Classification of EEG Signal by STFT-CNN Framework: Identification of Right-/left-hand Motor Imagination in BCI Systems

Yao Lu[12]

Brain Cognitive Computing Lab, School of Information Engineering, Minzu University of China
Beijing, 100081, China
E-mail: `5206661314@163.com`

Huiping Jiang

Brain Cognitive Computing Lab, School of Information Engineering, Minzu University of China
Beijing, 100081, China
E-mail: `jianghp@muc.edu.cn`

Wenqiang Liu

Brain Cognitive Computing Lab, School of Information Engineering, Minzu University of China
Beijing, 100081, China
E-mail: `liuwqleo@163.com`

This paper described the relationship between EEG signals and MI in BCI system. EEG signals are used to classify the direction of motioninto two kinds: left and right. We extracted features from original EEG data using STFT and put them into CNN. The result showed that the framework of STFT-CNN had higher average test accuracy. Furthermore, the generations of motor imagery were analyzed, and the result showed that better classification results will appear in the middle stage with its classification accuracy reaching 92.8%.

1Speaker

1.Introduction

The advanced computer technique makes it possible for people to communicate with computers or electronic devices by EEG signals, which are also called BCI systems. In general, BCI systems are composed of the following stages: brain signal acquisition, preprocessing, feature extraction, classification and external device communication. The key factor is the method used to process and extract meaningful information from the brain signal.

Motor imagery (MI) is a strategy with which BCI users can encode their intentions in the EFG. MI is linked to the sensory-motor cortex area of human brain which can be record by central, parietal and temporal electrodes of the international 10–20 electrode position system. The disabled patients can get benefit from a MI-based BCI systems since it could improve the quality of their life. In order to decode MI according to the EEG signals, the extraction of the most discriminative features is performed, followed by classification of such features into different directions.

For feature extraction, methods such as Fourier transform, wavelet transforms (WT) and common spatial patterns (CSP) are used.Particularly, time–frequency information is often extracted with the aid of WT due to its ability to achieve high resolutions in the time domain while also preserveresolution in the frequency domain of acquired EEG signals[1-4].For classification, different approaches have been used, including linear discriminant analysis (LDA)[5], support vector machines (SVM)[6],K-Nearest-Neighbor method (KNN)[7]and Deep Learning (DL)[8].

However, the performance of a new type of deep learning, convolutional neural network (CNN)[9] model, has not been widely tested for EEG classification in MI field.

For MI, while many previous works about deep neural network focused on learning discriminative features from input, they did not take disentangle factors of variation into account. On the one hand, Convolution Neural Networks (CNNs) are shown to have the capability to infer a hierarchical feature representation that facilitates categorization. On the other hand, existed works show that classical FFT does a good job in EEG feature extraction, and STFT overcomes the shortcoming of the FFT's local analysis capability and plays an important role in EEG signal processing. Currently, Yazhou Wu has used wavelet analysis and Back-propagation Neural Network method to process and analyze the experimental data of off-line.The average results of recognition rate are 72%.

From all above, we propose learning aboutdirection features for MI by combining short time Fourier transform (STFT) and convolutional neural net-work (STFT-CNN), in which simple features are learned with the use of STFT, and affect-salient, discriminative features and classification are handled with CNN layers.

This paper is organized as follows: The STFT and CNN are described in Section 2. The dataset and experiments are presented in Section 3. Finally, the results and conclusion are shown at length in Section 4 and 5.

2.Methods

2.1Short-Time Fourier Transform (STFT)

Due to EEG signals' instability, it is difficult to extract the key features from the noise. The

common used method is Short-Time Fourier Transform (STFT), which translates the time domain signal into the time-frequency domain for EEG signals. A detailed description is as follows.

Given a determined signal x(t), the power spectrum information can be obtained at each time point by the following formula (2.1).

$$\mathrm{STFT}_x(t,\omega) = \int x(\tau) g_{t,\omega}{}^*(\tau) d\tau$$

$$\int x(\tau) g^*(\tau - t) e^{\ j\omega\tau} d\tau = \left\langle x(\tau), g(\tau - t) e^{j\omega\tau} \right\rangle$$

$$g_{t,\omega}(\tau) = g(\tau - t) e^{j\omega\tau}, \left\| g(\tau) \right\| = 1, \left\| g_{t,\omega}(\tau) \right\| = 1$$

(2.1)

In which, $g(\tau)$ is the window function, and which respectively intercepts $x(\tau)$ in the time domain. The Fourier transformation is performed on the intercepted local signal. Keep moving $g(\tau)$ to intercept different signal $x(\tau)$, we can get the power spectrum characteristics of different periods.

