Characterizing EEG Dynamic Features for Motor Imagery Classification

Cheng Wan and **Liqing Zhang**
Brain-Like Computing Lab,
Shanghai Jiao Tong University,
P. R. China

Abstract

Electroencephalography (EEG) is an effective channel for brain computer interaction (BCI). One of the critical issues in BCI is to identify specific patterns robustly from EEG measurement. Due to low signal-to-noise ratio, it is still challenging to recognize dynamic patterns of motor imagery from EEG data. In this work, we propose a new classification method named FVFBCSP (**F**ilter **B**ank **C**ommon **S**patial **P**attern for **F**eature **V**isualization), which is a variant of FBCSP. Feature visualization is conducted on FVFBCSP by calculating the gradient of the score of a specified class with regard to the input to the classifier. This method reveals intrinsic features of EEG signal in three main domains (space, frequency, and time) significantly. The experimental results on two datasets show that the proposed method achieves more significant performance than popular methods in their specialty.

1 Introduction

EEG is now a widely used approach to identify motor imagery patterns. A variety of methods have been developed to increase the classification accuracy in the past decades. Denoising methods are aiming to remove physiological artifacts such as electrooculogram (EOG) and electromyogram (EMG) in EEG data, and several methods such as principal component analysis (PCA), independent component analysis (ICA), wavelet, or mixed algorithms allows researchers to study clean data (Kandaswamy et al. 2005). For motor imagination task, common spatial patterns (CSP) (Müller-Gerking, Pfurtscheller, and Flyvbjerg 1999) is a classic method for extracting or selecting features. Some variants (Ang et al. 2008; Liu et al. 2016) have been developed based on it. Other methods such as tensor decomposition (Kolda and Bader 2009) and power spectral density (Gwin et al. 2010) are also employed as EEG feature extractors. Concerning EEG pattern classifiers, support vector machine, neural network, Bayes network, graphical models and nearest neighbor classifiers have been successfully applied to classify EEG patterns (Lotte et al. 2007). Based on data preprocessing, feature extraction algorithms, and advanced classifiers, we are able to construct a BCI system which controls facility only using brain waves.

However, recognizing dynamic patterns of motor imagery from EEG data is still a challenging problem. Visualizing features of EEG data for motor imagery allows researchers to understand the characteristic of brain waves. Neuroscientists could rely on the results to study the brain behavior during motor imagery. For example, doctors could analyze the status of the brain of a stroke patient. It also provides the significance of each component of the input data, with which people can select main components to speed up the classification process, reduce computation requirement, and mitigate overfitting problems. Some feature visualization methods for EEG data have been developed in the past decades. The result of some feature selection methods can be applied directly to feature display. CSP is one of the most popular methods among them (Müller-Gerking, Pfurtscheller, and Flyvbjerg 1999). However, it can only show the feature in the space domain, and we cannot get any information in frequency domain and time domain. Observing the energy of time-frequency distribution is another traditional method (Pfurtscheller and Da Silva 1999). This method is quite rude that sometimes we cannot distinguish whether the part of abnormalities is related to a particular motor imagination task. Moreover, this approach can only be used to capture dynamic patterns, and cannot describe features in the space domain or frequency domain quantitatively.

In computer vision field, several methods have been developed to elucidate the significance of visual features on the contribution to pattern classification. Computing the derivative of the score of a specified class with regard to the input (Simonyan, Vedaldi, and Zisserman 2013) and occluding parts of the image (Zeiler and Fergus 2014) are two most obvious ways, which can be applied to a variety of classifiers. Other works are also useful, but can only be adopted for specific classification models (Dai et al. 2017; Zhou et al. 2016).

In this paper, we specify computing the gradient of the score of a specified class with respect to the input to the classifier as the feature visualization method. To achieve reasonable result, we also modify the framework of filter bank common spatial pattern (FBCSP) (Ang et al. 2008), and propose FVFBCSP (**F**ilter **B**ank **C**ommon **S**patial **P**attern for **F**eature **V**isualization). The proposed method visualizes features in three main domains (space, frequency, and time). Experiments are conducted on two datasets. The experimental results show that the proposed method achieves more significant performance than popular methods in their specialty.

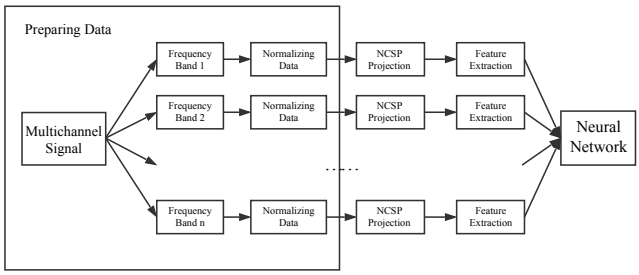

Figure 1: The framework of FVFBCSP

The paper is organized as follows: Section 1 is the introduction. Section 2 presents our feature visualization method. The experimental results are given in Section 3. Further discussions and conclusions are given in Section 4 and Section 5 respectively.

