Mutual Information-Based Feature Selection for Low-Cost BCIs Based on Motor Imagery

L. Schiatti, L. Faes, J. Tessadori, G. Barresi, and L. Mattos

Abstract*—In the present study a feature selection algorithm based on mutual information (MI) was applied to electro-encephalographic (EEG) data acquired during three different motor imagery tasks from two dataset: Dataset I from BCI Competition IV including full scalp recordings from four subjects, and new data recorded from three subjects using the popular low-cost Emotiv EPOC EEG headset. The aim was to evaluate optimal channels and band-power (BP) features for motor imagery tasks discrimination, in order to assess the feasibility of a portable low-cost motor imagery based Brain-Computer Interface (BCI) system. The minimal sub set of features most relevant to task description and less redundant to each other was determined, and the corresponding classification accuracy was assessed offline employing linear support vector machine (SVM) in a 10-fold cross validation scheme. The analysis was performed: (a) on the original full Dataset I from BCI competition IV, (b) on a restricted channels set from Dataset I corresponding to available Emotiv EPOC electrodes locations, and (c) on data recorded with the EPOC system. Results from (a) showed that an offline classification accuracy above 80% can be reached using only 5 features. Limiting the analysis to EPOC channels caused a decrease of classification accuracy, although it still remained above chance level, both for data from (b) and (c). A top accuracy of 70% was achieved using 2 optimal features. These results encourage further research towards the development of portable low cost motor imagery-based BCI systems.

I. Introduction

A Brain-Computer Interface (BCI) extracts meaningful information from brain signals and converts it into control signals for an external device. Such system has the potential to allow people suffering from mobility impairments to restore a non-muscular channel for communication and control [1]. In this context low cost, ease of use, and portability are key factors for the development of 'out-of-the-lab' systems, accessible to people in their home environment and usable in everyday life. Among existent BCI paradigms, detection of motor imagery is well suited for the development of self-paced systems, allowing the user to perform the control at will at any time [2]. These paradigms are based on the detection of event related desynchronization/ synchronization (ERD/ERS), indicating decrease/increase of signal spectral power in specific frequency bands (μ= 8-13 Hz and β= 13-30 Hz), during movement imagination or intent [3]. Recently the interest towards low cost EEG systems has been growing, as attested by an increasing number of related publications [4,5]. The most popular is the Emotiv EPOC[1], a wireless EEG headset including 14 sensors positioned as shown in Fig. 1 under label E. Recent works explored the possibility to develop motor imagery based on the EPOC system, obtaining encouraging results [5,6].

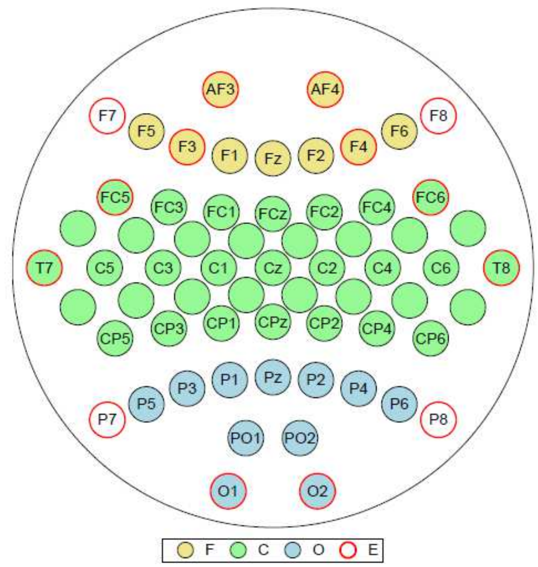

Figure 1. Scalp electrodes setup: frontal area (F); central area (C); occipital area (O), Emotiv EPOC channels (E).

Keeping a minimal number of electrodes is essential to develop a low cost and portable system. To this aim, several algorithms have been proposed to reduce the number of channels in BCIs [7]. Furthermore, optimal choice of frequency band and time segment reflecting ERD/ERS activation is highly subject-specific, and greatly affects the performance of the resulting BCI. The optimal combination of time, frequency and channel position, i.e. feature selection, is therefore essential in BCI design to achieve a good classification performance. In [8] a relevance analysis was applied offline to time and frequency domain parameters extracted from EEG data recorded with EPOC, and the resulting feature set was shown to allow left and right hand motor imagery discrimination.

