Motor Intention Recognition in EEG: In Pursuit of a Relevant Feature Set

Pablo A. Iturralde[1], Martín Patrone[1], Federico Lecumberry[2], and Alicia Fernández[2]

[1] Department of Physics, School of Engineering, UdelaR, Montevideo, Uruguay
[2] Department of Electrical Engineering, School of Engineering, UdelaR, Montevideo, Uruguay
{iturral,mpatrone,fefo,aliciaf}@fing.edu.uy

Abstract. Brain-computer interfaces (BCIs) based on electroencephalograms (EEG) are a noninvasive and cheap alternative to get a communication channel between brain and computers. Some of the main issues with EEG signals are its high dimensionality, high inter-user variance, and non-stationarity. In this work we present different approaches to deal with the high dimensionality of the data, finding relevant descriptors in EEG signals for motor intention recognition: first, a classical dimensionality reduction method using Diffusion Distance, second a technique based on spectral analysis of EEG channels associated with the frontal and prefrontal cortex, and third a projection over average signals. Performance analysis for different sets of features is done, showing that some of them are more robust to user variability.

1 Introduction

In recent years, Brain-Computer Interfaces (BCIs) have become an active topic of research. Such interfaces could provide an alternative communication channel between humans and computers, replacing the normal output channel of nerves and muscles. As its main application, BCIs could be used by paralyzed patients or others suffering some type of motor impairment but who are cognitively intact as a means to interact with the environment.

BCIs register brain states (relating to thoughts and intentions) as signals that are interpreted and translated into actions. The signal acquisition process is critical to the performance of the whole system, and several technologies have been proposed to carry out such a task. Signal registering through electroencephalograms (EEG) are one of the most promising systems because of its noninvasive nature that allows for simpler and cheaper devices with almost no associated risks, as opposed to invasive technologies such as electrocorticographic (ECoG) signals which require medical procedures for its implantation. However, EEG signals provide only a diffuse access to brain signals, since currents in the brain cortex are volume conducted through the skull before being sensed at the scalp. This means that more sophisticated processing and recognition systems are needed in order to obtain information about brain states from such signals [12,10].

One of the main issues with EEG signals is its high dimensionality. Usually, signals from over ten and up to hundreds of channels are acquired at sampling rates of at least 100 Hz to ensure that no aliasing occurs. This implies that even for short trials that last for about a second, the raw feature space has a dimension between 1000 and 60000, making it very difficult to work with. One of the most used [3,2] is to perform component analysis (PCA, ICA or equivalent) in order to reduce redundancy in both time (over-sampled signals) and space (EEG electrodes that are nearby have similar EEG signals). This type of analysis (particularly ICA) has also been used to deduce the actual location of brain activity from EEG signals, see [3].

Another common approach [13] after the component analysis is to divide the remaining signals in several frequency bands along different time windows for the duration of the trial (time-frequency analysis [11]), using these spectral components to perform the classification. While the amplitude of the spectral components is widely used particularly in the mu and beta bands [2], the phase of the components has received special attention because of its relation to Event Related (De)Synchronization (ERS/ERD) [8,11].

Dimensionality reduction is a major topic of research, and several algorithms have been proposed to perform such task. Data laying in high dimensional spaces generally presents complex geometries, and probably not enough samples for accurate statistics. This kind of problem requires new strategies to deal with it instead of the classical PCA or Multi-Dimensional Scaling (MDS). One reason is the non linearity of the manifold where the original high dimensional data points lies. Another is the computational cost of traditional methods applied to data of high dimension, as it grows exponentially with dimension. Additionally, the sparse sampling leads to poor convergence of the algorithms, a phenomenon referred to as "the curse of dimensionality". Although input data may present a high dimensionality, it is common that the "real" or intrinsic dimensionality of the source that generates this data is much lower due to significant correlations between many of the coordinates. Finding meaningful structures in the data and obtaining those "principal coordinates" is one of the goals of the machine learning algorithms. In past years kernel based methods have concentrated the attention with good results and a solid background theory; a brief list of these methods include Locally Linear Embedding [9], Laplacian Eigenmaps [1], Hessian Eigenmaps [4] and Diffusion Maps [5].