A key factor ofSTFT is how to select the reasonable window functionto achievean appropriate compromise between time and frequency resolution. The Joint Time-Frequency resolution of STFT is determined by the selection of the Window function. There exists many window functions.For example, Hamming or Rectangular window are the most commonly used. The time domain form of the rectangular window function can be expressed as Formula (2.2).

$$\omega(n) = R_N(n) = \begin{cases} 1, 0 \le n \le N-1 \\ 0, \quad \text{others} \end{cases}$$

(2.2)

Its frequency domain form can be expressed as Formula (2.3).

$$W_R(e^{jw}) = e^{-j(\frac{N-1}{2})\omega} \frac{\sin(\frac{\omega n}{2})}{\sin(\frac{\omega}{2})}$$

(2.3)

The time domain form of the Hamming window can be expressed as Formula (2.4).

$$W(k) = 0.54 - 0.46\cos(2\pi \frac{k}{N-1})$$

$k=1,2,...,N$(2.4)

Its frequency domain form can be expressed as Formula (2.5):

$$W(\omega) = 0.54\, W_R(\omega) + 0.23[W_R(\omega - \frac{2\pi}{N-1}) + W_R(\omega + \frac{2\pi}{N-1})]$$

(2.5)

In which $W_R(\omega)$ is the amplitude-frequency characteristic of the rectangular window function.

2.2Convolution Neural Networks (CNN)

A classifier based on a CNN seems to be a good approach for EEG classification, since the signal to detect contains a lot of variations over time and person. CNN are hierarchical models whose convolution layers alternate with sub-sampling layers, reminiscent of simple and complex cells in the primary visual cortex. The network architecture consists of three basic building blocks to be stacked and composed as needed. CNN is one of the most successful models to supervise image classification and set the state-of-the-art in many benchmarks.

In the convolution layer, the feature map of the upper layer is convoluted by a learning convolution kernel and then output feature map is obtained by an activation function. Each output feature map can be combined to convolve and the values of multiple feature maps are described as Formula (2.6):

$$x_j^1 = f(u_j^l)$$
$$u_j^1 = \sum_{i \in M_j} x_i^{l-1} * k_{ij}^1 + b_j^1$$

(2.6)

Where u_j^l is the j-th channel's activation value of the convolution layer l, which is obtained by convolving the output feature map x_i^{l-1} of the upper layer and bias it, x_j^l is the j-th channel's output value of the convolution layer l. F(.) is the activation function, usually can be used sigmoid() and tanh() functions. M_j is a subset of the input feature map used to compute M_j^l .K is a convolution kernel matrix, and b_j^l is a bias of the convolved feature map, "*" is the convolution operator.

The Down-sample Layer and every input feature map and output a feature map are described by the following Formula (2.7):

$$x_j^1 = f(u_j^l)$$
$$u_j^1 = \beta_j^1 \text{down}(x_j^{l-1}) + b_j^l$$

(2.7)

Where u_j^l is the j-th channel's activation value of the down-sample layer l, which is obtained by down-sample weighing the upper layer's output feature map and then biasing it, and B is the weight of the down-sample layer, b_j^l is the bias of down-sample layer. Down (.) is a down-sample function that uses the sliding window to divide the input feature map x_j^{l-1} into n×n non-overlapping blocks, and then find the mean or maximum value of each block, ultimately reducing the data dimension.

In the fully-connected layer, all 2D matrices are concatenated into a one-dimensional feature as the fully-connected layer's input. The output of fully-connected Layer l can be obtained by weighing the input and responding to the activation Formula (2.8):

$$x_j^1 = f(u^l)$$
$$u^1 = w^1 x^{l-1} + b^1$$

(2.8)

Where u^l is the activation value of the fully-connected layer l, which is obtained by

weighing the upper layer's output feature map x^{l-1} and biasing it, and w^l is the weight of the fully-connected layer, b^l is the bias of fully-connected Layer l.

3.Experiments

3.1Stimulus File Setup & Datasets

EEG data were record from 3 women and 3 men aged between 22 and 36, who were healthy and right-handed. All of the subjects were undergraduate students from Minzu University and were informed of the purpose of this experiment.

This experiment,was mainly concerned with Left-Right. The protocol designed for our experiment is shown in Fig. 1. Stimuli file contain 2 sections: Section I of 5 tasks for practicing; Section II of 100 pictures for Left-Right task. Each trail is shown for 3s. Before the task, a red cross in the center of screen appears to attract subject's attention. After the task, a black screen of 2s is shown to clear the impact of the last task. Left-Right tasks are shown randomly to subjects in each section. Between sections, subjects can choose to take a break or continue on him/her own.