2 Methods

Filter bank common spatial pattern (FBCSP) is one of the most successful methods for classifying EEG data of motor imagination. Calculating the gradient of the score of a specific class with regard to the input is another successful method for visualizing features in computer vision field. However, applying this feature visualization method to FBCSP directly will achieve unreasonable results. Here we introduce a variety of FBCSP named FVFBCSP. At the end of this section, we provide the advantages of FVFBCSP for feature visualization compared with FBCSP.

FBCSP for Feature Visualization

Collecting EEG data is notoriously costly. More often than not, subjects ought to concentrate on the experiment for hours to generate only small amount data. Therefore, the number of trials for each subject is not large. To train a good model on validation data, the number of parameters in this model should be appropriate. In our work, we choose FBCSP, which is a commonly used method for classifying EEG data for motor imagination, as the classifier. To achieve reasonable results, we also modify the framework of FBCSP and propose FVFBCSP. The framework of FVFBCSP is presented in Figure 1 and the detailed implementation is described as follows:

Frequency Band Filter. For each input EEG signal, we filter it with several band-pass filters. In our work, we specify the band as [4Hz, 8Hz], [6Hz, 10Hz], [8Hz, 12Hz], $\cdots$, [42Hz, 46Hz]. After this process, we transfer the data into three domain: space domain, time domain, and frequency domain. Now, we regard this three-mode data as input data.

Normalizing Data. Normalizing the data in the same frequency band and same channel. Suppose the mean value and standard deviation over all values in frequency band f and channel c is $\mu_{f,c}$ and $\sigma_{f,c}$ respectively. For the value $x_{f,c,t}$ at frequency band f, channel c, and time t, we replace it by $\frac{x_{f,c,t}-\mu_{f,c}}{\sigma_{f,c}}$.

Normalized Common Spatial Pattern. Common spatial pattern (CSP) method is a commonly used method for extracting features of EEG data. The main idea is to use a linear transform to project the EEG data into spatial subspace. This kind of transform maximizes the ratio of variance between two classes. To be more precise, suppose the average of preprocessed multichannel data in two different classes are X_1 and X_2 with t_1 and t_2 samples respectively. We are to find a full-rank transform matrix W such that the covariance matrix of $Z_1 = WX_1$ and $Z_2 = WX_2$ is diagonal, and satisfies

$$\frac{Z_1 Z_1^T}{t_1} + \frac{Z_2 Z_2^T}{t_2} = I.$$

When we get W, for each data X, the variance of values in each row of WX can be regarded as a feature. In practice, people choose $2m$ rows of W which achieve m largest variances of WX_1 or m smallest variances of WX_1. Since variance is usually propotional to the energy of a signal, people use $f_i = \log \frac{\mathrm{var}(w_i X)}{\sum_{j=1}^{2m} \mathrm{var}(w_j X)}$ as a feature, where w_i is the ith row of W.

In our work, we constrain the variance of each row in the projection matrix as 1, which is called *Normalized Common Spatial Pattern (NCSP)*. More precisely, for any given row vector w in W, we replace it with $w_{\mathrm{N}} = \frac{w}{||w||_2}$, where $||w||_2$ is the L2 norm of vector w. After this process, the ratio between two variances will not change.

In this paper, for k-classes classification, where $k > 2$, we use one-versus-rest NCSP features for the normalized data in each frequency band. For each class, we calculate the projective matrix to separate this class and all other $k-1$ classes. In this work, we specify $m = 2$. Therefore, we will generate $4k$ features for each filter band.

One thing needs to be highlighted is that during the process of calculating W, the validation data should not be included in the computing.

Fully Connected Neural Network. Taking advantage of NCSP features, we can build a small classifier to complete the classification task. In our work, we use a two-layer fully connected neural network with a softmax layer as the final classifier. A fully connected neural network contains several layers of nodes. The output of each node is given by

$$x_j^l = f\left(\sum_i k_{ij}^l x_i^{l-1} + b_j^l\right)$$

where x_j^l and x_i^{l-1} denotes the output value at level l and $l-1$ respectively, and $f(\cdot)$ is an activation function. k_{ij}^l and b_j^l are trainable parameters. In our two-layer fully connected neural network, we specify the activation function as ReLu function. The first layer of units are extracted features, and the last layer of units denotes the scores of specific input for each class.

Softmax layer is keeping up with the last layer. The output of it is

$$p_i = \frac{\exp\left(x_i^m\right)}{\sum_j \exp\left(x_j^m\right)}$$

where p_i is the final probability of input data to be classified as class i, and x_i^m is the output of the final layer.

At last, we use the negative log likelihood as the loss function and use stochastic gradient descent to learn the parameters of the network.

Feature Characterization

Calculating gradient features is a trustworthy method to visualize the feature of an image in computer vision field, which is easy to implement and highly efficient. The main idea is to visualize the data gradient of a score function.

More precisely, let $S_c(X)$ be the score of class c with input X. Here, the definition of the score is the input of softmax layer rather than normalized probability. For any specific input X_0, we regard

$$w = \left.\frac{\partial S_c}{\partial X}\right|_{X_0}$$

as the weight vector of the input, whose absolute value indicates the salience map of X_0.