In this paper we deal with a motor imagery task, consisting of several trials in where a user decided to release, and actually released, a button or not. The situation has the same high dimensionality problem that was described before, consisting of instances containing 31 EEG signals originally sampled at 1kHz for a time frame of one second. To reduce the amount of data preserving the spectral components of interest, downsampling to 100Hz was performed. Data also presents high inter-user variance, and non-stationarity.

We present different approaches to deal with the high dimensionality of the data, finding relevant descriptors in EEG signals for motor intention recognition: first, a classical dimensionality reduction method using Diffusion Distance,

second a technique based on spectral analysis of EEG channels associated with
the frontal and prefrontal cortex [11], and third a projection over average signals
as proposed in [3].

This paper is structured as follows. On section Methods we describe the
dataset that was used, and we detail the three different approaches used. On
section Results we present the performance analysis in all three cases for all the
users in the dataset, as well as some other results obtained from different vari-
ants of these strategies. Discussion regarding the results and future lines of work
are mentioned on the last section.

2   Methods

2.1   Data Description

The EEG data files of the experiment were made available by A. Delorme in
(`http://sccn.ucsd.edu/~arno/fam2data/data/`). Fourteen subjects of several
ages and both sexes were tested in two recording phases on different days. Each
day consisted of at least 10 series with 100 images per series. The subjects were
asked to identify target images and non-target images, equally likely presented.
In each instance, the image was shown for 20 msec. in a computer (avoiding the
use of exploratory eye movement), and the subject had to gave their response
following a go/nogo paradigm. For each target they had to lift their finger from
a button as quickly as possible, and for each distractor had to keep pressing the
button during at least 1 sec (nogo response).

Scalp EEGs were recorded from 31 channels (placed according to the inter-
national 10/20 system [6], see figure 1) with a 1 kHz sampling rate during 1s
for each instance. Every instance consists of a single image experiment; it starts
with the image being displayed for 20 msec., and lasts for a second, which was
the maximum time the users had to respond. The database includes at least
1000 images per subject and fourteen different subjects, which results in 14000
instances with a feature set of 31000 dimensions. However, since EEGs depends
on personal physiological factors [6], there is an important variability in the
characteristics of the signals between subjects (signal mean and energy). This is
the reason to use the data as fourteen independent sets, one for each subject.

Pre-processing of the data was performed to correct for a DC drift between
instances. Therefore, signals from every channel in each instance were forced
to start at zero by substracting a DC level. Signals were also downsampled to
100 Hz, since spectral analysis showed that no energy was present over 50 Hz,
making the elevated original sampling rate unnecesary.

2.2   High-Dimensionality Analysis Tools

A first step for dimension reduction is to estimate the intrinsic dimension of the
problem.

We found the best results (regarding consistency in the estimation) using the
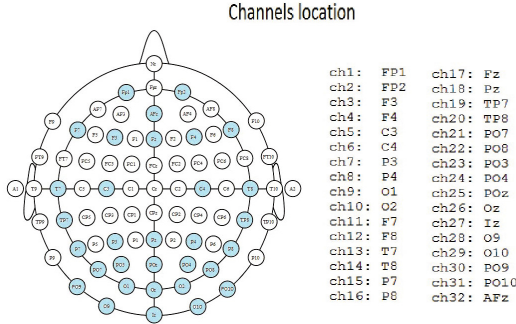Maximum Likelihood Estimator for dimensionality, which gave values ranging

