explained in detail in sections 4.3 - 4.4, we observed that both using all 64 channels, and using only channels C3, C4, and Cz produced approximately equal prediction accuracies ($\sim 60\%$ correct classification). However, our initial experiments using PCA performed significantly worse ($\sim 52\%$ correct classification). We believe that these results occur

because PCA combines the information across all the 64 channels, when in fact, the differences between the channels are the ones that are important for classifying motor and imagery signals. Following these initial experiments, we decided to consider using only channels C3, C4, and Cz in the remainder of our experiments for the following reason: using only channels C3, C4, and Cz had better prediction accuracy rate than using PCA, and though it had an equivalent predictive accuracy rate versus the approach which used all 64 channels, it was computationally more efficient.

4.3 Feature Selection

4.3.1 EEG Frequency Band Selection

As shown by previous research [13], neural activity can be grouped into the 5 following frequency activity bands:

- Delta (4 Hz) – These occur in deep sleep, childhood and in serious organic brain disease.
- Theta(4-7 Hz) – These occur during childhood, and during drowsiness and idling in adults.
- Alpha(8-12 Hz) – These occur in normal persons when they are awake in a quiet, relaxed state.
- Beta(12-30 Hz) – These are affected by mental activity and occur during the alert/working state state.
- Gamma(30-100Hz) – These occur in certain cognitive or motor functions.

We applied a band pass filter on the EEG signals which admits only the frequencies from 3-30 Hz. This is done to reduce noise at uninteresting frequency ranges – previous research [14] has shown that brain activity is limited to the 3-30 Hz spectrum for awake subjects. Since, only the alpha, theta and beta waves correspond to the brain activity of an awake adult, we selected just these 3 frequency bands for feature selection. We used bandpass filters to extract these three frequency bands from the EEG channels.

4.3.2 Signal power in different frequency bands.

After we filtered out the required frequency bands from the EEG, we computed the average power for each channel. We then normalized the energies across all three frequency bands per channel which thus forms our final feature vector.

Let $P_{c3,alpha}$, $P_{c3,beta}$ and $P_{c3,theta}$ be the average power in *alpha*, *beta*, and *theta* frequency bands in channel C3. Then, the normalized power in *alpha* band for a channel is given by:

$$NP_{c3,alpha} = \frac{P_{c3,alpha}}{P_{c3,alpha} + P_{c3,beta} + P_{c3,theta}}$$

The tasks in the experimental trials are repetitive. As a subject progresses in the trial, the alpha, beta and theta frequency bands decrease in amplitude as the patient prepares for their next movement, thus resulting in reduced cortical activity and lower

energy in the EEG signals in all frequency bands. This observation motivated the normalization of power across frequency bands in a channel. High and low values of average power in different frequency bands do not necessarily correspond to a particular task; it may also mean that the subject has become used to the task and hence requires less cognitive processing to execute those tasks.

Different individuals have different activations and therefore the average power in their EEG signals may not be on the same scale. This too, motivated using the normalized powers across frequency bands as features for classification. The expectation here is, even if the average power may be in frequency bands may vary greatly across subject, the percentage of energy contained in frequency bands might be a common feature across the group.

For comparison, we also present findings in which we did not normalize the average power in the frequency bands.

4.3.3 Feature Space

To summarize, every data segment is characterized by a feature vector given by the tuple:

$$\langle NP_{c3,alpha}, NP_{c3,beta}, NP_{c3,theta}, NP_{c4,alpha}, NP_{c4,beta}, NP_{c4,theta}, NP_{cz,alpha}, NP_{cz,beta}, NP_{cz,theta} \rangle$$

In cases where we did not normalize the average power in the frequency bands, every data segment is characterized by a feature vector given by the tuple

$$\langle P_{c3,alpha}, P_{c3,beta}, P_{c3,theta}, P_{c4,alpha}, P_{c4,beta}, P_{c4,theta}, P_{cz,alpha}, P_{cz,beta}, P_{cz,theta} \rangle$$

4.4 Classifier Selection

We used the libSVM implementation [14] of the support vector machine algorithm with a Gaussian kernel for all experiments. We also performed some initial experimentation with other kernels and classification algorithms; however, all methods appeared to give similar results, thus we present only our usage of the SVM with Gaussian kernel in this paper.

We performed two types of experiments: across the global population and individually on a per patient basis. These approaches differ in that the global experiments train a single classifier which contains multiple patients' data, while individual experiments train a new classifier per patient.

