Filter bank riemannian-based kernel support vector machine for motor imagery decoding

Yueqi Zhang[1], and *Jiaming* Chen[2,*]

[1]Fan Gongxiu Honors College, Beijing University of Technology, Beijing, China
[2]Faculty of Information Technology, Beijing University of Technology, Beijing, China

Abstract. Brain computer interface (BCI) enables the communication between the brain and external machines through Electroencephalography (EEG) signals, which has attracted lots of attention. Motor Imagery-based BCI (MI-BCI) is one of the most important paradigms in the BCI field. In MI-BCI, machine learning algorithms can be employed for identifying the target limb of motor intention effectively. As a typical machine learning algorithm for motor imagery decoding, the Riemannian-based kernel support vector machine (RK-SVM) algorithm is not capable of feature extraction from multiple frequency bands, which limits its performance. To solve this problem, the Filter Bank Riemannian-based Kernel Support Vector Machine (FBRK-SVM) method that combines the filter bank structure and Riemannian-based kernel was proposed. In comparative experiments on two commonly used public datasets, it is found that the proposed algorithm can yield higher decoding performance, which provides a new option for the classification of motor imagery.

Keywords: Brain computer interface, Motor imagery, Riemannian geometry, Machine learning.

1 Introduction

In recent years, brain computer interface (BCI) has attracted much attention from researchers. Brain computer interface is a system that realizes the communication between the brain and machines through biological signals such as Electroencephalography (EEG) instead of the common way of information output such as peripheral nerves and muscles [1]. Motor imagery-based BCI (MI-BCI) is a representative BCI paradigm that can be adjusted by users and reflect autonomous motor intention without external stimulation [2]. MI-BCI is one of the most important and widely studied paradigms.

The classification of Motor Imagery-based Electroencephalography (MI-EEG) signals is one of the key problems. MI-EEG signals induced by motor intention of specific limbs are used as the input in MI-BCI. A classification model is built based on features of MI-EEG to identify the label of MI-EEG trials, so as to identify the target limb of the subject. MI-EEG signals have five rhythms according to frequency bands: δ (0.5-4Hz), θ (4-7Hz), α (8-13Hz),

* Corresponding author: billchen@emails.bjut.edu.cn

β (14-30Hz), and γ (30-100Hz), and Event-related desynchronization (ERD) and event-related synchronization (ERS) occurred in α and β bands are considered to be induced by motor imagery [3]. Machine learning algorithms like Common Spatial Patterns (CSP) [4], Filter Bank Common Spatial Patterns (FBCSP) [5], and Riemannian geometry-based methods [6] are commonly used in MI-BCI for feature extraction based on MI-EEG [7]. In Riemannian geometry-based methods, the Riemannian-based Kernel Support Vector Machine algorithm [8] only extracts features from a single frequency band instead of features from fine-divided frequency bands, which limits the decoding performance. To solve this problem, the Filter Bank Riemannian-based kernel (FBRK) method was proposed to extract features from multiple frequency bands, and an MI decoding model based on FBRK and the support vector machine algorithm [9] was built to achieve higher classification performance.

The structure of this paper is as follows. Section 2 describes the Filter Bank Riemannian-based kernel Support Vector Machine method proposed in this paper. In Section 3, the performance of the three existing methods and the proposed method are compared through experiments. In Section 4, the conclusion of this paper is drawn.

2 Methods

The structure of the proposed method is shown in figure 1. MI-EEG signals are filtered by 9 band-pass filters (with frequency bands of 4-8Hz, 8-12Hz, 12-16Hz, 16-20Hz, 20-24Hz, 24-28Hz, 28-32Hz, 32-36Hz and 36-40Hz respectively) of a filter bank structure to obtain 9 groups of signals. This design enables feature extraction from different fine-split frequency bands, in order to extract the spatial information effectively from θ, α, β, and low γ bands. Then, the covariance matrices of filtered signals are estimated by the Oracle approving shrinkage (OAS) method, and then mapped to the tangent space by the Riemannian-based kernel (RK) to yield tangent space features (TS features). After concatenated, TS features are classified by the support vector machine algorithm to identify the target limbs.

