A BCI System with Motor Imagery Based on Bidirectional Long-Short Term Memory

Jzau-Sheng Lin and Bing-Hong She

Dept. of Computer Science and Information Engineering, Nat'l Chin-Yi University of Technology, Taichung, Taiwan, R.O.C
Email: jslin@ncut.edu.tw and tombear07@gmail.com

Abstract. A bidirectional long short term memory (BiLSTM) neural network was embedded into a brain-computer interface (BCI) system based on motor-imagery (MI) in this paper. The MI-based electroencephalogram (EEG) signals were used to recognize different imagery actions. The dynamic characteristics of MI signals in EEG are usually low signal-to-noise ratio as non-stationary time series. A lot of strategies have been proposed to clustering MI-EEG signals. However they are not considering the concept of series features of the signal in time domain with forward and backward manners, so the recognition results are not promising. The discrete wavelet transform (DWT) was also used to get the frequency feature from transforming each channel of MI-EEG in this paper. Then the proposed BiLSTM is proposed as a classifying system to identify the MI-EEG data. BiLSTM can extract dependencies of different time points by each recurrent unit with an adaptive manner. Besides the forward manner of time series signals in the LSTM unit, the BiLSTM also puts the output signals into previous layers with backward manner. The BiLSTM system can get more promising results in the classification of MI-EEG than those obtained by other strategies shown as in experimental results.

1. Introduction

The BCI system is a different way of intercourse through EEG signals to support one of the most important aspects. The response of brain wave is translated into activities by using of electroencephalogram (EEG) signals extracted from electrodes contacting on scalp in a BCI system. These EEG signals can be processed to provide a communication channel through system's hardware and software to control several systems such as computer games, electric wheelchair and so on. Being one of a topic in the research field of BCI, motor imagery (MI) emulates a given mental activity, e.g., imaging the motions of the legs [1]. MI refers to the visualization of any moving action which responses the various voltage changes in the connectivity between neurons in the cortex without any actual executing activity that is in either an event related desynchronization (ERD) or event-related synchronization (ERS) of mu rhythms. These effects are resulted from the change of intrinsic membrane properties of local neurons, the change in strength between the interconnections or the change of micro voltage in the chemical synapses between neurons. A lot of strategies were used to control electric wheelchair. An MI-based BCI system to control electric wheelchair was constructed by Lin and Lo [2]. In 2017, Sreeja et al. [3] presented two different techniques and modeling using Gaussian Naive Bayes (GNB) classifier for the purpose of selecting optimal features to construct a machine-learning classification platform for MI-EEG signals. They proved that their strategy can provide improved accuracy than LEA and SVM methods. A classification framework and data reduction method for classifying MI-EEG signals was proposed by Guan et al. [4] in 2019. They used the manifold of covariance matrices in a Riemannian perspective through decision tree framework

with filter geodesic minimum distance to recognize MI tasks.Xie et al. [5] proposed a symmetric positive-definite (SPD) covariance matrix for EEG signals to convey important discriminant information for the MI BCI system in 2016. In 2015, some classification strategies for MI-Based BCI system were surveyed by Jois et al. [6]. They pointed out that general features such as band power values in the EEG signals can be extracted by proper classification methods like neural networks, SVM or ensemble classifiers. The different classifiers can obtain different performances and are compared to find the better techniques for using equal number of features. They proved that the neural network strategies can get the most efficient performance. One barrier of the general neural networks for their wide application is the initial conditions must be set carefully. Generally, small initial value could result in the multilayer network un-trainable due to weight diffusion, and large weight values could make wicked local minima [7]. For the purpose of resolving this obstacle and organize higher promising neural networks, new novel model of techniques and methodologies, called deep learning (DL), have been effectively proposed and become dominant in some areas [8]. Recently, one of models in DL such as recurrent neural networks (RNNs) have been certified that they can obtain promising results in many research field [9] especially in time-sequence processing with variable-length input/output. For the problems in EEG signals classification, Petrosian et al. [10] used wavelet transform to extract the compact features and RNN to classify EEG signals into several categories. Because the scalp EEG containing external noises causes not promising results, the RNN is not suitable to classify EEG signals from scalp. Several researchers have proved that the RNN network can results in gradient explosion and gradient vanished when the weights are updated layer by layer. The RNNs with Long short-term memory (LSTM) [11] using the time-series features of signals is an effective deep learning model in several sequential-data applications. Not only the problems in RNN such as gradient explosion and gradient vanished can be solved but also the long time information can be stored by the memory cell in the LSTM-based RNN. The gating mechanism is organized in the LSTM to avoid vanishing gradients. In addition to occupy the LSTM unit, BiLSTM also feedback transfers the output into previous layers and sequence data with variable-length manner can be captured by each recurrent unit adaptively.