Channels location

ch1:	FP1	ch17:	Fz
ch2:	FP2	ch18:	Pz
ch3:	F3	ch19:	TP7
ch4:	F4	ch20:	TP8
ch5:	C3	ch21:	PO7
ch6:	C4	ch22:	PO8
ch7:	P3	ch23:	PO3
ch8:	P4	ch24:	PO4
ch9:	O1	ch25:	POz
ch10:	O2	ch26:	Oz
ch11:	F7	ch27:	Iz
ch12:	F8	ch28:	O9
ch13:	T7	ch29:	O10
ch14:	T8	ch30:	PO9
ch15:	P7	ch31:	PO10
ch16:	P8	ch32:	AFz

Fig. 1. Channel locations used in the EEG recordings, according to the 10/20 system. Image adapted from `http://www.mariusthart.net/`

from 29 to 35 for the different users. Independently from the estimator used, a mapping tool was selected in order to reduce the feature space. Locally Linear Embedding, Laplacian Eigenmaps, Hessian Eigenmaps and Diffusion Maps (DM) were considered. In this case the best results were obtained with the use of Diffusion Maps. DM is a general framework for data analysis based on a diffusion process over an undirected weighted graph, defining a new metric on the data called Diffusion Distance [7]. This distance is equivalent to the Euclidean distance in the space with coordinates given by the mapping function. In order to compute the weights on the graph we used a Gaussian kernel with an adequate (manually) selected variance, fixed for all subjects.

Once the reduced feature space was obtained, a classifier was trained with these new features.

2.3   Channel and Time Window Selection from Active Zones

Plotting the data corresponding to the EEG signals over a single trial, different zones of activity (both in the spatial and temporal senses) become visible. Approximately 100msec after the image is shown, the occipital region becomes active. Since the occipital lobe is where visual information is processed, this seems to indicate that an analysis of the image is being performed. After about 550msec pre-frontal cortex becomes active, which is consistent with the expected motor control actions that need to be carried out. While it could be expected that the activity be more noticeable when a motor action is indeed executed (target class), both classes seem to present some of this activity (see figure 2, which show average signals for target and non-target instances). Furthermore, it is in these channels and time-window that the greatest differential activity is observed.

From both the functional analysis of the cerebral lobes and the activity zones shown in the signals, follows that considering a feature space consisting only of the signals corresponding to frontal and pre-frontal channels (F-channels, see figure 1) might allow to reduce dimensionality while still allowing to differentiate
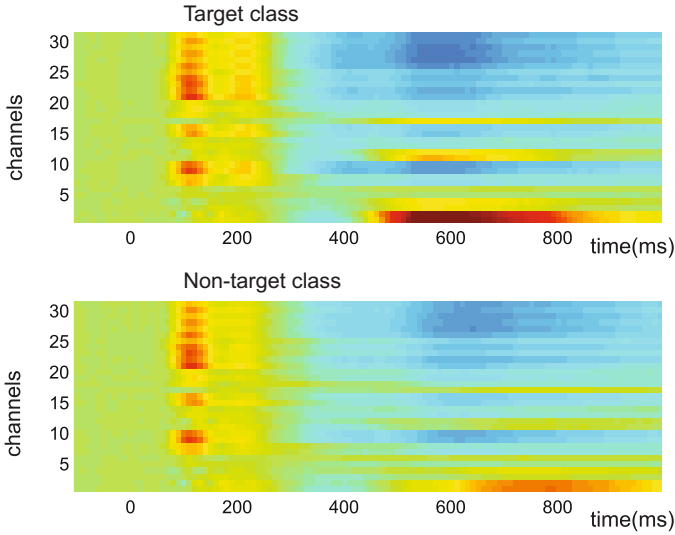
Fig. 2. Average signals for target and non-target classes

between the two types of instances. Spectral analysis was then performed over these channels. Frequencies finally considered for the training of classifiers where in the 0-10Hz band, resulting in a feature vector of dimension 275.