In the present study a feature selection algorithm based on the conditional mutual information, originally presented in [9], is applied to the Dataset I from the BCI competition IV and on new data recorded using the Emotiv EPOC headset. The method allows the selection of optimal features that are at the same time maximally relevant to the class variable, and minimally redundant among each other. Band-power (BP) features are extracted from both datasets in different time windows during task execution. Optimal subsets of features are than selected to discriminate two out of three motor imagery classes, i.e. left hand, right hand, and foot, and results are validated in terms of classification accuracy.

L. Schiatti, J. Tessadori, G. Barresi and L. De Mattos are with the Istituto Italiano di Tecnologia, Genova, 16164 Italy (corresponding author e-mail: Lucia.Schiatti@iit.it).

L. Faes is with the Fondazione Bruno Kessler and BIOtech, University of Trento, Trento, 38122 Italy.

[1] https://emotiv.com/

TABLE I. RESULTS OF FEATURE SELECTION ON (A) DATASET I-BCI COMPETITION IV

	left hand-right hand						*left hand-foot*					
	Sb_all			**Sg_all**			**Sa_all**			**Sf_all**		
	TW1	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***
Mu	2(O)	2(C);3(C); 5(C)	1(C); 2(O)	-	-	-	1(C);2(C); 6(C)	1(C);2(O); 3(C);5(C);	1(C); 2(C)	1(C);3(C)	1(C);2(C); 6(O)	1(C)
B1	3(O); 4(F)	-	3(O)	1(C);3(O); 4(C)	**4(O)**; **5(F)**	-	**4(O)**	4(C); 7(C)	**4(F)**;5(C); 6(O)	**4(C)**	-	2(O); 4(C)
B2	5(C)	-	4(O)	2(C)	1(C); 2(C)	1(C);2(C); 3(C)	5(C)	6(C)	3(O)	-	5(C)	5(O)
B3	**1(O)**	**1(O)**;**4(C)**; 6(C)	-	5(O)	-	-	-	-	-	2(C);6(C)	3(C);**4(O)**	6(C)
B	6(C)	-	5(C);6(C)		3(O)	4(O)	3(F)	-	-	5(O)		3(F)

TABLE II. RESULTS OF FEATURE SELECTION ON (B) EPOC CHANNELS FROM DATASET I- BCI COMPETITION IV

	left hand-right hand						*left hand-foot*					
	Sb_EE			**Sg_EE**			**Sa_EE**			**Sf_EE**		
	TW1	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***
Mu	-	-	-	2(O)	-	-	1(C)	1(C);2(F)	1(C)	1(F);2(O)	1(C)	-
B1	2(F)	-	-	-	2(F)	-	-	-	-	-	-	2(F)
B2	3(O)	-	-	1(C)	-	-	3(F);5(F)	-	3(O)	-	-	1(C)
B3	1(O)	1(O)	-	-	-	-	2(C);4(F)	-	2(O)	-	2(C)	-
B	-	-	-	-	1(F)	-	-	-	-	-		-

II. MATERIALS AND METHODS

A. Dataset

The first considered dataset is the 100 Hz version of the data from calibration session of the BCI Competition IV Dataset I [10]. It consists of four subjects performing 100 trials for each one of two motor imagery tasks, respectively left hand and foot for subjects Sa and Sf, and left and right hand for subjects Sb and Sf. At the beginning of each trial, a visual cue was provided in a computer screen to the subject, who then started to perform a motor imagery task for 4 s. Task periods are interleaved with 4 s long rest periods. EEG signals were recorded from 59 channels, mostly distributed over sensorimotor areas, labeled as F, C, and O in Fig. 1. The second dataset was recorded at 128 Hz using the Emotiv EPOC headset (channels labeled as E in Fig. 1). It encompasses data from three subjects, S01, S02 and S03, performing 100 trials for each one of three motor imagery classes: left hand, right hand, and foot. The experimental paradigm was the same described for Dataset I. All subjects provided their informed consent to participate to the experiment.

B. Pre-processing

Raw signals were spatially filtered using common average re-referencing, and band-pass filtered using a zero-phase digital finite impulse response filter of 50th order, with Hamming window, to remove high frequency noise, slow artifacts, and extract information within the frequency bands of interest. Five frequency bands were considered for BP features extraction, covering μ and β bands, in which ERD/ERS phenomena connected to motor imagery are known to take place. Frequency bands were split as follows: *Mu={8-13}Hz*, *B1={13-18}Hz*, *B2={18-23}Hz*, *B3={23-28}Hz*, and *B={13-30}Hz*. Three overlapping epochs after cue presentation were extracted from each record for BP analysis: *TW1= [0-2]s*, *TW2= [1-3]s*, and *TW3=[2-4]s*. Logarithmic BP features were extracted for each frequency band and time window, using Welch periodogram estimation.