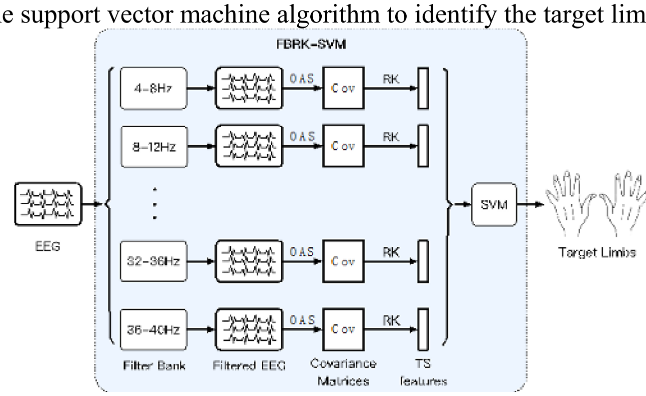

Fig. 1. The structure of FBRK-SVM method.

2.1 Feature extraction based on Riemannian-based kernel

Trials and labels of MI-EEG are expressed as a $N \times E \times T$ matrix $\boldsymbol{X}$ and a vector $\boldsymbol{y}$ with length N respectively, where N is the number of trials, E is the number of channels, and T is the number of sampling points. The data of trial i is recorded as X_i whose label is y_i, where $i \in \{1,2,\ldots,N\}$. Given a set of MI-EEG trials and labels, a model f is trained to predict the category $\hat{y}_i$ of an unseen trial. The covariance matrix of each trial is obtained before defining Riemannian-based kernel. The sample covariance matrix $\boldsymbol{\Sigma}_i$ of trial $\boldsymbol{X}_i$ ($\boldsymbol{X}_i \in \mathbb{R}^{E\times T}$) is computed, as shown in equation (1),

$$\boldsymbol{\Sigma}_i = \frac{1}{N-1}\boldsymbol{X}_i\boldsymbol{X}_i^T \tag{1}$$

Where $\boldsymbol{X}_i^T$ represents transpose of matrix $\boldsymbol{X}_i$. To obtain more stable spatial information representation, we employed the OAS method [10] to estimate the covariance $\boldsymbol{C}_i$, as shown in equation (2),

$$\boldsymbol{C}_i = (1-\hat{\rho})\boldsymbol{\Sigma}_i + \hat{\rho}\widehat{\boldsymbol{F}}_i$$

$$\widehat{\boldsymbol{F}}_i = \frac{Tr(\boldsymbol{\Sigma}_i)}{E}\boldsymbol{I}$$

$$\hat{\rho} = min\left(\frac{\left(\frac{1-2}{E}\right)Tr(\boldsymbol{\Sigma}_i{}^2)+Tr^2(\boldsymbol{\Sigma}_i)}{\left(\frac{T+1-2}{E}\right)\left[Tr(\boldsymbol{\Sigma}_i{}^2)-\frac{Tr^2(\boldsymbol{\Sigma}_i)}{E}\right]}, 1\right) \tag{2}$$

Where $\hat{\rho}$ is the coefficient, $\hat{F}_i$ is the transformed sample covariance, and Tr is the trace operation. Since the covariance matrix is a symmetric positive definition (SPD) matrix, the space of covariance matrices is an matrix space, which forms an Riemannian manifold of dimension $E(E+1)/2$. To employ a vector-based classification algorithm, the symmetric positive definite covariance matrices are transformed by the modified half-vectorization [8] by taking the lower triangular matrix of symmetric positive definite matrix $\boldsymbol{C}_i$ (denoted as $vect(\boldsymbol{C}_i)$) to obtain an column vector with length $E(E+1)/2$.