2.4 Projection over Average Signals

An average signal –over instances– was extracted for each of the 31 available channels, for each class. Thus, 62 reference signals were obtained, 31 corresponding to the average in all the EEG channels in the target class (T) and 31 for the channels in the non-target class (NT). Let us call s_i^j the time signal corresponding to the i-th channel for the j-th instance, s_i^T the time signal corresponding to the i-th channel in the average over the target class in the training set, and s_i^{NT} the time signal corresponding to the i-th channel in the average over the non-target class in the training set (see Equations 1, 2). For each instance, a projection over the reference signals was performed (channel by channel, by means of a scalar product over time), resulting in a feature vector $c = (c_i^T, c_i^{NT})$ of 62 scalar features for each instance (see Equations 4, 3). The first 31 of these features (c_i^T) refer to the projection of the instance's signals into the averaged signals for the target class, and the last 31 (c_i^{NT}) for the non-target class. From this new feature space different classifiers were trained. Results were validated over the test set.

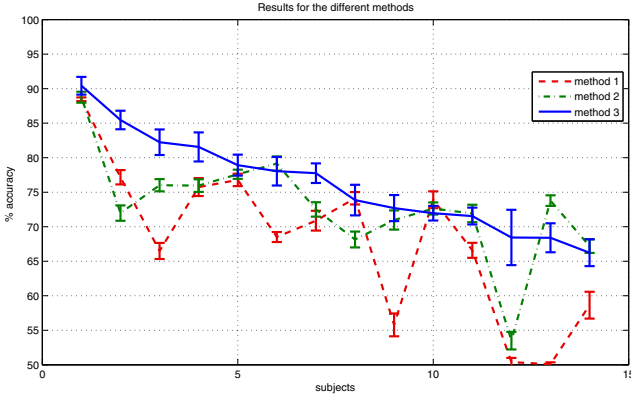$$s_i^{NT} = \frac{1}{N} \sum_{j \in NT} s_i^j, \, i \in [1, 31] \tag{1}$$

Fig. 3. Results obtained for the different users. Method 1: Diffusion Maps. Method 2: Channel and time window selection from active zones. Method 3: Projection over average signals. The classification was made with a two-layer feedforward perceptron, with 3 neurons in each layer, a learning rate of 0.3 and backpropagation as the training algorithm.

$$s_i^T = \frac{1}{N} \sum_{j \in T} s_i^j, \, i \in [1, 31] \tag{2}$$

$$c_i^{NT} = < s_i^j, s_i^{NT} >, \, i \in [1, 31] \tag{3}$$

$$c_i^T = < s_i^j, s_i^T >, \, i \in [1, 31] \tag{4}$$

Experiments were also conducted considering only F-channels, as is suggested by the analysis of the previous section. However results were worse in that case than when considering all channels.

3   Results

The results obtained with the methods described in the previous section are shown in figure 3. For all cases different classifiers were trained and tested (multilayer perceptrons, C4.5 trees) but since results did not differ significantly among them, only best results concerning multilayer perceptrons are presented.

There are two major observations that arise from figure 3. First of all, for most subject the best results are obtained by making the projection over average signals (Method 3), even over Method 2 (spectral analysis in frontal and prefrontal channels) which is the most common approach found in literature. This indicates that there is useful information in the shape of the signals (in time) that could complement the frequency analysis.

Second, the inter-user variance is significantly lower for Method 3 which indicates that this method is more consistent among the different subjects. On

Table 1. Average results and variance for the tested methods over all users

	Method 1	Method 2	Method 3
mean	68.1%	72.9%	76.3%
variance	11.0%	7.7%	7.0%

the other hand, DM (Method 1) presents an extreme inter-user variance (with accuracy going over 85% for some users and as low as 50% for others).

Method 3 presents the best results regarding both classifier performance and inter-user variance (see Table 1). Although this method uses a feature set dependent on information extracted from the actual signals (averages over instances) it can be generalized, which allows an automatic process of feature extraction and classification for new users.