C. Feature Selection and Classification

A feature selection algorithm based on estimation of conditional mutual information was used to select optimal features sub-sets, considering BP features extracted in three cases: (a) the set of all 59 channels from Dataset I; (b) the selection of 10 out of 59 channels corresponding to available EPOC electrodes locations from Dataset I; (c) the set of 14 Emotiv EPOC channels from the second dataset.

Given a set of M features $F=\{f_1, f_2, \ldots, f_M\}$, and N cases (trials), a data matrix of size NxM is defined. In addition, the value of the corresponding discrete class variable C ($C=\{1,2\}$ for the two-class problem) is assigned to each case. This allows defining the process of optimal feature selection as the problem of finding a feature subset $S \subseteq F$ that best classifies the N cases. Here, mutual information (MI) and conditional MI (CMI) quantities were used to provide a feature selection criterion that evaluates two key properties of optimal features: the relevance of a feature to the class variable, and the redundancy of similar features. MI can intuitively be described as the information one variable carries about another variable and vice versa. For discrete variables, MI between each feature f_i and the class variable C, can be computed in terms of entropies:

$$\begin{aligned} I(f_i;C) &= H(f_i) + H(C) - H(f_i, C) \\ &= H(C) - H(C|f_i) \end{aligned} \qquad (1)$$

for $i=1,\ldots,M$, where $H(f_i)$, $H(C)$, $H(f_i,C)$ are entropies of f_i, C, and joint entropy respectively. Redundancy and relevance, which together constitute the criterion for features selection, can be jointly formulated as the conditional mutual information $I(f_i;C|S)$, i.e. the MI of the candidate feature f_i and the class variable C accounting for the information shared between each of them and the currently selected features set S. The CMI can be expressed in terms of entropies as:

TABLE III. RESULTS OF FEATURE SELECTION ON (C) EMOTIV EPOC DATASET

	left hand-right hand									*left hand-foot*								
	S01			**S02**			**S03**			**S01**			**S02**			**S03**		
	TW1	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***
Mu	-	-	-	-	-	1(F)	2(O)	-	-	1(O); 2(C)	-	1(C)	1(C)	-	-	1(O)	1(O)	2(F)
B1	2(C)	-	-	-	-		-	-	1(F)	-	-	2(C)	2(O)	1(F)	1(O)	-	-	-
B2	1(F)	2(F)	1(F); 2(O)	1(O)	-	2(C)	-	-	-	-	-	-	-	2(F)	-	-	3(O)	-
B3	-	1(C)	3(F)	-	1(O); 2(C)	-	-	-	-	-	1(C); 2(O)	-	-	-	2(F)	2(C)	2(F)	1(O)
B	-	-	-	2(O)	-	-	1(O)	-	-	-	-	-	3(O)	-	-	-	-	-

	right hand-foot								
	S01			**S02**			**S03**		
	TW1	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***	***TW1***	***TW2***	***TW3***
Mu	3(O)	1(F);3(O)	-	2(O);3(F)	3(O)	-	1(O)	1(O)	-
B1	1(F)	2(O)	1(C)	1(O)	1(C); 2(C)	1(O);2(F)	-	-	-
B2	2(C)	-	-	2(C)	-	-	-	-	1(C)
B3	-	-	2(C)	-	-	-	-	2(C)	-
B	-	-	-	-	-	-	-	-	2(O)

$$\begin{aligned} I(f_i;C|S) &= I(f_i;(C,S)) - I(f_i;S) \\ &= H(f_i,S) + H(C,S) - H(S) - H(f_i,C,S) \end{aligned} \tag{2}$$

The progressive selection of features on the basis of MI and CMI can be summarized in the following steps:

1. Among all features $f_i \in F$, find the feature f^* most relevant to the class variable C as the one which minimizes the conditional entropy $CE_i=H(C|f_i)=H(f_i,C)-H(f_i)$, thus maximizing the mutual information $I(f_i;C)$, computed as shown in (1). Set $S=\{f^*\}$.
2. To find the next optimal feature f^* to be added to S, compute for all candidate features $f_i \in F \backslash S$ the CMI, $I(f_i;C|S)$ as in (2), and select the one which maximizes it, $f^*=argmax_{fi}\, I(f_i;C|S)$.
3. Add f^* to S and repeat step 2 if CMI for f^* exceeds a threshold value, computed as the 95th percentile of a distribution of 100 estimations of CMI, each one derived using the values of f^* randomly permuted; otherwise stop the algorithm.