The Riemannian-based kernel allows computing the scalar product on the Riemannian manifold by the scalar product on the tangent space. At point $\boldsymbol{C}$ on the Riemannian manifold, a tangent space $\mathcal{T}_C\mathcal{M}$ can be constructed. With the projection of the Riemannian manifold to the tangent space, the tangent vector $\boldsymbol{p}_i$ on the tangent space (high-dimensional Euclidean space) is used to approximate $\boldsymbol{C}_i$ on the Riemannian manifold, as shown in figure 2.

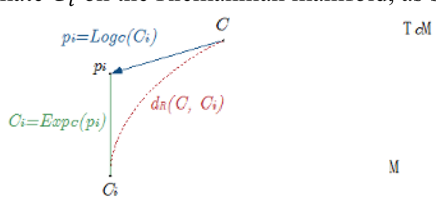

Fig. 2. Riemannian manifold and the tangent space.

In Figure 2, $d_R(C, C_i)$ indicates Riemannian distance [11] from point $\boldsymbol{C}$ to point $\boldsymbol{C}_i$. The projection from points at Riemannian manifold to the tangent space can be computed by the mapping function ϕ, as defined in equation (3),

$$\boldsymbol{p}_i = \phi(\boldsymbol{C}_i) = Log_{\boldsymbol{C}}(\boldsymbol{C}_i) = \boldsymbol{C}^{1/2}logm(\boldsymbol{C}^{-1/2}\,\boldsymbol{C}_i\boldsymbol{C}^{-1/2})\boldsymbol{C}^{1/2} \tag{3}$$

Where $\boldsymbol{C}_i$ is a symmetric positive definite matrix on a Riemannian manifold, $\boldsymbol{p}_i$ is the tangent vector on tangent space $\mathcal{T}_C\mathcal{M}$, and $logm(\cdot)$ denotes the logarithm of a matrix. Also, the mapping function from high-dimensional tangent space to Riemannian manifold is defined in equation (4):

$$\boldsymbol{C}_i = Exp_{\boldsymbol{C}}(\boldsymbol{C}_i) = \boldsymbol{C}^{1/2}expm(\boldsymbol{C}^{-1/2}\,\boldsymbol{p}_i\boldsymbol{C}^{-1/2})\boldsymbol{C}^{1/2} \tag{4}$$

Where $expm(\cdot)$ denotes the exponential of a matrix. The scalar product of two tangent vectors $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ on the tangent space at point $\boldsymbol{C}$ are defined as:

$$<\boldsymbol{p}_1, \boldsymbol{p}_2>_{\boldsymbol{C}} = Tr(\boldsymbol{p}_1\boldsymbol{C}^{-1}\boldsymbol{p}_2\boldsymbol{C}^{-1}) \tag{5}$$

If the point to form the tangent space is denoted as $\boldsymbol{C}_t$ (called the reference point [8]), the mapping function can also be written as $\phi(\boldsymbol{C}) = Log_{\boldsymbol{C}_t}(\boldsymbol{C})$.

The Riemannian-based kernel (RK) K_R can be obtained by projecting the scalar product of covariance matrices to scalar product of tangent vectors on the tangent space, and K_R can be denoted as the product of half-vectorized tangent vectors according to the definition of the trace operator, as shown in equation (6).

$$K_R\big(vect(\boldsymbol{C}_i), vect(\boldsymbol{C}_j); \boldsymbol{C}_t\big) =< \phi(\boldsymbol{C}_i), \phi(\boldsymbol{C}_j) >_{\boldsymbol{C}_t}= vect(\boldsymbol{p}_i)^T vect(\boldsymbol{p}_j) \tag{6}$$

where $\boldsymbol{p}_i$ is the tangent vector on the tangent space at point $\boldsymbol{C}_t$. $\boldsymbol{p}_i$ is mapped from point $\boldsymbol{C}_i$, as shown in equation (7).

$$\boldsymbol{p}_i = \boldsymbol{C}_t{}^{-1/2} Log_{\boldsymbol{C}_t}(\boldsymbol{C}_i) \boldsymbol{C}_t{}^{-1/2} = logm(\boldsymbol{C}_t{}^{-1/2} \boldsymbol{C}_i \boldsymbol{C}_t{}^{-1/2}) \tag{7}$$

Since a valid scalar product is employed in the Riemannian-based kernel, the Riemannian-based kernel is a positive definite kernel function and respects the Mercer condition [8]. Methods based on the Riemannian-based kernel usually map the covariance matrices to the tangent space to obtain the tangent space features for effective feature extraction in the spatial domain.

