

A Method for Classifying Mental Tasks in the Space of EEG Transforms

Renato Amorim
Birkbeck, University of London
renato@dcs.bbk.ac.uk

Boris Mirkin
Birkbeck, University of London
mirkin@dcs.bbk.ac.uk

John Q. Gan
University of Essex
jqgan@essex.ac.uk

Abstract

In this article we describe a new method for supervised classification of EEG signals. This method applies to the power spectrum density data and assigns class-dependent information weights to individual pixels, so that the decision is defined by the summary weights of the most informative pixel features. We experimentally analyze several versions of the approach. The informative features appear to be rather similar among different individuals, thus supporting the view that there are subject independent general brain patterns for the same mental task.

1. Introduction

Electroencephalography (EEG) signals contain useful information about the state or intention of the mind [1, 2, 9, 12] and are considered to be one of the best non-invasive approaches to acquiring information for classifying mental tasks [5].

EEG signals may provide an individual with an alternative channel for communication with the external environment [5, 8, 12]. It could be the only possibility to communicate with other people, if the individual is completely motor paralyzed but has intact sensory and cognitive brain functions (locked-in-syndrome), with the communication being conducted via a brain computer interface (BCI). Other interesting applications of EEG signals include the diagnosis of neurological disorders and other abnormalities of the human body [9, 12] and even monitoring the depth of anesthesia [7].

Although EEG signals may provide a number of benefits, their processing is far from being trivial. Indeed each electrode records the activity of thousands of neurons simultaneously [5], which makes EEG recording very noisy, and thus EEG patterns are difficult to discern.

Although most of the noise is supposed to come from either within the brain or over the scalp [9], the truth is that there can be many other sources of noise,

sometimes rather significant, such as eye movement, muscle activity, cardiac activity, respiration and skin potential. Also, even if the pure biological EEG source were to be noise free, amplification and digitalization would add noise (systematic bias) [1]. As a matter of fact noise components of a signal can have different origins, biological or not.

More irregularities in the EEG patterns may also be generated by the use of devices such as mobile phones, as exposure to pulse modulated electromagnetic fields generated by them affects the cerebral blood flow in certain areas of the brain [3, 13].

Another problem faced by EEG signal processors is that classifying these signals is an intrinsically high dimensional task [10]; a recording of one hour using 128 electrodes at 500 samples per second would generate around 0.45 GB [9].

One well-known technique used to smooth data and reduce variability is signal averaging [2]. This also allows estimation of the amplitude of signals that may be buried in noise, which involves the following, not necessarily realistic, assumptions [1]:

- The signal and the noise are uncorrelated;
- The timing of the signal is known;
- A consistent signal component exists to be extracted using repeated measurements;
- The noise is truly random with zero mean.

The averaging technique has proven sufficiently robust to survive minor violations of the above assumptions and it is currently used by researchers, see, for example, [2, 12].

Our proposed method takes advantage of the above benefit by applying averaging to the power spectrum density data that belong to the same class. In this way we can find information weights of the most informative pixels and use them for classification.

2. Description of the Data and the Power Spectrum Density

The raw EEG data were recorded from three healthy subjects, with five bipolar electrodes (channels) and

sampling frequency 250Hz. In each trial of eight seconds there are 2001 samples. The electrodes were placed according to positions standardized in the extended 10-20 system using fc3 to pc3, fc1 to pc1, cz to pz, fc2 to pc2 and fc4 to pc4. In spite of suggesting that dense electrode arrays may enhance the signals for classification purposes [8, 15], only five channels were utilized here because our previous work [11] has shown that this configuration is able to achieve good classification for online BCI.

In each trial, the subject was instructed to imagine a body movement task which could be either

- (T1) Moving left hand, or
- (T2) Moving right hand, or
- (T3) Moving feet.

This set of tasks will be denoted as $\Omega = \{T1, T2, T3\}$.