EEG signals are acquired by using the Neuroscan 64 system. Subjects were fitted with a 64-channel electrode cap during the experiment. The Ag/AgCl electrodes are placed according to the international 10-20 system. The contact impedance between electrodes and skin was kept below10kΩ. The EEG data were recordat a sampling rate of 1000Hz. Experiment went on in a quiet and comfortable room, without being disturbed by people talking or phone rings.
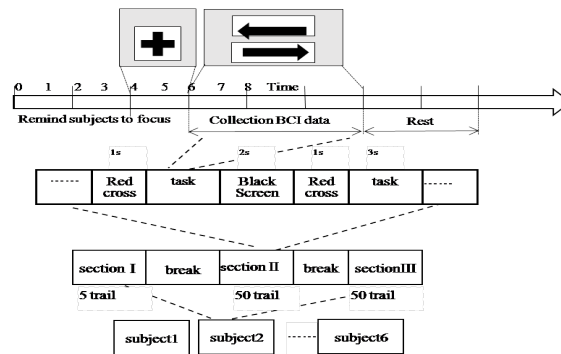

Figure 1: Acquisition Protocol for EEG Recording

3.2EEG Signal Pretreatment

During the experiment, the collection of the EEG signals was very weak and susceptible to interference from other signals. Related studies have shown that the interference sources from the neck facial muscle movement and eye movement and these noise will cause signal changes. Therefore, it is necessary to preprocess the EEG signal to remove artifacts contained in the original signal.

In this experiment, subjects were asked to remain motionless as possible as they canto minimize the number of blinks and they were told to turn off mobile phones and other communications equipment to decrease signal interference during the experiment. Duringthe data preprocessing stage, we use the manual method to ensure that most of the interference signals are removed; filtering use low-pass filter and the signal frequency is limited to 50Hz.

Because the multi-band characteristic frequency of the post-extraction is under 50Hz, it will not affect the motor image classification result.

3.3STFT+CNN

The EEG data power spectrum is usually divided into five frequency bands: Delta band(δ:1-4Hz), Theta band(θ:4-8Hz), Alpha band (α:8-13Hz), Beta band (β:13-30Hz), and Gamma band(γ:36-44Hz). According to the stimulus file, each picture of the switch will produce an event, so we take the incident as a marker interception of EEG signal, and then each segment of the EEG signal short-time Fourier transforms to obtain EEG time-frequency distribution, Spectrum of EEG signal.

We selected the Hamming window with the length of 3000, and averaged the frequencies in these five frequency bands. In this paper, we used the 64-electrode cap to get the sample of the 320-dimensional vector.

After feature extraction of EEG data by short-time Fourier was transformed, we put these data into convolution neural networks to train the convolution layer, the max-pooling layer and the classification layer. Convolution neural networks structure is shown in Fig. 2.

The first convolution layer filters 64×5 input matrix with 16 kernels of size 3×3, the first max-pooling layer takes as input the output of the first convolution layer and filters it with the size of 2×1, and then outputs a matrix of size 32×5; The second convolution layer filters 32×5 input matrix with 25 kernels of size 3×3, the second max-pooling layer takes as input the output of the second convolution layer and filters it with the size of 2×1, and then outputs a matrix of size 16×5; The third convolution layer filters 16×5 input matrix with 400 kernels of size 3×3, the third max-pooling layer takes as input the output of the third convolution layer and filters it with the size of 2×1, and then outputs a matrix of size 8×5; The neurons in the fully-connected layer are connected to all neurons in the previous layer, and the fully-connected layer has 8×5×400 neurons.

As a deep neural network, CNN can use the classical training method: from low to high layer, convolution kernel of each layer is given the initial value randomly, and then adjustedto get a better value.
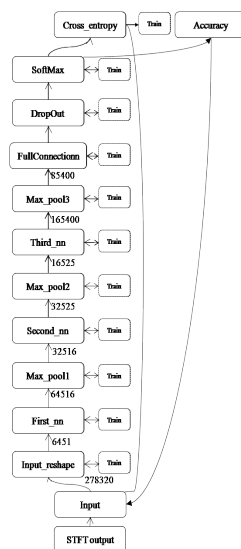