In this paper, we proposed a platform that the DWT was used to extract compact data and the BiLSTM was applied to classify the EEG signals. Each channel, the MI-EEG signalt is extracted and converted by a DWT with an effective time-frequency characteristics. Then we calculated the average power spectrum of MI-EEG signals and determined the effective time segment. The experimental results showed that BiLSTM method can make full use of the time-frequency information of MI-EEG as well as time sequence information, and can get promising classification results.

The organization of this paper is listed as follows. The system architecture is described in Section 2, Section 3 discusses the DWT. The LSTM recurrent network is presented in Section 4. The architecture of BiLSTM is shown in Section 5. Section 6 discusses the classification accuracies obtained by the BiLSTM compared with other strategies. Finally, the conclusion is given in Section 7.

2. System Architecture

The Emotiv EPOC chip and g.tec dry electrodes were combined an extracting subsystem to extract EEG signals from electrodes located on C3, C4 and Cz in the proposed BCI system. MI brain signals including right and left hand actions were recognized in the proposed MI-EEG based BCI system respectively. For the purpose of establishing a sampling model, The DWT is used to transform the extracted brainwave signal to get the spectrums in frequency domain. Then the BiLSTM is used to classify the features with frequency manner into different labels. The NVIDIA Jetson TK1 was also embedded in the proposed platform in order to promote the calculating speed. Additionally, the Bluetooth 2.1 was also used to transfer to EEG signals from extracting subsystem to computer with wireless manner. Figure 1 shows the proposed BCI platform.
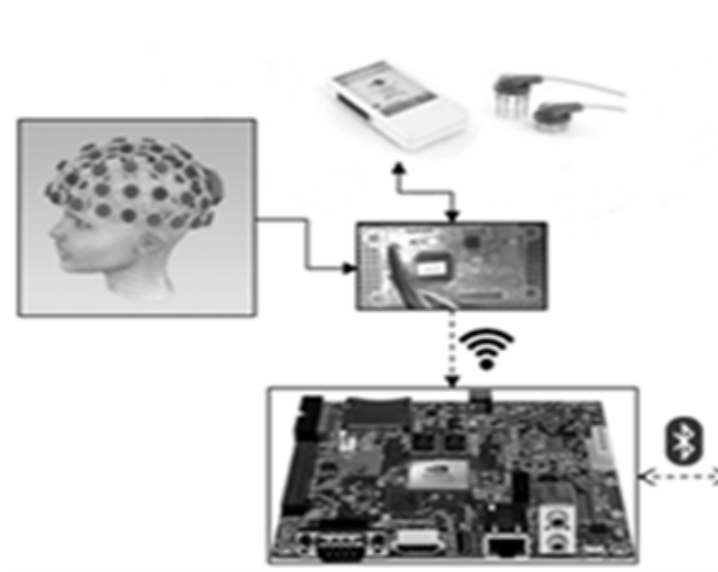
Figure 1. The proposed BCI platform

3. Wavelet Transform

In 1981 Jean Morlet proposed the strategy of wavelet. In applications of noise filtering, digital signal analysis, and signal compression and so on, the wavelet is always used. Several series db with Daubechies wavelet [12] can obtain better results in signal analysis. db4 wavelets were used in order to capture main features from EEG signals in this paper. The single-resolution WT is not easy to get detailed features when a signal has an altitude-manner diversity in a proper field. Instead, the multi-resolution methodology can decompose the lower-degree signals to obtain more message. Therefore, the signals with lower frequency can be continuously decomposed to show more characteristics. However, a lot of decomposition iterations of the signal can results in few number of samples that make less obvious signal features. Therefore, the number of layers for the decomposition of signals is limited. A high-pass filter and a low-pass filter are respectively used to transform the original signal in wavelet decomposition. The consistency of the original signal is retained in the low-pass filter while the variability of the original data is reversed high-pass filter. The wavelet and scale functions can be combined with the DWT transform. It has a lower-frequency resolution and a higher-time resolution in the high frequency. And in the low frequency part, it occupies a high frequency resolution and low temporal resolution.