The EEG signal obtained in a trial was recorded as a three way function $a=f(t, c, r)$ in which a represents the amplitude, t corresponds to a sample and ranges from 1 to 2001, c to one of five channels, $c=1, \dots, 5$, and r to a trial. The three subjects from which the data have been collected are denoted as A (240 trials), B (120 trials) and C (350 trials) respectively.

The raw data have been transformed into power spectrum density (psd) to generate a dataset which can be viewed as a set of images. Each image is a raster of 71×80 pixels to represent a trial on the scale of its psd as a function of time and frequency-channel, in which 71 is the number of samples within second 1 and second 8 in a trial, 80 is the number of features in each sample consisting of psds over 8~45Hz (psds in 16 frequency bands from each channel). Each of these images will be referred to as a trial in the remainder.

Consider a set S of N trials S_i ($i=1, 2, \dots, N$). This set is partitioned into three subsets $S(\omega)$ corresponding to the mental tasks $\omega=T1, T2, T3$, so that $S = S(T1) \cup S(T2) \cup S(T3)$. This allows us to average trials over the tasks as follows.

$$S_\omega = \frac{1}{N_\omega} \sum_{i \in S(\omega)} S_i \quad (1.1)$$

Where N_ω is the number of trials in $S(\omega)$.

3. The Method

Our method involves two phases: (i) feature extraction and (ii) classification, which is similar to most BCI techniques [10].

As defined by [4, 14] a pattern can be seen as the opposite of a chaos; it is an entity, vaguely defined, that could be given a name. What is important is that

patterns emerging at solving the same mental tasks much depend on personal circumstances (compare Figures 2 (a) and 2 (b)) and a number of other aspects, including the aging process as it changes the response of brain to stimuli [9]. That is why feature extraction is especially important for BCI systems based on EEG signals.

3.1 Feature extraction

Since EEG data are likely to have a certain degree of noise, assuming that this noise is truly random, it can be minimized by averaging the trials belonging to the same tasks by using formula (1.1), which leads to three averaged trials S_ω , one for each of the tasks $\omega \in \Omega$.

In the data used for experiments, each trial $s \in S$ has 5680 pixels (71×80), so that it becomes important to identify a small group of pixels to be used as features in the follow-up classification.

In order to create such a group one needs a measure of importance of a pixel (m, n) ($m=1, \dots, 71; n=1, \dots, 80$) for classification purposes. We measure the distance of a pixel in trial $s \in S$ to task $\omega \in \Omega$ by comparing its brightness with the brightness of the corresponding pixel in the average trial S_ω :

$$d(s(m, n), \omega) = |s(m, n) - S_\omega(m, n)|^2 \quad (1.2)$$

We refer to the pixel as being good in trial s if the difference of brightness (1.2) is smaller to the mean of the subset $S(\omega)$ that has the correct label. The goodness value for pixel (m, n) is defined then as the proportion of trials in which it is good. After computing the goodness values for all pixels, one can use them as the importance weights.

Moreover, our experiments have shown that classification results can be improved if pixels with low goodness values are discarded. Hence we introduce a threshold θ such that all pixels whose goodness value is less than θ are removed from the process of classification. Features remaining after removal of all those pixels whose goodness is less than $\theta = 0.45$ can be seen in Figures 1 (a), 1 (b) and 1 (c) for subjects A , B and C respectively. These figures show the pixels found as being the most important ones for classification in the three subjects. It can be seen that the pixels before second 4 in each trial tend to be unimportant to the classification (check the y-axis, about the 35th time point), psds over 8~17Hz (the first five frequency bands in each channel) are most useful, and all the 5 channels provide important pixels and thus make contributions to the classification (check the

x-axis).