Intuitively, the gradient of the input of softmax layer shows how greatly one part of input influences the classification. Mathematically, the first-order Taylor expansion of $S_c(X)$ is

$$S_c(X) \approx w^T X + b$$

Hence, w shows the importance of each part of the input, which can be used as features.

Interpretation

In this part, we explain why we adopt FVFBCSP as the classifier rather than FBCSP.

We can separate calculating the gradient process above into three independent parts $\mathcal{P}_{\text{NCSP}}$, $\mathcal{P}_{\text{rule}}$ and $\mathcal{P}_{\text{NN}}$, which indicates calculating the gradient of NCSP process, the gradient of feature extraction process, and the gradient of classification process respectively.

The main difference between FBCSP and FVFBCSP is that FVFBCSP adds data normalization part, and replace CSP matrix with NCSP matrix. We classify the above three processes into two classes: calculating data-based features ($\mathcal{P}_{\text{NCSP}}$ and $\mathcal{P}_{\text{NN}}$) and calculating rule-based features ($\mathcal{P}_{\text{rule}}$). We should reduce the contribution of the process in the second class. Rule-based features do not reflect characteristics of data directly. Moreover, calculating rule-based features might introduce singularity. For example, in FBCSP, it needs to calculate the logarithm of variance. If the variance is quite small, its derivative would be a huge number, which influences the final saliency map tremendously. We should find a way to overcome these problems. Here we can only deal with the framework of FBCSP rather than that of extracting features.

More specifically, the task of $\mathcal{P}_{\text{rule}}$ is as follows: given the gradient of the classification score with respect to the $2m$ extracted features in every frequency band, calculate the gradient of the classification score with respect to the projected signal in each frequency band.

For any frequency band, suppose its projected multichannel signal is y_1, y_2,$\cdots$, y_{2m}, then the features are $f_i = \log \frac{\text{var}(y_i)}{\sum_{j=1}^{2m} \text{var}(y_j)}$. Suppose the gradient of the classification score with respect to the extraced $2m$ features in this frequency band is $w_{\text{NN}} = (s_1, s_2, \cdots, s_{2m})$, and the output of $\mathcal{P}_{\text{rule}}$ is w_{rule}. According to the chain rule of derivatives,

$$w_{\text{rule},(i,j)} = \sum_{k=1}^{2m} s_k \cdot \frac{\partial \text{var}(y_i)}{\partial y_{i,j}} \cdot \frac{\partial f_k}{\partial \text{var}(y_i)}$$

where $w_{\text{rule},(i,j)}$ is the ith row jth colum of w_{rule}, which denotes the significance of timestamp j of the ith channel.

For the second part $\frac{\partial \text{var}(y_i)}{\partial y_{i,j}}$, although it can be arbitrary small, we do not think it introduces singularity. Two most important features in time domain for EEG features are event-related synchronization (ERS) and event-related desynchronization (ERD) (Pfurtscheller and Da Silva 1999). For both phenomena, after normalizing the data, its absolute value would be greater if it is event-related timestamp. In this way of view, the absolute value of $\frac{\partial \text{var}(y_i)}{\partial y_{i,j}}$ reflects the salience at each timestamp. Now we focus on $\frac{\partial f_k}{\partial \text{var}(y_i)}$.

$$\frac{\partial f_k}{\partial \text{var}(y_i)} = \begin{cases} \frac{\sum_{l \neq i} \text{var}(y_l)}{\text{var}(y_i) \sum_l \text{var}(y_l)} & , \quad i = k \\ -\frac{1}{\sum_l \text{var}(y_l)} & , \quad i \neq k \end{cases}$$

Based on this observation, we figure that it will introduce singularity ł the scale of $\text{var}(y_k)$ will influence the result tremendously. If its value is close to 0, the contribution of $\mathcal{P}_{\text{rule}}$ will become very large, and singularity is introduced. For EEG data, this phenomenon possibly happens. After normalizing the input, the variance of the signal in the high-frequency band is always significantly lower than that in the low-frequency band. Moreover, since signals in the high-frequency band contain a lot of noise, NCSP matrix is often irrelevant to it. The result will probably show that data in the high-frequency band is much more important for motor classification than data in the low-frequency band, which is not acceptable. Therefore we impose some constraints on $\text{var}(y_i)$ in a reasonable bound, both the problems that rule-based features influence final results significantly and that $\mathcal{P}_{\text{rule}}$ introduces singularity will be solved. In our FVFBCSP, the experiments below show that the variance is concentrated under most cases.

Therefore, we can say that in our model the frequency domain features are extracted by $\mathcal{P}_{\text{NN}}$, the time domain features are derived by $\mathcal{P}_{\text{rule}}$, and the space domain features are derived by $\mathcal{P}_{\text{NCSP}}$.

3 Experiments

Experiment Set Up

We use two datasets to evaluate our method.