In experiments which are performed across the global population, we partitioned data into a training set (the first 79 patients' data) and a validation set (the last 30 patients' data). We then performed 4-fold cross validation using the training set to select the SVM parameters (C and the kernel's sigma value). After cross validation, we trained

on SVM using the entire training set and the derived SVM parameters, and tested the SVM using the validation set. Our results are shown only from the testing on the validation set.

In experiments which are performed individually on a per patient basis, we partitioned each patient’s data into a training set (the patient’s first 8 trials’ data) and a validation set (the patient’s last 4 trials’ data). We then performed 4-fold cross validation for each patient using the patient’s training set to select SVM parameters for that patient. After cross validation, we trained the SVM using the patient’s entire training set and the derived SVM parameters for that patient, and tested the SVM using that patient’s validation set – yielding a classification accuracy for that patient. Our results are shown as an average of classification accuracies obtained from the individual patients.

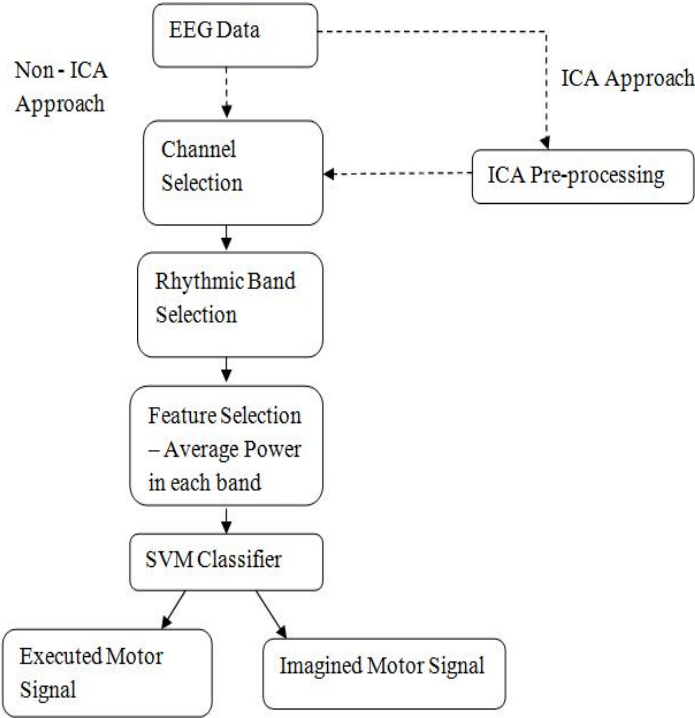


Figure 1: The flowchart of feature extraction and classification process

5 Results

Table 1 depicts the effects that each portion of our approach has on the classification accuracy. Of particular significance, we note that performing classification individually on a per patient basis yields significantly higher classification rates than performing classification globally across all patients. This intuitively makes sense since patients are highly unique, and it therefore seems unlikely that a global classification algorithm

could express these personal discrepancies.

We also observe that the effects of using ICA as a pre-processing step are statistically insignificant. These findings come in contrast to other research, which indicates that ICA is well suited for use on EEG signals. We attribute our findings to the following possible explanations: other more successful applications of ICA have been performed under the eye of expert neurologists, who are capable of distinguishing which ICA components are of importance. We, on the other hand could not make these distinctions, and thus were forced to make the – perhaps incorrect – assumption that components which did not largely contribute the original variance were simply noisy artifacts.

Finally, we notice that normalizing the energies across all three frequency bands per channel also yields a statistically insignificant difference in classification accuracy. We attribute these findings to the observation that in the case of individually classifying on a per patient basis, there is an inherent normalizing effect on the average power in the frequency bands since a single patient’s data is relatively similar. In the case of global classification, we would expect that normalizing the average power in the frequency bands would yield a better classification accuracy than by not normalizing, since patients should appear more regular. However, our findings still show that normalization does not produce significantly better classification accuracies, indicating that not only does the amplitude of the average powers differ across patients, but also the ratio of average power in the respective frequency bands.