2.2 Filter bank structure combined with the Riemannian-based kernel

In the filter bank structure, the filtered signals of each frequency band are transformed into tangent space features through the mapping function defined by the Riemannian-based kernel. Specifically, the covariance matrix $\boldsymbol{C}_i^b$ of trial $\boldsymbol{X}_i^b$ filtered by band $b(b \in \{1, \cdots, B\})$ are computed, mapped to the tangent space, and then half-vectorized to yield the tangent space feature TF_i^b of frequency band b, as shown in equation (8),

$$TF_i^b = \mathrm{vect}(\boldsymbol{p}_i^b) = vect\left(\mathrm{logm}\left(\boldsymbol{C}_t^{b\,-1/2} \boldsymbol{C}_i^b \boldsymbol{C}_t^{b\,-1/2}\right)\right) \tag{8}$$

where $\boldsymbol{C}_t^b$ is the reference point for covariance matrices of the trials filtered by band b, and $\boldsymbol{p}_i^b$ is the corresponding tangent vector. Finally, the tangent space feature TF_i can be obtained by concatenating B groups of tangent space features, as shown in equation (9).

$$TF_i = \begin{bmatrix} TF_i^1 \\ \vdots \\ TF_i^B \end{bmatrix} \tag{9}$$

With the feature extraction method based on Filter Bank Riemannian-based kernel, the input sample $(\boldsymbol{X}, \boldsymbol{y})$ can be transformed into $(\boldsymbol{TF}, \boldsymbol{y})$, where $\boldsymbol{TF}$ ($\boldsymbol{TF} \in \mathbb{R}^{N \times B \times E(E+1)/2}$) represents the tangent space feature extracted from MI-EEG data. Finally, combining $(\boldsymbol{TF}, \boldsymbol{y})$ with the support vector machine algorithm, the Filter Bank Riemannian-based kernel Support Vector Machine (FBRK-SVM) algorithm is employed for classification.

3 Experiment

3.1 Experiment setup

BCI Competition IV Dataset IIa [12] (denoted as BCIC IV IIa below) and BCI Competition IV Dataset IIb [13] (denoted as BCIC IV IIb below) were used for validation. BCIC IV IIa is a public dataset with four MI tasks including left hand, right hand, feet, and tongue of 9 subjects. MI-EEG signals were collected from 22 electrodes at a sampling rate of 250Hz, including 576 trials from two days (with one session in each day). Each session contains four types of trials, and each class contains 72 trials. The data from 0.5 to 2.5 seconds after the cue were extracted as a trial. BCIC IV IIb contains two MI tasks including left and right hands of 9 subjects. MI-EEG signals were collected from three electrodes, including five sessions. The first three sessions were used for training, including 400 trials (200 for the left hand and 200 for the right hand), and the last two sessions were used for testing, including 320 trials (160 for the left hand and 160 for the right hand). The data from 0 to 4 seconds after the cue were extracted as a trial. All trials were labelled (i.e., the target limb of each trial was marked).

Since features of MI-EEG between subjects are different, the classification accuracy of each subject was computed separately, and the mean accuracy of multiple subjects were used as the performance measurement of a model. All trials were pre-processed by the 1st order Butterworth band-pass filter. The frequency band of CSP and RK-SVM is 7-35Hz, and the frequency bands of methods with a filter bank structure are 4-8Hz, 8-12Hz, 12-16Hz, 16-20Hz, 20-24Hz, 24-28Hz, 28-32Hz, 32-36Hz, and 36-40Hz.