(1) Dataset I: Dataset I comes from BCI Competition IV dataset IIa (Brunner et al. 2008). EEG data is collected by 22 electrodes, bandpass-filtered between 0.5Hz and 100Hz, and sampled at 250Hz. Nine subjects participated in this experiment. Each subject is recorded 576 trails of four kinds of motor imagery: the left hand, the right hand, the foot and the tongue. A time segment from 0.5s to 2.5s is chosen for analysis. Data with artifacts are removed before we use them.

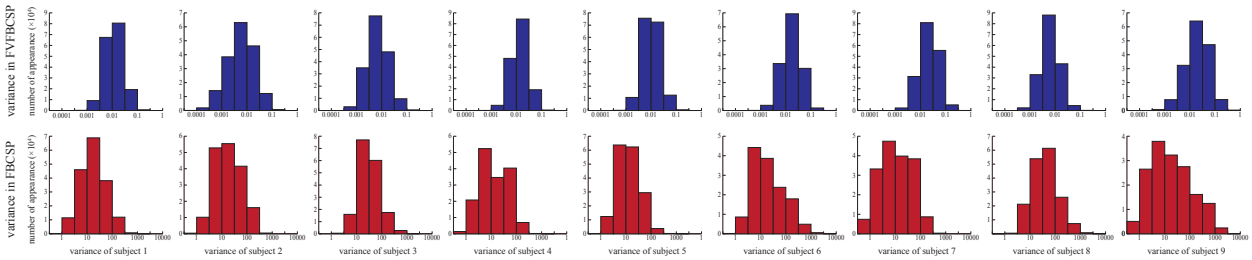

Figure 2: The variances of projected signals.

(2) Dataset II: Dataset II is recorded by Liu et al. 2016. This dataset is collected from five unilaterally paralyzed stroke patients during two months' rehabilitation who perform motor imagination of their disabled (left or right) arm in a BCI combined with Functional Electrical Stimulation (BCI-FES) rehabilitation training system. EEG data is recorded by a 16-channel (FC3, FCZ, FC4, C1-C6, CZ, CP3, CPZ, CP4, P3, PZ, and P4) g.USBamp amplifier at a sampling rate of 256 Hz. Patients have to complete basic motor imagery related tasks for five sessions, and each session comprises forty trials which last for 240 seconds. We extract a time segment starting from 0s to 4s after the visual cue for analysis. Artifacts in EEG data has been removed before we use them.

Results of Dataset I

We first evaluate the effectiveness of our method on the dataset IVa from famous BCI competition IV. We split the data into five groups randomly and run cross validation. It proves that FVFBCSP can constrain the variance of projected data in a small range under most cases. The comparison with popular methods shows that the features captured by our methods are reliable.

The Distribution of Variance for Projected Data Figure 2 shows the variance distribution in FVFBCSP. The results are conducted on test data. We also display that in FBCSP. The comparison verifies that the ranges of variances in FVFBCSP are highly concentrated. Except for subject 2, all variances are larger than 10^{-4} in FVFBCSP. In summary, under most cases, FVFBCSP removes the singularity successfully.

Training Accuracy Table 1 compares the average accuracy for cross validation between FBCSP and FVFBCSP. The results show that the accuracy for FVFBCSP is a lower than FBCSP. However, if we just replace CSP matrix with NCSP matrix in FBCSP, the accuracy will not be influenced significantly. One possible reason is that the standard deviation of the original signal reflects its energy. The higher the energy is, the more likely it will influence the final result. Therefore, the energy for each channel in each frequency band can be regarded as a kind of attention model (Denil et al. 2012). After removing this attention mechanism, the accuracy becomes lower.

Extracted Features Now we focus on the extracted features by FVFBCSP.

Table 1: Accuracy comparison between FBCSP, FVFBCSP and FBCSP with NCSP.

subject	FBCSP	FVFBCSP	FBCSP with NCSP
1	82.65±1.37	72.84±1.80	82.05±1.67
2	59.72±0.92	53.74±1.85	58.56±1.54
3	84.50±1.76	76.98±1.89	84.52±1.83
4	61.14±1.41	49.57±2.47	59.48±1.39
5	63.42±1.09	53.32±2.91	60.06±2.34
6	48.40±1.49	38.68±1.72	48.60±1.81
7	89.68±1.08	75.34±1.46	87.55±2.08
8	80.34±1.22	75.91±1.41	80.70±1.35
9	73.98±1.09	64.22±1.43	73.75±1.10

(1) Features in frequency domain and space domain. For each subject, we pick five features with the highest score for each class during every cross validation and take the average of all results as the three-dimension features of the particular class. To visualize the features in space domain and frequency domain, we add all absolute values of values in the same channel or frequency band together and regard the sum as its importance. Figure 3 and Figure 4 show the features in frequency domain and space domain respectively. The score for each feature denotes the average score calculated by the classifier.

To test whether the result is stable, we also draw the features when the classification task is binary classification. In our experiment, we split the whole data into two groups: left hand vs. right hand and foot vs. tongue.

The results show that in most cases (except for the subject 2), the result of quaternion classification is quite similar with that of binary classification, and the higher the average score is, the higher the salience (the maximum importance for each feature) is. One possible reason for the exception of subject 2 is that the variance of subject 2 does not concentrate enough, and could be very small, which introduces singularity.