	Normalized by Frequency Band				Not Normalized by Frequency Band			
	Per patient		Group		Per Patient		Group	
	ICA	Non-ICA	ICA	Non-ICA	ICA	Non-ICA	ICA	Non-ICA
Accuracy	0.6400	0.6900	0.5604	0.5367	0.6406	0.6286	0.566	0.5654
Standard Deviation	0.05	0.0665	0.0027	0.0236	0.05	0.0734	0.0076	0.0145

Table 1: Comparison between classification various classification schemes

6 Conclusion and Future Work

Our experiments have shown the using EEG signals to perform classification of executed and imagined motor movements can be performed effectively through the use of sophisticated machine learning techniques such as ICA and SVM, despite the close similarities between the two signals. However, it is important to note that results differ significantly across patients, suggesting that BMI is highly specialized and must be uniquely determined for each individual patient.

While features based average signal power have been successfully used to classify neural activities like seizure from regular activity in the brain [15], it is not a good feature for classifying motor imagery. These results support the theory that motor imagery and execution of motor movement are very similar neurological phenomenon, and therefore accurate classification is a challenge.

To improve the accuracy rate of the classifiers further investigation needs to be carried out on other methods of feature selection which might promise a better removal of redundant information from the EEG signal channels. Also, it is possible that the subjects of this experiment didn't execute their tasks in the approved manner (more probability of such occurrences in the imagined motor movement tasks), resulting in an imprecision within the dataset. Thus, experimenting with our methods on a different dataset will help give more insights to the features distinguishing the executed and imagined motor movement EEG signals. Additional sources of error include differences among patients including electrode placement and relative motor cortex location.

7 References

- [1] V. Morash, O. Bai, S. Furlani, P. Lin; M. Hallett, "Classifying EEG signals preceding right hand, left hand, tongue, and right foot movements and motor imageries." *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 119(11):2570-8, 2008
- [2] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *Rehabilitation Engineering, IEEE Transactions on Neural Systems and Rehabilitation*, vol. 8, no. 4, pp. 441-446, 2000.
- [3] Hung C. I., Lee P. L., Wu Y. T., Chen L. F., Yeh T. C., Hsieh J. C. "Recognition of motor imagery electroencephalography using independent component analysis and machine classifiers." *Ann. Biomed. Eng.* 33(8):1053–1070, 2005.
- [4] J. Decety, "Do executed and imagined movements share the same central structures?," *Cognitive Brain Research*, 3, 87-93, 1996.
- [5] S Makeig , S Enghoff, T-P Jung, TJ Sejnowski, "Independent Components of Event related Electroencephalographic Data", *Cognitive Neuroscience Society Abstracts*, 2000
- [6] A. J Bell, T. J. Sejnowski, " An Information- Maximization Approach to Blind Separation and Blind Deconvolution", *Neural Computing*, 1995, 7: pp1129-1159.
- [7] A. Hyvarinen, E. Oja, " Independent Component Analysis: Algorithms and Applications", *Neural Networks*, 2000
- [8] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, "BCI2000: A General-Purpose Brain-Computer Interface (BCI) System.", *IEEE Transactions on Biomedical Engineering* 51(6):1034-1043, 2004.
- [9] AL Goldberger , LAN Amaral, L Glass, JM Hausdorff, PCh Ivanov, RG Mark, JE Mietus, GB Moody , C-K Peng , HE Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals." *Circulation* 101(23):e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>]; 2000 (June 13).
- [10] A Delorme, S Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis", *Journal of Neuroscience Methods*, 2004.
- [11] L Deecke , H Weinberg , P Brickett , "Magnetic fields of the human brain accompanying voluntary movements: Bereitschaftsmagnetfeld." *Exp Brain Res* 48: 144–148, 1982.
- [12] C. Neuper, G. Pfurtscheller, "Evidence for distinct beta resonance frequencies in human EEG related to specific sensorimotor cortical areas," *Clinical Neurophysiol.* 112 (11), pp. 2084–2097, 2001.
- [13] A. Tzelepi, T. Bezerianos, I. Bodis-Wollner, "Functional properties of sub-bands of oscillatory brain waves to pattern visual stimulation in man.," *Clinical Neurophysiology Volume 111, Issue 2, 1 February 2000, Pages 259-269*
- [14] C. C. Chang , C. J. Lin, "LIBSVM: a library for support vector machines", 2001. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [15] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echaz, B. Litt, "Epileptic Seizure Prediction Using Hybrid Feature Selection Over Multiple Intracranial EEG Electrode Contacts: A Report of Four Patients", *IEEE Transactions on Biomedical Engineering*, Vol. 50, No. 5, May 2003