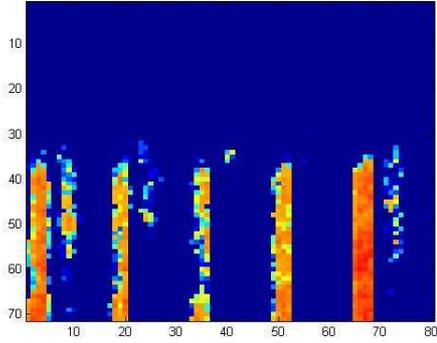


Figure 1 (a)

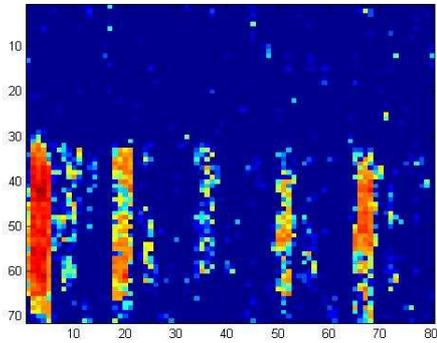


Figure 1 (b)

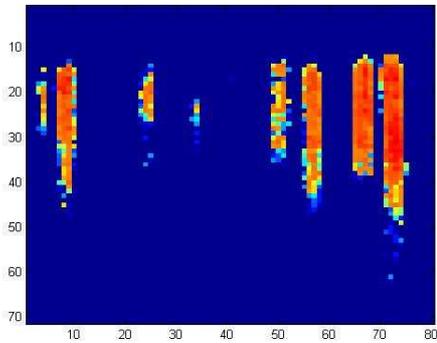


Figure 1 (c)

The feature extraction algorithm is as follows

- 0.0. Initialize all goodness values $g(m, n)=0$.
- 1.0 Calculate the mean of all trials per task S_ω using formula (1.1).
- 2.0 For each trial s
 - 2.1 For each pixel (m, n)
 - 2.1.1 Calculate its distance to the same pixel in S_ω using formula (1.2).
 - 2.1.2 If the distance is of pixel $(m,n) \in s$ is closer to its peer pixel in

the correct average trial $S(\omega)$, increase its goodness value $g(m, n)$ by 1.

- 3.0 Update the goodness values by dividing them by the total number of trials and remove all pixels whose goodness values are smaller than θ (i.e., set those initially as $g(m,n)$ as 0).

3.2 Classification

Given a trial s , its distances to tasks $\omega \in \Omega$ are calculated using the following formula:

$$d(s, S_\omega) = \sum_{n=1}^{N_s} \sum_{m=1}^{N_\omega} |s(m, n) - S_\omega(m, n)|^2 g(m, n) \quad (1.3)$$

The trial is assigned to its closest task $\omega \in \Omega$.

4. Experimental Results

In an attempt to get the best possible value for the threshold θ with subjects A , B and C , a number of experiments have been run. The accuracies given in figures 2 (a), 2 (b) and 2 (c) are the averages of three runs using the ten-fold cross validation method. The relation between the threshold and the accuracy with subject A is depicted in Figure 2 (a). The best result, 67.22%, is at $\theta=0.45$.

The accuracy for subject B as function of the discarding threshold which had a maximum of 84.17% at $\theta=0.45$ is presented in Figure 2 (b).

Subject C had a maximum accuracy of 82.34% at $\theta=0.48$ and 82.07% at $\theta=0.45$, as shown in Figure 2 (c).

One can notice that subject B 's data allow to achieve a higher accuracy rates including the maximum of 84.17% at the same optimal value of threshold $\theta=0.45$ as subject A . Also the difference of accuracies for subject C using its optimal threshold of 0.48 and 0.45 is not that great either.

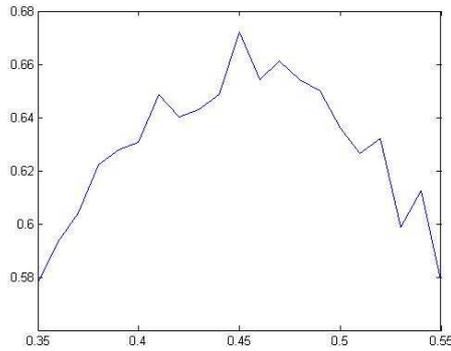


Figure 2 (a) Accuracy rates at different thresholds for subject A.