The results in spatial domain show that for hand motor imagination, the most important component locates at both sides, for foot motor imagination, it locates near the center, and for tongue motor imagination, the result is similar to that of foot motor imagination to some extent. The frequency features show that α rhythm and β rhythm contribute more to motor imaginary. Sometimes frequency band

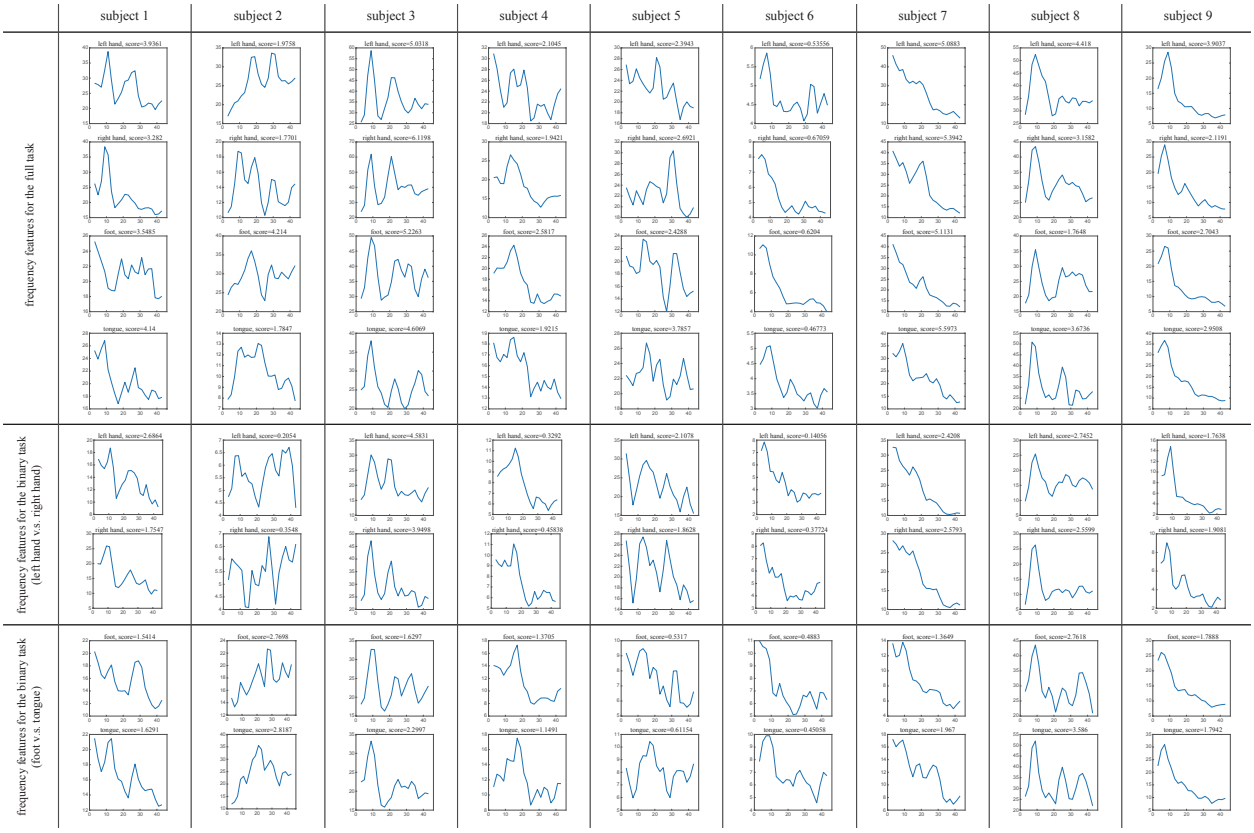

Figure 3: Features of all subjects in dataset I in frequency domain. X-axis denotes frequency in hertzes and y-axis denotes the importance of the corresponding frequency band.