The CMI criterion in step 2 for optimal feature selection accounts for both relevance and redundancy: in fact, the feature selected is the one giving the largest amount of information about the class (relevance), which is not contained in the features already selected (redundancy). In this study, MI and CMI were estimated computing entropies from the probabilities assessed on the discretized values of the features. Discretization was performed by means of equidistant binning, using a number of quantization level $c=4$. Such small value for c was chosen because it allows the selection of a higher number of features with an equal information gain.

The optimal sub-sets of features, selected for each subject and time window as described above, were used to train and test a linear SVM classifier, implemented according to a 10-fold cross-validation scheme, in order to show the corresponding optimal classification rates.

III. RESULTS AND DISCUSSION

Results of the feature selection step are shown in Tables I-III, considering the cases (a), (b), and (c) as starting datasets for the feature extraction and selection algorithms. For each time window and frequency band, the feature selected is indicated by the number matching its rank in the selection procedure, followed by a letter within brackets, indicating the area where the channel is located (F: frontal, C: central, O: occipital, see Fig.1). In Table I, features relative to Emotiv EPOC electrodes locations are evidenced in bold type. Classification accuracies obtained using the optimal features sets in Tables I-III are reported in Fig. 2.

When using all channels from Dataset I, the highest accuracies for subjects Sa and Sg (0.79 and 0.83 respectively) were obtained for the window *TW2*, 1-3 s after cue (Fig.2). A high accuracy was shown also in *TW3*, 2-4 s, while a lower accuracy was observed immediately after the cue (*TW1*, 0-2 s). On the contrary, Sb and Sf showed higher accuracy in *TW1* than in later time intervals (0.67 and 0.765 respectively), but comparable with Sa and Sg classification accuracies in the same time window. This suggests that features selected in late time intervals are more effective in capturing the ERD/ERS mechanisms leading to discriminate between motor imagery classes, independently from the considered task, (left-right hand, or left hand-foot discrimination). Furthermore, the resulting accuracy depends more on the specific subject's brain patterns than on the type of imagined movements. Table I shows that the difference in task is captured by the frequency band of the optimal BP features selected. For left hand-foot discrimination the features with the highest rank are found in the μ band, while for left-right hand the most discriminative information is carried by features in low β band. Spatially, for both tasks the channels in central areas result as the most effective for discrimination. Higher classification accuracies were obtained when the first selected features are in central areas (Sa, Sf, Sg). For subject Sb, for whom the best accuracy was much lower than for other subjects, the first optimal feature was selected in occipital area, in β band (*TW1*).

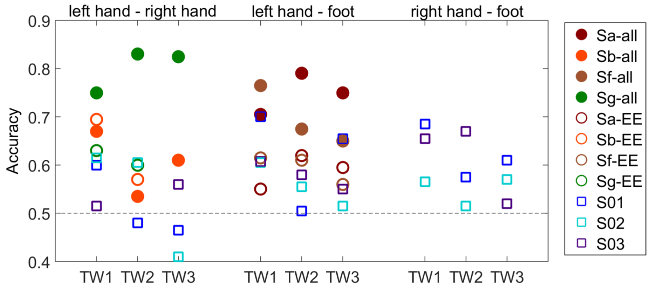

Figure 2. Classification accuracy using optimal features sets computed in time windows TW1-3 for subjects: Sa, Sb, Sf, Sg from full 59 channels (-all) and reduced to EPOC channels (-EE) Dataset I- BCI Competition IV; S01, S02, S03 from Emotiv EPOC dataset.

When restricting the analysis to Emotiv EPOC channels, the best classification accuracies for subjects Sa, Sf, and Sg showed a decrease of about 20%, though remaining above the chance level (0.62, 0.615, and 0.63 respectively). On the contrary, for subject Sb, the best classification accuracy increased by 4% (0.695). This can be explained looking at the results of feature selection in Table I-II. For subjects Sa, Sf, and Sg, when starting from the entire 59 channels set (Table I), features in EPOC channels were selected with rank from 4 to 5, mostly in O and F areas. This indicates that for the tasks considered the 10 EPOC channels carry much less information than other channels of the full set of electrodes. On the other hand, for subject Sb most informative features where found in O, and coincident with EPOC electrodes also when starting from the complete 59 channels dataset (Table I). Thus accuracy does not differ significantly when restricting the analysis to EPOC channels. The observation above highlights the influence of the first selected feature on the composition of the resulting optimal features set. When using the algorithm presented in Section II-C, all subsequent features are selected based on their relevance and redundancy compared to the already selected ones. This constraints the search potentially preventing the selection of the absolute best features set, as shown by the increase in Sb accuracy. The dimensionality of the optimal features set, leading to the best classification accuracy, when starting from the EPOC channels from Dataset I, ranged from 2 to 3. This suggests that the EPOC channels are able to determine features explaining most of the information connected to class discrimination, and that the first one or two selected features are mostly responsible for the final classification accuracy. Relevant information carried by subsequent features rapidly decreases with the number of features selected.

