Random Forest and Filter Bank Common Spatial Patterns for EEG-Based Motor Imagery Classification

Maouia Bentlemsan, ET-Tahir Zemouri, Djamel Bouchaffra, Bahia Yahya-Zoubir, and Karim Ferroudji

Design and Implementation of Intelligent Machines DFM
Centre for Development of Advanced Technologies
Baba Hassen, Algiers, Algeria
{mbentlemsan, tzemouri, dbouchaffra, byahyazoubir, kferroudji}@cdta.dz

Abstract—We propose using filter bank common spatial pattern (FBCSP) feature extraction algorithm, and random forest (RF) technique for classification of EEG motor imagery signals. FBCSP algorithm allows extracting features and dealing with subject variability by automatic selection of frequency bands. Performing random forest in the classification avoid the use of feature selection step, since RF combine a bagging (bootstrap aggregation) and a random selection of features. We evaluate our system on the dataset 2b of the Brain-Computer Interface BCI Competition IV. The proposed method is promising since it has outperformed the results obtained in BCI Competition for some subjects in term of accuracy and kappa.

Keywords—brain-machine interface (BMI); brain-computer interface (BCI); electroencephalogram EEG; motor imagery; random forest (RF); filter bank common spatial pattern (FBCSP)

I. INTRODUCTION

Brain machine interface (BMI) is a direct communication pathway between the brain and an external device. It allows individuals with motor disabilities to interact with the outside world. Electroencephalography (EEG) is the most widely neuroimaging technique used in BMI. This acquisition modality has several advantages: non-invasive, high temporal resolution, relative low cost, and high portability. Once the intentions have been measured by EEG, an appropriate feature vector is extracted from the preprocessed EEG signals. The acquired signal is corrupted with background noise and different artifact: eye blinking, cardiac activity, muscular activity, etc. The preprocessing is an important stage, since it reduces the artifacts influence, and allows the extracting meaningful information after artifacts removal from the recorded signal. The obtained features vectors are translated into commands by using classifiers.

The EEG activity investigated in this paper is sensorimotor activity, specifically motor imagery, which corresponds to the situation when a person imagine some movements, for example, right hand, left hand, etc.

The classification of mental activity related to motor imagery is very common in the literature. Movement or imaging movement causes a decrease/increase in µ activity over sensorimotor cortex, named event-related desynchronization (ERD)/event-related desynchronization (ERS) [1]. When a motor movement is imagined, the bands frequencies µ and beta respectively 8-15 Hz and 15-25 Hz rhythms are activated, and exhibit a decrease in amplitude prior to the actual movement (ERD). An increase in the amplitude in beta band is observed when the movement ceases (ERS).

Feature extraction in motor imagery is an important step, which allows selecting the most relevant features in the frequency bands specific to the imaging movement.

Common spatial pattern CSP [2] is the widely feature extraction method used in motor imagery, subject specific band is chosen manually and it is sensitive to non-stationary and noisy data. To address with this problem, modified several CSP were proposed: FBCSP using a bank of frequencies [3], Discriminative CSP [4], Regularizing CSP (RCSP) which adds prior information into the learning process [5].

Several types of classification procedures have been used in BMI systems. Linear discriminant analysis (LDA) has been exploited with success in motor imagery based BMI [1]. Support vector machine (SVM) [6] and hidden Markov models (HMM) [7,8] have also been a focus of investigation [9]. Multi-layer perceptron (MLP) is the widely neural network (NN) utilized in BMI applications [10,11]. K-nearest neighbors (K-NN) have also been exploited and were efficient in the case of a low dimensional feature space [12].

Another approach to enhance brain signals classification consists of combining several classifiers into one. In fact, Boosting have been used in a few BMI applications, it has been integrated within MLP [13]. Classifiers ensemble is the most widespread technique for combining classifiers in BMI research. For example, voting with learning vector quantization neural network (LVQNN) [14], or MLP [15]. Stacking is another way to combine classifiers, this technique has been integrated in BMI using HMM as level-0 classifiers, and SVM as meta-classifier [16].

Furthermore, EEG signal is inherently non-stationary, and very noisy. Some classifiers are sensitive to noise, others to overfitting. To cope with these specific problems in BMI systems, regularized terms are often added to the classifier and combination of different classifiers is recommended. To deal with the particularity of the EEG signal, we propose to use FBCSP feature extraction algorithm, with a classification
method based on classifiers ensemble: Random forest (RF). Unlike many other classifiers RF [17]:

- Achieves a good accuracy in classification and is robust to outliers and noise.
- Is faster than bagging or boosting.
- Gives useful internal estimates of error, strength, correlation and variable importance.
- Does not overfit.

II. PROPOSED METHOD

In this work, we used FBCSP for feature extraction and Random forest for classification. We don't perform feature selection step in FBCSP, we use directly the classification stage by the Random forest algorithm. In the sequel, we present the FBCSP and the Random Forest classifier.