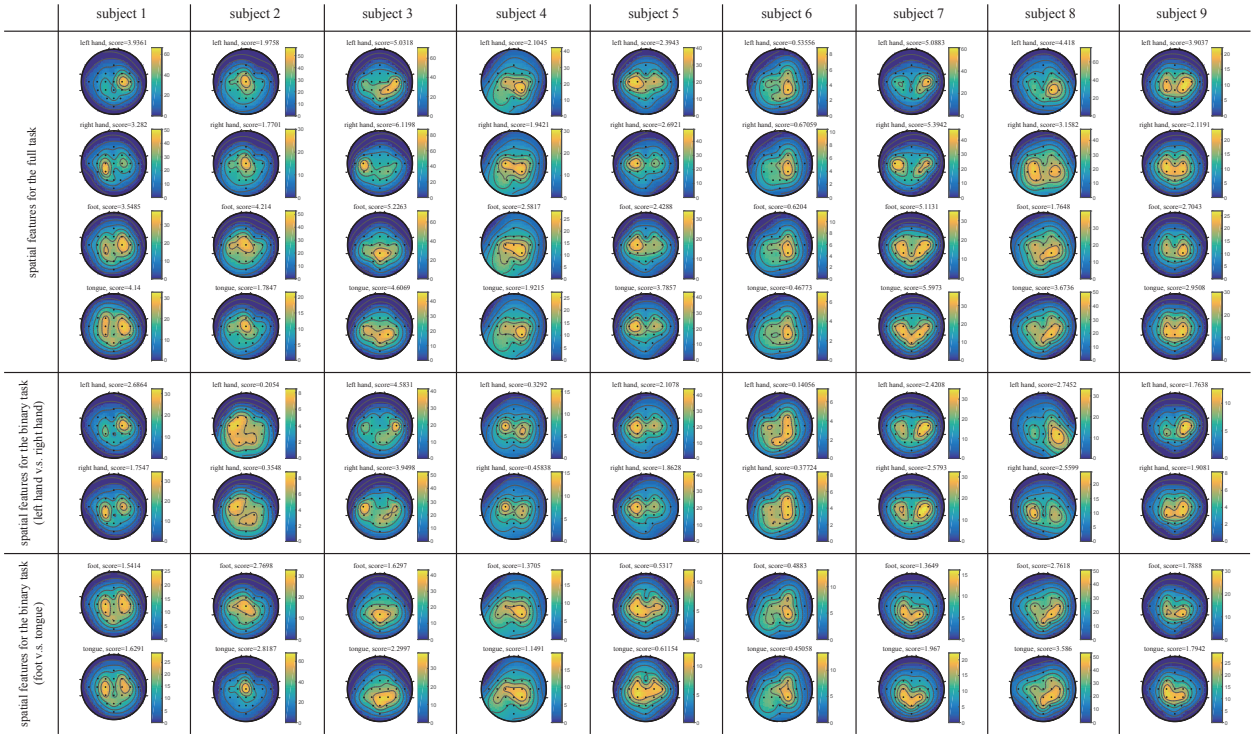

Figure 4: Features of all subjects in dataset I in spatial domain.

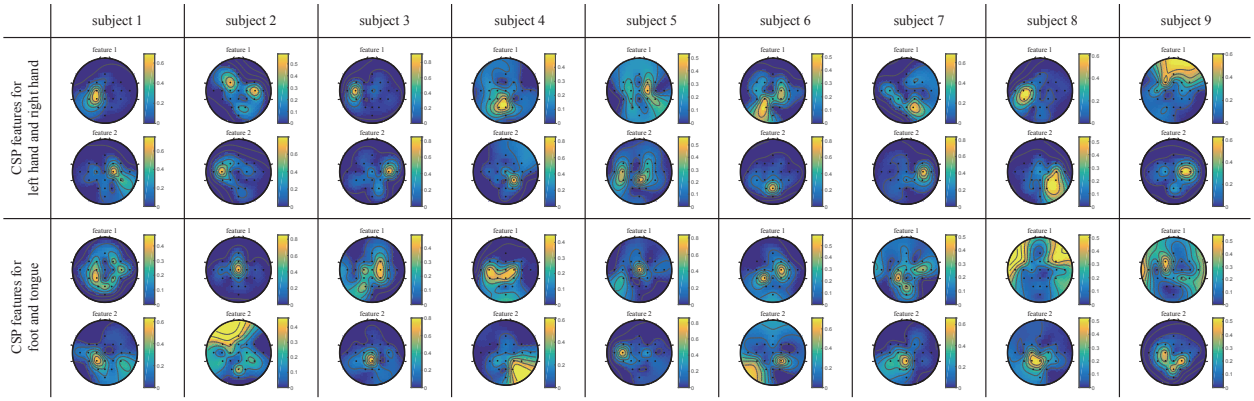

Figure 5: CSP features of all subjects in dataset I in spatial domain.

between 30Hz and 40Hz is also likely to make some contributions. Similar results can be found in other work (Liu et al. 2016).

For comparison, we also implement common spatial pattern for characterizing features (Müller-Gerking, Pfurtscheller, and Flyvbjerg 1999). Figure 5 shows the result. It should be highlighted that for CSP features, both of the two results indicate the most relevant part of both classes.

Firstly we compare the result of left hand and right hand. For subject 1, subject 3, subject 7 and subject 8, the results are similar. For subject 2, our method is unconvincing, which has been discussed above. For subject 4, subject 5, subject 6 and subject 9, the results of CSP are not satisfying. More specifically, the result of subject 5 is in significant dispersion, and for the result of subject 4, subject 6 and subject 9, the highlighted regions are not reasonable.

Then, we compare the results of foot and tongue. The results of CSP are implausible. For different subjects, the results of CSP are entirely different. Moreover, some results are not continuous. These features contradict with the physiological characteristics.

In summary, our method derives reasonable results for both frequency features and space features. For space features, our method generates much more convincing results than the traditional method. However, under some special cases like subject 2, our method does not perform well.