3.2 Results

The performance of CSP-SVM, FBCSP-SVM, RK-SVM, and the proposed FBRK-SVM method were evaluated on the test set of BCIC IV IIa and BCIC IV IIb datasets. The results are shown in table 1, where the highest accuracy and kappa value are shown in bold. In both datasets, the mean accuracy and kappa value of FBRK-SVM are the highest among compared methods, which indicates that the proposed FBRK-SVM method can yield higher performance than other methods.

Table 1. Performance of compared methods on the test set of MI dataset.

Dataset	Methods	Mean Accuracy	Mean Kappa Value
BCIC IV IIa	CSP-SVM	0.6380	0.5173
	FBCSP-SVM	0.6524	0.5366
	RK-SVM	0.6259	0.5012
	FBRK-SVM	**0.7222**	**0.6296**
BCIC IV IIb	CSP-SVM	0.7183	0.6244
	FBCSP-SVM	0.7598	0.6797
	RK-SVM	0.7251	0.6335
	FBRK-SVM	**0.7621**	**0.6828**

4 Conclusion

Combined with the filter bank structure and Riemannian-based kernel support vector machine algorithm, the FBRK-SVM method was proposed. Multiple groups of band-pass filters were employed to filter MI-EEG signals, and the OAS method was used for estimating the covariance matrix for higher robustness. Nine sets of tangent space features were extracted by Riemannian-based kernel and finally classified by the SVM algorithm. The performance of proposed FBRK-SVM algorithm on two MI datasets was compared with that of typical algorithms including CSP-SVM, FBCSP-SVM, and RK-SVM. It is found that the proposed algorithm has higher classification accuracy, up to 72.22%, and 76.21% on the two datasets respectively. Future work includes: (1) further improving the classification accuracy with the ensemble learning algorithm; (2) exploring the performance of the proposed method in inter-subject tasks.

References

1. Wolpaw J R, Birbaumer N, Heetderks W J, *et al* 2000 Brain-computer interface technology: a review of the first international meeting *IEEE Trans. Rehabil. Eng.* **8** 164-173
2. Yi W, Qiu S, Wang K, *et al* 2017 Enhancing performance of a motor imagery based brain–computer interface by incorp orating electrical stimulation-induced SSSEP *J. Neural Eng.* **14** 02600
3. Pfurtscheller G, Da Silva F H L 1999 Event-related EEG/MEG synchronization and desynchronization: basic principles *Clin. Neurophysiol.* **110** 1842-1857
4. Ramoser H, Muller-Gerking J, Pfurtscheller G 2000 Optimal spatial filtering of single trial EEG during imagined hand movement *IEEE Trans. Rehabil. Eng.* **8** 441-446
5. Ang K K, Chin Z Y, Zhang H and Guan C 2008 Filter bank common spatial pattern (FBCSP) in brain–computer interface *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) IEEE Int. Joint Conf. Neural Networks* (IEEE) pp 2390–2397
6. Yger F, Berar M, Lotte F 2016 Riemannian approaches in brain-computer interfaces: a review *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 1753-1762
7. Lotte F, Bougrain L, Cichocki A, *et al* 2018 A review of classification algorithms for EEG-based brain–computer interfaces: a 10-year update *J. Neural Eng.* **15** 031005
8. Barachant A, Bonnet S, Congedo M, *et al* 2013 Classification of covariance matrices using a Riemannian-based kernel for BCI applications *Neurocomputing* **112** 172-178
9. Cortes C, Vapnik V 1995 Support-vector networks *Mach. Learn.* **20** 273-297
10. Chen Y, Wiesel A, Eldar Y C, *et al* 2010 Shrinkage algorithms for MMSE covariance estimation *IEEE Trans. Signal Process.* **58** 5016-5029.
11. Moakher M 2005 A differential geometric approach to the geometric mean of symmetric positive-definite matrices *SIAM Journal on Matrix Analysis and Applications* **26** 735-747
12. K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang 2012 Filter bank common spatial pattern algorithm on BCI competition iv datasets 2a and 2b *Front. Neurosci.* **6** 39
13. Tangermann M, Müller K R, Aertsen A, *et al* 2012 Review of the BCI competition IV *Front. Neurosci.* **6** 00055