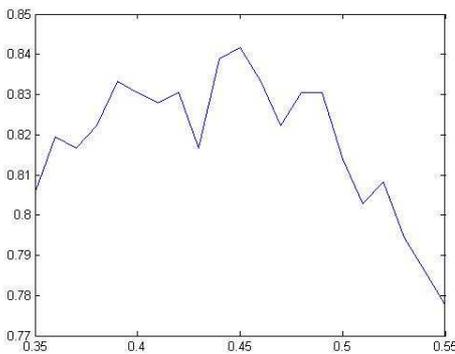


Figure 2 (b) Accuracy rates at different thresholds for subject B.

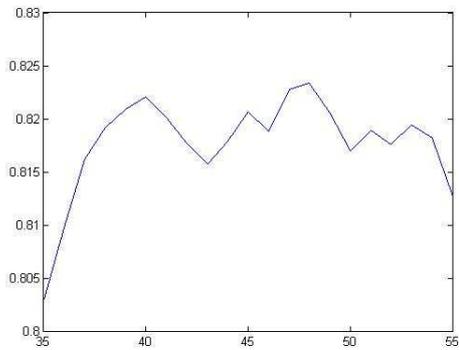


Figure 2 (c) Accuracy rates at different thresholds for subject C.

5. Comparative Results

We have experimented with several versions of the method on the same data.

5.1. Contrast exponent

In the first modification we raise the goodness values to the power of the parameter e . The objective was to make the difference between pixels more contrast, so

that the best pixels would have even greater influence in the classification process using formula (1.3).

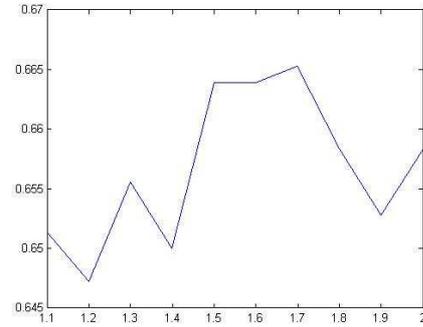


Figure 3 (a) Accuracy rates for subject A at different values for e .

The accuracies shown in figure 3 (a) for subject A were obtained at the best threshold θ values. Similar results for subject B and C are presented in Figures 3 (b) and 3 (c) respectively. These show that the “contrast” exponent gives no subject independent improvement to the classification results.

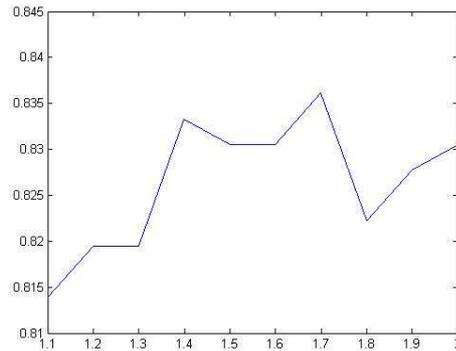


Figure 3 (b). Accuracy rates for subject B at different values for e .

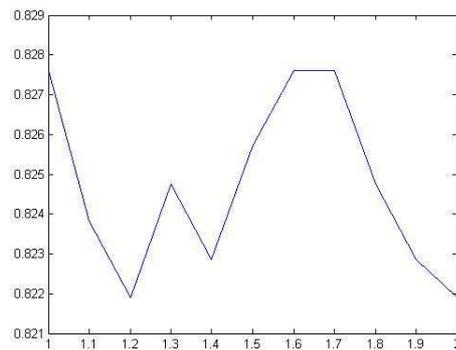


Figure 3 (c). Accuracy rates for subject C at different values for e .