(2) Features in time-frequency domain. Figure 6 shows the frequency-time features in channel C3 of some trials for subject 1 and their variation of the energy of the signal in each frequency band, which is widely used for characterizing EEG features in frequency-time domain. Under most cases, our result accords with energy band. However, for some cases, there are some slight differences. For example, for the second case of right-hand class, the most significant parts in different frequency bands do not appear simultaneously. Concerning the features, our method successfully aligns them. Under another situation, say the third case of left-hand class, frequency band around 36Hz is highlighted while its energy is not large. The reason is that our method assigns a higher weight to this frequency band. Despite the low energy, it is still important for classification. On the other hand, although frequency band 4-6Hz in the fourth case

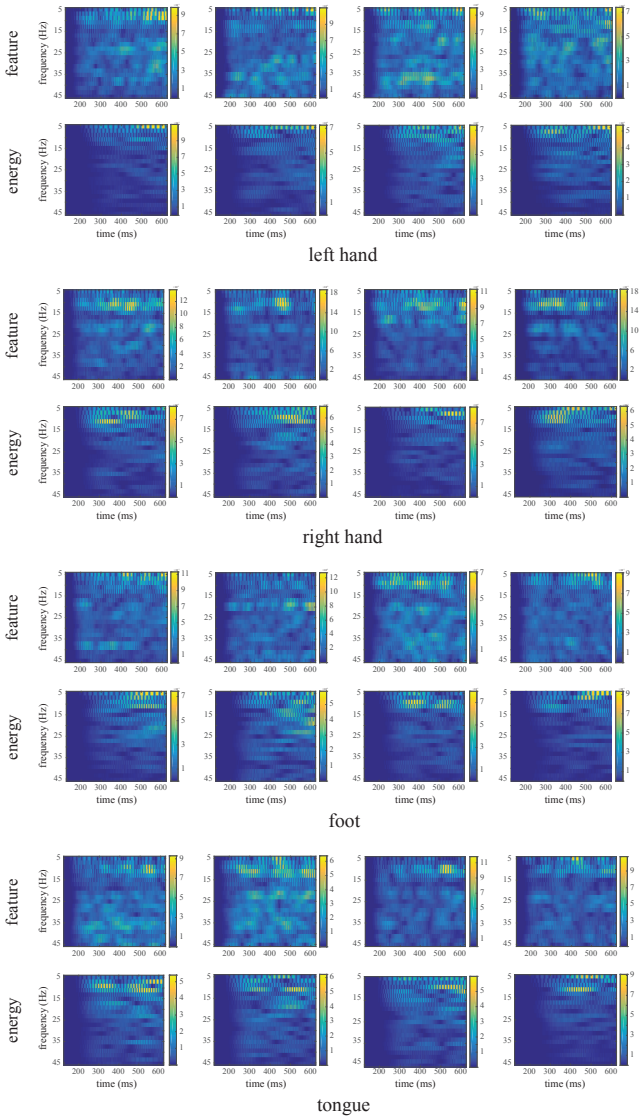

Figure 6: Features in frequency-time domain in channel C3 of some trials for subject 1.

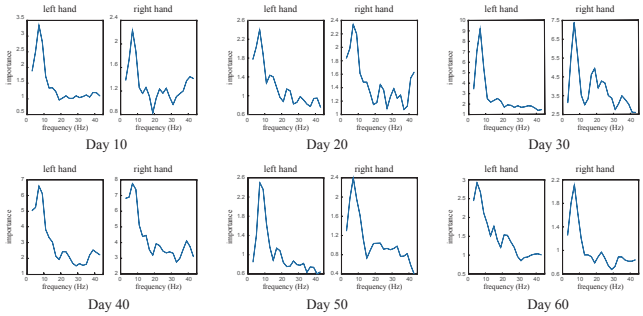

Figure 7: The evolution of frequency features for patient 1.

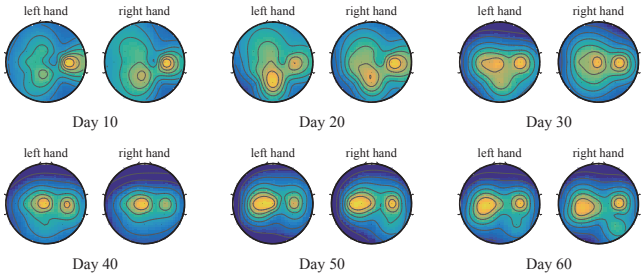

Figure 8: The evolution of spatial features for patient 1.

for the right-hand class is salient, its importance is quite small. However, for lack of prior knowledge, we can not tell which method generates the better result than the other.

Results of Dataset II

For this dataset, we focus on the evolution of the EEG features during the patients' rehabilitation training process. In this paper, we show the results of patient 2 who suffers right hemiplegia and has the most plentiful records.

Evolution of Features in the Frequency Domain Figure 7 shows the evolution of features in the frequency domain for patient 2 during the rehabilitation training process. For motor imagination of left hand, the features are stable: frequency band around 8Hz is the most significant part, and frequency band around 15Hz is also crucial for classification. For the right hand, which can not be moved for subject 2, the frequency band is getting concentrating towards low-frequency band around 8Hz during the rehabilitation training process. This phenomenon of concentrating of frequency specificity was also reported in (Liu et al. 2016).