Best classification accuracies obtained on Emotiv EPOC dataset were 0.7, 0.615 and 0.67 for subjects S01, S02 and S03, achieved respectively for left hand-foot (*TW1*), left-right hand (*TW1*) and right hand-foot (*TW2*) discrimination (Fig. 2). As visible in Table III, the corresponding first feature selected belonged to O area for all three subjects, in μ band for subjects S01 and S03, and in β band for subject S02. This confirmed that features in μ band are more effective in discriminating tasks involving foot motor imagery, while features in β band bring more useful information for left-right hand motor imagery discrimination. The importance of channels in C area was confirmed by the observation that best accuracies (0.7, 0.685) were obtained for S01 in both hand-foot tasks, in *TW1*, when also features in C are selected.

IV. Conclusions

Results of feature selection on the 59 channels Dataset I from BCI competition IV showed that optimal features vary considerably across subjects, thus confirming the importance of subject-adaptive feature selection algorithms in the development of effective BCIs. The number of optimal features selected as relevant and not redundant varied from 5 to 7 for the best classification accuracy achieved on each subject data. This confirmed the feasibility of motor imagery-based BCI systems using a low number of EEG channels. Results also attested the usefulness of low cost EEG system like Emotiv EPOC, for the development of such BCIs. Even not encompassing electrodes in the motor area, BP features extracted from EPOC channels were found to be relevant, especially in hand-foot discrimination, allowing an offline classification accuracy of 70% using only 2 features. Future work will be devoted to assess the relevance of different time and frequency domain features, both on healthy and paralyzed people, and to evaluate their effectiveness when implemented in a real time low cost BCI system, providing the user with continuous online feedback.

Acknowledgment

The research leading to these results has received funding from Fondazione Roma under grant agreement TEEP-SLA.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain–computer interfaces for communication and control,” Clinical neurophysiology, vol. 113, no. 6, pp. 767–791, 2002.

[2] B. Blankertz, G. Curio, and K.-R. Muller, “Classifying single trial eeg: Towards brain computer interfacing,” Advances in neural information processing systems, vol. 1, pp. 157–164, 2002.

[3] C. Neuper, M. W¨ortz, and G. Pfurtscheller, “Erd/ers patterns reflecting sensorimotor activation and deactivation,” Progress in brain research, vol. 159, pp. 211–222, 2006.

[4] M. Duvinage, T. Castermans, M. Petieau, T. Hoellinger, G. Cheron, and T. Dutoit, “Performance of the emotiv EPOC headset for p300-based applications,” Biomedical engineering online, vol. 12, no. 1, p. 56, 2013.

[5] J.-A. Martinez-Leon, J.-M. Cano-Izquierdo, and J. Ibarrola, “Are low cost brain computer interface headsets ready for motor imagery applications?” Expert Systems with Applications, vol. 49, pp. 136–144, 2016.

[6] T. Mladenov, K. Kim, and S. Nooshabadi, “Accurate motor imagery based dry electrode brain-computer interface system for consumer applications,” in Consumer Electronics (ISCE), 2012 IEEE 16th International Symposium on. IEEE, 2012, pp. 1–4.

[7] A. Barachant and S. Bonnet, “Channel selection procedure using riemannian distance for bci applications,” in Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on. IEEE, 2011, pp. 348–351.

[8] J. Hurtado-Rincon, S. Rojas-Jaramillo, Y. Ricardo-Cespedes, A. M. Alvarez-Meza, and G. Castellanos-Dominguez, “Motor imagery classification using feature relevance analysis: An emotiv-based bci system,” in Image, Signal Processing and Artificial Vision (STSIVA), 2014 XIX Symposium on. IEEE, 2014, pp. 1–5.

[9] A. Tsimpiris, I. Vlachos, and D. Kugiumtzis, “Nearest neighbor estimate of conditional mutual information in feature selection,” Expert Syst. Appl., vol. 39, no. 16, pp. 12 697–12 708, Nov. 2012.

[10] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. M¨uller, and G. Curio, “The non-invasive berlin brain–computer interface: fast acquisition of effective performance in untrained subjects,” NeuroImage, vol. 37, no. 2, pp. 539–550, 2007.