5.2. Different weighting

In our second modification we used different weighting in formula (1.3). We have made experimentations with

the same method but changing the weight to $1/(1+g)$. The accuracy results using the respective best values for the threshold θ found in the previous section were much worse:

- Subject A: 0.5542
- Subject B: 0.7194
- Subject C: 0.7200

5.3. Different distance

Another modification involved the weighted Pearson correlation instead of the Euclidian squared distance:

$$p(s, S_\omega) = \frac{\sum_{m=1}^{N_s} \sum_{n=1}^{N_m} (s(m,n) - \bar{s}) \cdot (S_\omega(m,n) - \bar{S}_\omega) g(m,n)}{\sqrt{\sum_{m=1}^{N_s} \sum_{n=1}^{N_m} (s(m,n) - \bar{s})^2 g(m,n) \sum_{m=1}^{N_s} \sum_{n=1}^{N_m} (S_\omega(m,n) - \bar{S}_\omega)^2 g(m,n)}} \quad (1.4)$$

While in terms of the distance a trial should be labeled to its closest averaged trial, in this modification it should be labeled to the averaged trial with the highest weighted correlation index.

This modification has also provided a lower accuracy than the original method itself:

- Subject A: 0.4639
- Subject B: 0.5611
- Subject C: 0.5495

The previous test was also performed with an extra step in the data preprocessing stage: subtract the given trials by the mean of all trials, which lead to the following accuracies.

- Subject A: 0.6333
- Subject B: 0.8056
- Subject C: 0.8171

One more experiment was conducted using the above index (formula (1.4)) with the extra preprocessing step explained in the previous paragraph and the weight as $1/(1+g)$. The accuracies are as follows:

- Subject A: 0.5625
- Subject B: 0.7139
- Subject C: 0.7524

A summary of the experiments is presented in Table 1.

Table 1. Accuracy rates of methods using different modifications. Methods	A(%)	B(%)	C(%)
Goodness Matching Algorithm	67.2	84.1	82.3
1 st version: contrast exponent	66.5	83.6	82.7
2 nd version: different weighting	55.4	71.9	72.0

3 rd version: Pearson correlation	46.3	56.1	54.9
4 th version: Pearson correlation combined with shift to grand mean	63.3	80.5	81.7
5 th version: 4 th version using the different weighting	56.2	71.3	75.2

6. Results of Using Data within Shorter Windows

In real world applications a BCI system should respond as fast as possible rather than at the end of a trial. This requirement fits well with our proposed method, because it does not need to have a full trial as input. One can use, for instance, only the trial data within a temporal window. The results using different window sizes for subjects A, B and C are listed in Tables 2, 3 and 4 respectively.

Table 2. Accuracy rates for subject A using different window sizes

Subject	Threshold	Window Size (Time Points)	Accuracy
A	0.47	5	63.33%
A	0.48	10	62.92%
A	0.47	15	63.06%
A	0.49	20	64.72%
A	0.47	25	64.31%
A	0.47	30	65.28%

Table 3. Accuracy rates for subject B using different window sizes

Subject	Threshold	Window Size (Time Points)	Accuracy
B	0.47	5	72.5%
B	0.47	10	75.00%
B	0.49	15	77.78%
B	0.49	20	80.83%
B	0.48	25	83.61%
B	0.47	30	83.61%

Table 4. Accuracy rates for subject C using different windows sizes

Subject	Threshold	Window Size (Time Points)	Accuracy
C	0.49	5	79.90%
C	0.47	10	81.43%
C	0.48	15	82.86%
C	0.49	20	82.48%
C	0.49	25	82.48%
C	0.47	30	82.86%

Even though the optimal threshold value now depends on the size of the window, we can still see 0.47 as being an optimal value.

When the optimal threshold is 0.48 or 0.49 the increase in accuracy in relation to the use of 0.47 is very small and in most cases cannot be proved.

7. Conclusion and Discussion

We proposed a classification method involving a measure of the goodness of pixels as both the pixel

weighting coefficient and the pixel rejection base.

The results of EEG classification heavily depend on subjects, as was pointed out by other researchers [5].

As the reader can see in Figures 1 (A), 1 (B), and 1 (C), although subjects *A*, *B*, and *C* provide for very different accuracy rates, their good pixels are somewhat similar. This suggests that there may be subject independent general brain patterns for the same tasks. This could also mean that finding 0.45 (when using the whole data of a trial) and 0.47 (windowed version) as the best parameter for the threshold θ for all subjects, *A* *B* and *C*, is not entirely by chance.