Evolution of Features in the Space Domain Figure 8 shows the evolution of features in the spatial domain for patient 2 during the rehabilitation training process. For the right part of the brain, the critical region is stable, which is always around C4 channel. For the left part of the brain, the most significant area changes continuously. At first, the most important part is around CPz channel and Pz channel, which is not quite significant. This result shows that the left part suffers severe lesion, the active region for the right-hand movement moves to the middle of the brain. At day 20, this part begins to be much more salient. At day 30, the feature shows that the active region moves toward C3 channel. In the following days, this part keeps moving, which indicates the effect of the training for motor imagination. In the end, this region is around C3 channel and C1 channel. Similar results are reported by a magnetoencephalography (MEG) study (Feydy et al. 2002).

4 Discussion

This study presents a new characterizing feature method based on a variety of FBCSP, namely FVFBCSP. After calculating the gradient of the class score of a specific class with regard to the input, this method extracts features in three main domains (space, frequency and time). We further analyze why FVFBCSP is supposed to be applied rather than FBCSP.

Extensive experiments are conducted on the dataset containing data acquired from normal people. Experimental results show that our method yield relatively better performance compared with popular methods. In particular, compared with CSP methods, under most cases, our method identifies more reasonable spatial features. To explore the reasons, it could be the fact that our method utilizes more information. With a close look of FVFBCSP for binary classification, it firstly separates the data into different frequency band and then extracting CSP features for each band. During the characterizing feature process, the importance of each frequency band and timestamp would be calculated and finally contributes to the final result. With this information, our method achieves better performance than CSP. For features in the time-frequency domain, we compare our result with the energy distribution. They will generate similar results with slight differences. For lack of prior knowledge, we cannot tell which one performs better than the other.

Our experiments show that the concentration of the variance of the projected signal is the key to yield reliable results. If this variance is in dispersion like the situation of subject 2, the result is not stable, and will likely achieve bad performance.

A further experiment is conducted on another dataset related to unilaterally paralyzed stroke patients. In this experiment, we apply our method for capturing the evolution features during the rehabilitation training. Our results confirm the results of the previous study, and it also shows one potential application of our method.

5 Conclusion

In our work, we propose a novel classifier FVFBCSP for feature visualization. The experiments show that results are reliable on all three dimensions. There are still several problems need to be solved. For different feature visualization methods, how to compare these methods quantitatively? How to modify FBCSP so that the training accuracy is still held? Can we improve the result of features without adding more data?

References

Ang, K. K.; Chin, Z. Y.; Zhang, H.; and Guan, C. 2008. Filter bank common spatial pattern (fbcsp) in brain-computer

interface. In *International Joint Conference on Neural Networks*, 2390–2397. IEEE.

Brunner, C.; Leeb, R.; Müller-Putz, G.; Schlögl, A.; and Pfurtscheller, G. 2008. Bci competition 2008–graz data set a. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology* 136–142.

Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; and Wei, Y. 2017. Deformable convolutional networks. *CoRR* abs/1703.06211.

Denil, M.; Bazzani, L.; Larochelle, H.; and de Freitas, N. 2012. Learning where to attend with deep architectures for image tracking. *Neural computation* 24(8):2151–2184.

Feydy, A.; Carlier, R.; Roby-Brami, A.; Bussel, B.; Cazalis, F.; Pierot, L.; Burnod, Y.; and Maier, M. 2002. Longitudinal study of motor recovery after stroke. *Stroke* 33(6):1610–1617.

Gwin, J. T.; Gramann, K.; Makeig, S.; and Ferris, D. P. 2010. Removal of movement artifact from high-density eeg recorded during walking and running. *Journal of neurophysiology* 103(6):3526–3534.

Kandaswamy, A.; Krishnaveni, V.; Jayaraman, S.; Malmurugan, N.; and Ramadoss, K. 2005. Removal of ocular artifacts from eegła survey. *IETE journal of research* 51(2):121–130.

Kolda, T. G., and Bader, B. W. 2009. Tensor decompositions and applications. *SIAM review* 51(3):455–500.

Liu, Y.; Zhang, H.; Chen, M.; and Zhang, L. 2016. A boosting-based spatial-spectral model for stroke patients eeg analysis in rehabilitation training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24(1):169–179.

Lotte, F.; Congedo, M.; Lécuyer, A.; Lamarche, F.; and Arnaldi, B. 2007. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of neural engineering* 4(2):R1.

Müller-Gerking, J.; Pfurtscheller, G.; and Flyvbjerg, H. 1999. Designing optimal spatial filters for single-trial eeg classification in a movement task. *Clinical neurophysiology* 110(5):787–798.

Pfurtscheller, G., and Da Silva, F. L. 1999. Event-related eeg/meg synchronization and desynchronization: basic principles. *Clinical neurophysiology* 110(11):1842–1857.

Simonyan, K.; Vedaldi, A.; and Zisserman, A. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *CoRR* abs/1312.6034.

Zeiler, M. D., and Fergus, R. 2014. Visualizing and understanding convolutional networks. In *European conference on computer vision*, 818–833. Springer.

Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; and Torralba, A. 2016. Learning deep features for discriminative localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2921–2929.