A. Filter Bank Common Spatial Pattern

The filter bank common spatial pattern (FBCSP) is a feature extraction algorithm [3], based on the common spatial pattern (CSP) used in motor imagery [2].

FBCSP performs in 4 progressive steps: filter bank, spatial filtering using CSP algorithm, feature selection and classification of the selected features.

We estimate band power (BP) features from EEG signal using at a first step multiple Chebyshev type II band-pass filters. In the second step, spatial filtering using CSP is performed on each frequency band in order to compute the relevant features of the EEG signal. CSP algorithm is powerful in calculating spatial filters for detecting ERD/ERS [2]. The EEG measurement is linearly transformed using the spatial filter obtained with the CSP algorithm (1)

$$Z = \tilde{W}^T E.$$ (1)

$E \in \mathbb{R}^{c \times n}$ denotes the single-trial EEG measurement, $Z \in \mathbb{R}^{c \times \omega}$ are the filtered signals, and $\tilde{W} \in \mathbb{R}^{c \times \omega}$ denotes the CSP projection matrix. $c$ is the number of channels, and $\omega$ the number of EEG samples per channel.

The transformation matrix $\tilde{W}$ yields features whose variance is optimal for discriminating the 2 classes of the EEG signal. $W$ is computed by solving the eigenvalue decomposition problem.

$$\Sigma_1 \tilde{W} = (\Sigma_i + \Sigma_c)WD.$$ (2)

$\Sigma_i$ and $\Sigma_c$ are estimates of the covariance matrix of the band-pass filtered EEG measurements of the respective motor imagery task. $D$ is the diagonal matrix that contains the eigen values of $\Sigma_i$.

The CSP features of the $i$th trial for the EEG measurements are then given by:

$$f_i = \log \frac{\text{diag}(\tilde{W}^T E_i E_i^T \tilde{W})}{\text{trace}[\tilde{W}^T E_i E_i^T \tilde{W}]}.$$ (3)

where $\tilde{W}$ is subtracted from $W$.

B. Random Forest Algorithm

Random Forest [17] is an ensemble of method-based learning algorithm. RF is composed of a set of tree classifiers. Each tree is constituted of nodes and edges. The obtained ensemble classifies new data points through a consensus obtained of the predictions of each classifier (refer to Fig 1).

This method combines a bagging (bootstrap aggregation) and a random split selection. Each tree is obtained through a separate bootstrap sample from the data set and each tree classifies the data. A majority vote among the trees provides the final result. The random forest algorithm is defined by the following steps:

- Build $k$ trees bootstrap samples from the data.
- Grow an un-pruned tree for each of the bootstrap samples.
- At each node, randomly sample $n$-try of the predictors and choose the best split from among those variables.
- Predict new data by combining the predictions of the $k$ tree.

III. EXPERIMENT AND RESULTS

A. Data Set Description:

The Graz data set B of BCI competition is downloaded from [18], its consist of EEG data from 9 subjects (22 EEG and 3 EOG electrodes) with two classes, namely the motor imagery of left hand (class 1) and right hand (class 2). Three bipolar recordings (C3, Cz, and C4) were recorded with a sampling frequency of 250 Hz (refer to Fig. 2). Each subject participated in two sessions recorded on two separated days within two weeks. Each session consisted of six runs with
ten trials each and two classes of imagery. This resulted in 20 trials per run and 120 trials per session [18].

Each subject participated in two screening sessions without feedback recorded on two different days within two weeks. For the three online feedback sessions four runs with smiley feedback were recorded (refer to Fig 3b).

![Figure 2. Position of EEG electrodes (from [19]).](image)

![Figure 3. (a) Screening](image)

(b) Smiley Feedback

Figure 3. Timing scheme of the paradigm.

B. Results

We have applied Chebyshev filter that covers 4-40 Hz, to obtain 9 band-pass filters. CSP is then used on each band to extract the pair of features. The 18 extracted features are then used as input of the classifiers. We have compared in this section the performance of RF with SVM, the results are shown in Table I.

SVM has been successfully applied to various pattern recognition problems such as character recognition [20], and medical imaging [21].

The parameters of Random forest model, as the number of trees, are determined for each subject to achieve the best classification rates.

Due to the subject variability, we have also searched for the best time segment for each subject.

The training data is used to fix the parameters of the classifiers. The results are based on the system’s performance using two unseen test sessions.

In order to evaluate the performance of our method, we have used the accuracy and kappa value criteria [22].