One can notice, too, that a good classification is not a matter of having less sparse goodness value tables: subject *A* even having a much denser clusters than subject *B* has led to poorer result.

This points to the conclusion that BCI devices should be supplied with a set of classification algorithms so that a learning device would not only fit the parameters but also choose the algorithm that better applies to a subject.

Since our goodness measure is task independent, it can also be applied to not only learning but also clustering trials using the k-means algorithm and its intelligent version [6].

References

- [1] Drongelen, W. V., Signal Processing for Neuroscientists: An Introduction to the Analysis of Physiological Signals, Academic Press/Elsevier, Burlington MA. USA, 2007
- [2] Geng, T., Gan, J. Q., Dyson, M., Tui, C. S. L. and Sepulveda, F., A novel design of 4-class BCI using two binary classifiers and parallel mental tasks, Journal of Computational Intelligence and Neuroscience, volume 2008, doi:10.1155/2008/437306
- [3] Huber, R., Treyer, V., Borbe, A. A., Schuderer, J., Gottselig, J. M., Landol, H. P., Werth, E., Berthold, T., Kuster, N., Buck, A., Achermann, P., Electromagnetic fields, such as those from mobile phones alter regional cerebral blood flow and sleep and awaking EEG, Journal of Sleep Research, vol. 11, no. 4, pp. 289-295
- [4] Jain, A. K., Duin, R. P. W., Mao, J., Statistical pattern recognition: A review, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, vol. 22, no. 1, pp. 4 - 37
- [5] Lee, F., Scherer, R., Leeb, R., Neuper, C., Bischof, H., Pfurtscheller, G. A Comparative analysis of multi-class EEG classification for brain computer interface, Proceedings of the 10th Computer Vision Winter Workshop, 2005, pp.195-204
- [6] Mirkin, B., Clustering for Data Mining: A Data Discovery Approach, Chapman and Hall/CRC, Boca Raton Fl. USA, 2005
- [7] Ortolani, O., Conti, A., Di Filippo, A., Adembri, C., Moraldi, E., Evangelisti, A., Maggini, M., Roberts, S. J., EEG signal processing in anaesthesia: Use of neural network technique for monitoring depth of anaesthesia, British Journal of Anaesthesia, 2002, vol. 88, no. 5, pp. 644-648
- [8] Peterson, D., Knight, J., Kirby, M. Anderson, C., Thaut, M., Feature selection and blind source separation in EEG-based brain-computer interface, EURASIP Journal on Applied Signal Processing, 2005, vol. 19, pp. 3128-3149
- [9] Sanei, S., Chambers, J. A., EEG Signal Processing. UK: WileyBlackwell, 2007
- [10] Tomioka, R., Aihara, K., Muller, K. R., Logistic regression for single trial EEG classification, Advances in Neural Inf. Proc. Systems, 2007, vol. 19, pp.1377-1384
- [11] Tsui, C., Gan, J. Q., Roberts, S., A self-paced brain-computer interface for controlling a robot simulator: an online event labelling paradigm and an extended Kalman filter based algorithm for online training, Medical and Biological Engineering and Computing, 2009, vol. 47, pp 257-265.
- [12] Ungureanu, M., Bigan, C., Strungaru, R., Lazarescu, V., Independent component analysis applied in biomedical signal processing, Measurement Science Review, 2004, vol. 4, Section 2, pp. 1-8
- [13] Von Klitzing, L., Low-frequency pulsed electromagnetic fields influence EEG of man, Physica Med, 1995, vol. 11, pp. 77-90
- [14] Watanabe, S., Pattern Recognition: Human and Mechanical. New York: Wiley, 1985
- [15] Wolpaw, J. R., McFarland, D. J., Neat, G. W., Forneris, C. A., An EEG-based brain-computer interface for cursor control, Electroencephalography and Clinical Neurophysiology, 1991, vol. 78, no. 3, pp. 252-259