Figure 2: Framework of STFT-CNN

4.Simulation & Results

We select 3/4 samples as training set and 1/4 samples as testing set. After 5000 training, the accuracy of the test set classification reached 89%. The accuracy of the test set classification is obtained by the visualization tool as is shown in Figure 3.
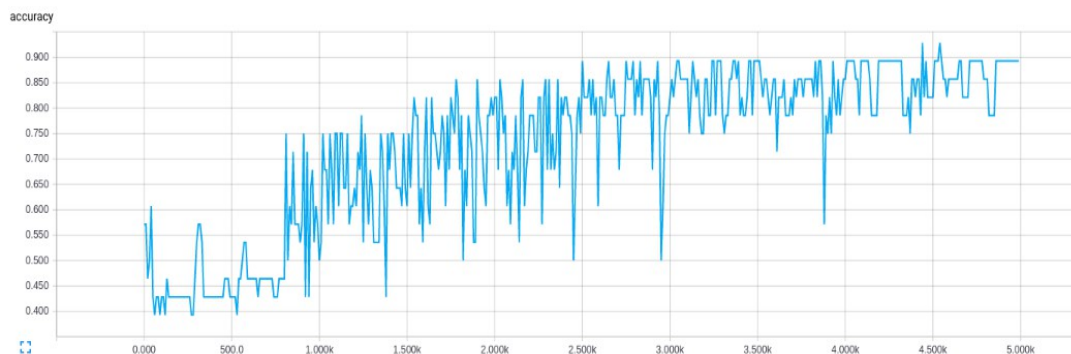

Figure 3: Accuracy of the Test Set Classification

In the field of MI, the classification speed and classification accuracy are equally important. According to our experiments, data were collected from 0 second and last three seconds which contains not only the movement of imagination information (left and right) components but also picture stimulation, brain electrical noise and other signals. In order to extract the high-quality signal in a short time and reduce the data processing time. The data collected in the experiment are analyzed in 0-1s, 1-2s and 2-3s, and the same method and parameters are used for the analysis. Results are summarized in Table1.

Time period	Recognition rate
0-1s	57%
1-2s	92.8%
2-3s	78%

Table1: Influence of the Time Period on the Highest Classification Accuracy

As seen from the table, the recognition rate increases with the time period of the selected data and the recognition rate of the sample increasing from 57% of the time period 0-1s to 92.8% of the time period of 1-2s. But the recognition rate does not increase over time. During the period of 0-1s, when the movement of imagination components in the EEG signal is not the main component, the recognition rate is low; in 1-2s period, the recognition rate has reached 92.8%, which indicates that motor imagery has begun to become the main component; from the 2-3s time period, we can see that the movement of imagination in the sample components has begun to subside.

5.Conclusion

In this paper, after the energy spectrum feature of signal is extracted by STFT which is classified by CNN, the classification accuracy is 89%, which is calculated by intercepting different time periods. From the results we can see that the best classification is in the period of 1-2s, and the accuracy rate is 92.8%. It indicates that valid motor imagery thinking is gradually

stimulated, and will gradually subside over time.

Themethod of feature vectors in the time-frequency domain, which can effectively remove noise, decrease dimension and extract feature of signal, was proposed and comformed as a effective and practical technique. The ideal correct rate of classification and the external control signal of BCI system were obtained by using pattern recognition and classification method based on CNN to recognize features of mental EEG for imaging left-right hands movement. It is an available method of signal pattern recognition. This study has provided a new idea and method for feature extraction and classification of different mental tasks for BCI.

References

[1] Nguyen T, Khosravi T, Creighton D, Nahavandi S.*Fuzzy system with tabu search learning for classification of motor imagery data*[M].Biomed Signal Proces. 20, 61-70(2015)

[2] Muller-Putz GR, Kaiser V, Solis-Escalante T, Prurtscheller G.*Fast set-up asynchronous brain-switch based on detection of foot motor imagery in 1-channel EEG*[M]. Med Biol Eng Comput.48, 229-233(2010)

[3] Nguyen T, Khosravi A, Creighton D, Nahavandi S.*EEG data classification using wavelet features selected by Wilcoxon statistics*[J]. Neural Comput Appl. 26, 1193-1202(2015)

[4] Ahangi A, Karamnejad M, Mohammadi N, Ebrahimpour R, Bagheri N.*Multiple classifier system for EEG signal classification with application to brain–computer interfaces*[J]. Neural Comput Appl. 23, 1319-1327(2013)

[5] Xanthopoulos P, Pardalos PM, Trafalis TB.*Linear discriminant analysis*[M]. *Robust data mining,* 1st edn. Springer, Berlin. 27-33(2013)

[6] Xanthopoulos P, Pardalos PM, Trafalis TB Support vector machines. In: *Xanthopoulos P, Pardalos PM, Trafalis TB* [M]. Robust data mining, 1st edn. Springer, Berlin.35-48(2013)

[7] H. Parvin, H. Alizadeh, and B. MinaeiBidgoli. MKNN: *Modified k-nearest neighbor*[M]. In Proc. World Congress on Engineering and Computer Science. San Francisco, USA.(2008)

[8] G. E. Hinton, R. Salakhutdinov. *Reducing the Dimensionality of Data with Neural Networks*[J]. Science, vol. 313, 504-507(2006)

[9] P.Y. Simard, D. Steinkraus, and J.C. Platt.*Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis*[M]. Document Analysis and Recogni-tion,958-962(2003)