<table>
<thead>
<tr>
<th>Subject</th>
<th>SVM</th>
<th>RF</th>
<th>SVM</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>59.65</td>
<td>72.81</td>
<td>0.17</td>
<td>0.46</td>
</tr>
<tr>
<td>2</td>
<td>53.88</td>
<td>66.53</td>
<td>0.08</td>
<td>0.33</td>
</tr>
<tr>
<td>3</td>
<td>52.17</td>
<td>65.65</td>
<td>0.08</td>
<td>0.30</td>
</tr>
<tr>
<td>4</td>
<td>95.11</td>
<td>97.07</td>
<td>0.90</td>
<td>0.94</td>
</tr>
<tr>
<td>5</td>
<td>66.67</td>
<td>87.91</td>
<td>0.34</td>
<td>0.76</td>
</tr>
<tr>
<td>6</td>
<td>64.14</td>
<td>78.88</td>
<td>0.29</td>
<td>0.58</td>
</tr>
<tr>
<td>7</td>
<td>62.07</td>
<td>75.43</td>
<td>0.24</td>
<td>0.51</td>
</tr>
<tr>
<td>8</td>
<td>65.22</td>
<td>90.43</td>
<td>0.31</td>
<td>0.81</td>
</tr>
<tr>
<td>9</td>
<td>75.10</td>
<td>83.27</td>
<td>0.50</td>
<td>0.69</td>
</tr>
</tbody>
</table>

The classification results illustrate that Random forest classifier produces higher classification rates and kappa values than SVM. This can be justified by the fact that SVM is sensitive to irrelevant features unlike RF classifier which doesn’t require a feature selection stage.

In order to compare our approach to others, Table II presents the results of BCI competition and our results in term of kappa value.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Method</th>
<th>Algorithm</th>
<th>kappa</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Zheng Yang Chin</td>
<td>FBCSP + Naïve Bayes classifier</td>
<td>0.60</td>
<td>0.40</td>
<td>0.21</td>
<td>0.22</td>
<td><strong>0.95</strong></td>
<td><strong>0.86</strong></td>
<td>0.61</td>
<td>0.56</td>
<td><strong>0.85</strong></td>
<td>0.74</td>
</tr>
<tr>
<td>2</td>
<td>Proposed method</td>
<td>FBCSP + RF</td>
<td><strong>0.59</strong></td>
<td><strong>0.46</strong></td>
<td><strong>0.33</strong></td>
<td><strong>0.30</strong></td>
<td>0.94</td>
<td>0.76</td>
<td>0.58</td>
<td>0.51</td>
<td>0.81</td>
<td>0.69</td>
</tr>
<tr>
<td>3</td>
<td>Huang Gan</td>
<td>Common spatial subspace + LDA</td>
<td><strong>0.58</strong></td>
<td>0.42</td>
<td>0.21</td>
<td>0.14</td>
<td>0.94</td>
<td>0.71</td>
<td><strong>0.62</strong></td>
<td><strong>0.61</strong></td>
<td>0.84</td>
<td><strong>0.78</strong></td>
</tr>
<tr>
<td>4</td>
<td>Damien Coyle</td>
<td>Log variance + LDA &amp; SVM</td>
<td>0.46</td>
<td>0.19</td>
<td>0.12</td>
<td>0.12</td>
<td>0.77</td>
<td>0.57</td>
<td>0.49</td>
<td>0.38</td>
<td>0.85</td>
<td>0.61</td>
</tr>
<tr>
<td>5</td>
<td>Shaun Lodder</td>
<td>Wavelet packet + LDA</td>
<td><strong>0.43</strong></td>
<td>0.23</td>
<td>0.31</td>
<td>0.07</td>
<td>0.91</td>
<td>0.24</td>
<td>0.42</td>
<td>0.41</td>
<td>0.74</td>
<td>0.53</td>
</tr>
<tr>
<td>6</td>
<td>Jaime Fernando Delgado Saa</td>
<td>Spectral features + NN</td>
<td><strong>0.37</strong></td>
<td>0.20</td>
<td>0.16</td>
<td>0.16</td>
<td>0.73</td>
<td>0.21</td>
<td>0.19</td>
<td>0.39</td>
<td>0.86</td>
<td>0.44</td>
</tr>
<tr>
<td>7</td>
<td>Yang Ping</td>
<td>Band power features + LDA</td>
<td><strong>0.25</strong></td>
<td>0.02</td>
<td>0.09</td>
<td>0.07</td>
<td>0.43</td>
<td>0.25</td>
<td>0.00</td>
<td>0.14</td>
<td>0.76</td>
<td>0.47</td>
</tr>
</tbody>
</table>
The best results appear in bold. Table II shows that our proposed method outperforms other existing methods in subjects 1, 2, and 3, and are close to the best for subjects 4, 6, 7, and 8. The average kappa value (kappa=0.59) is ranked second when compared to the competition results.

We can see that the data from subject 4 and subject 8 can be identified best, whereby subjects 1, 2, and 3 present some challenges. These results are consistent over all approaches.

IV. CONCLUSION AND FUTURE WORK

We have used FBCSP feature extraction algorithm and Random forest classifier, for two class EEG motor imagery-classification. The experiments show promising results but it is not worthy that the accuracy of the classifier is subject-dependent. Due to subject variability, no method achieved good results on all subjects. FBCSP can find the best band frequency for each subject, but the time segment in evaluation and training phase allows enhancing the classification rate.

As future work, we intend to modify FBCSP in order to find the best time segment for each subject automatically in the training and evaluation stages, and then we will focus on the exploitation of these features for a multi-class motor imagery classification.

REFERENCES