It is worth noting that results drop notoriously if instances are not randomly mixed before the train and test groups are separated. This seems to be due to the temporal non-stationarity that exists between signal from the same subject, where significant differences can be found between signals from different experiments. Even more, these differences seem to increase with time.

The results are consistent with the foreseen need to include expert knowledge in the feature extraction. The dimension of the original feature space has proved to be too big for most automatic algorithms of dimensionality reduction.

4   Conclusion

We proposed and tested three different approaches to perform feature selection/extraction in EEG signals. All the methods make a classifier independent selection. Performance evaluation of motor intention recognition, using the selected features with a two layer perceptron show that the results are user dependent for all the methods, but projection over average signals (Method 3) shows the least variability between users. The classifier based on DM has the same high performance for the best users but a very low one for others; further analysis in the parameter selection process is needed in order to generalize this method to new users. As an advantage, DM shows lower dependability with the training/test data set split.

Discarding channels that are not related with the frontal cortex works well for reducing dimensionality, and thus helps to increase classifier performance, as was shown with results for Method 2. However for Method 3 discarding channels seemed to show a decrease in performance, suggesting that there is indeed relevant information associated with these channels. It is worth noticing that in the latter case dimensionality is no longer an issue, since the projection over average signals yields a feature space of dimension 62, as opposed to the thousands originally present.

In the future we will try to obtain results while making embedded feature selection with different classifiers (SVM, C4.5, etc.) and try combining the three

tested methods in order to improve the performance for each user. Some promising early results along this line show that an improvement of 4% over the best single classifier is possible at least for some users.

References

1. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering (2002)
2. Bigdely-Shamlo, N., Vankov, A., Ramirez, R.R., Makeig, S.: Brain activity-based image classification from rapid serial visual presentation. IEEE Transactions on Neural Systems and Rehabilitation Engineering 16(5), 432–441 (2008)
3. Delorme, A., Makeig, S., Fabre-Thorpe, M., Sejnowski, T.: From single-trial eeg to brain area dynamics. Neurocomputing 44-46, 1057–1064 (2002); Computational Neuroscience Trends in Research 2002
4. Donoho, D.L., Grimes, C.: Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. Proceedings of the National Academy of Sciences 100(10), 5591–5596 (2003)
5. Lafon, S., Keller, Y., Coifman, R.R.: Data fusion and multicue data matching by diffusion maps. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(11), 1784–1797 (2006)
6. Malmivuo, J., Plonsey, R.: Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields. Oxford University Press (1995)
7. Nadler, B., Lafon, S., Coifman, R.R., Kevrekidis, I.G.: Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. Applied and Computational Harmonic Analysis 21(1), 113–127 (2006); Special Issue: Diffusion Maps and Wavelets.
8. Pfurtscheller, G., Lopes da Silva, F.H.: Event-related eeg/meg synchronization and desynchronization: basic principles. Clinical Neurophysiology 110(11), 1842–1857 (1999)
9. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000)
10. Vaughan, T.M.: Guest editorial brain-computer interface technology: a review of the second international meeting. IEEE Transactions on Neural Systems and Rehabilitation Engineering 11(2), 94–109 (2003)
11. Wang, T., Deng, J., He, B.: Classifying eeg-based motor imagery tasks by means of time–frequency synthesized spatial patterns. Clinical Neurophysiology 115(12), 2744–2753 (2004)
12. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain–computer interfaces for communication and control. Clinical Neurophysiology 113(6), 767–791 (2002)
13. Wolpaw, J.R., Birbaumer, N., Heetderks, W.J., McFarland, D.J., Peckham, P.H., Schalk, G., Donchin, E., Quatrano, L.A., Robinson, C.J., Vaughan, T.M.: Brain-computer interface technology: a review of the first international meeting. IEEE Transactions on Rehabilitation Engineering 8(2), 164–173 (2000)