4. LSTM Recurrent Network

The RNNs are called recurrent module because it performs a work for every element of sequential signals with the output being depended on the previous states in the hidden layers. Being popular networks, the RNNs have been shown great promise in many sequence-manner tasks. More complex types of RNNs have been created to resolve several problems recently. Same as the conventional the neural networks, the training process of RNNs have difficult learning process with long-term dependencies owing to the problems of gradient vanishing and exploding. Without a fundamentally different organization from RNNs, the LSTMs use a different manner to update the states in hidden layer. The memory cell in the LSTMs can be worked as a black box that input the previous state and current input data. The data in memory will be internally decided what to be kept in or erased from memory by this cell. Then the input data, the current memory, and the previous state to be combined to construct a neuron node. It is proved that these neuron node can capture long-term dependencies with a very efficient manner. The architecture of the LSTM is shown as in Figure 2. Eq. (1) shows the state of forget gate f_t which is calculated by a sigmoid function from previous cell state c_{t-1}, previous hidden layer state h_{t-i} and input data x_t .

$$f_t = \sigma(w_{c,f} C_{t-1} + w_{x,f} x_t + w_{h,f} h_{t-i}) + b_f \tag{1}$$

The function of cell state can be derived in Eq. (2), It is calculated by the forget-gate state f_t , previous cell state c_{t-1}, and $i_t * \widetilde{C}_t$.

$$\mathbf{C}_t = f_t * C_{t-1} + i_t \widetilde{C}_t \tag{2}$$

where

$$i_t = \sigma(w_{c,i} * C_{t-1} + w_{x,i} * x_t + w_{h,i} * h_{t-1} + b_i) \tag{3}$$

and

$$\widetilde{C}_t = \tanh(w_{x,c} * x_t + w_{h,c} * h_{t-1} + b_c) \tag{4}$$

Then, the states of hidden layer and output gate can be calculated by Eq. (5) and Eq. (6), respectively.

$$h_t = o_t \tanh(\mathbf{C}_t) \tag{5}$$

$$o_t = \sigma(w_{c,o} * C_t + w_{x,o} * x_t + w_{h,o} * h_{t-1} + b_o) \tag{6}$$
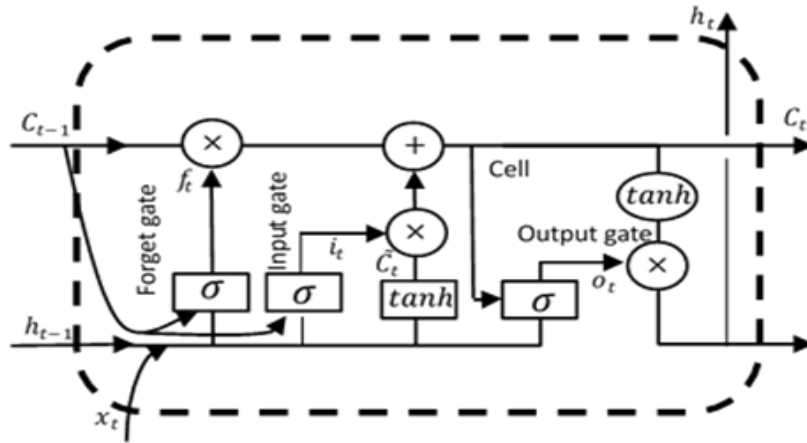

Figure 2. Long-Short term Memory

5. Bidirectional LSTM

A recurrent neural network (RNN) can iteratively compute the hidden vector sequence h_t and output vector sequence y_t for an input sequence x_t with Eqs. (7) and (8).

$$h_t = \sigma(w_{xh} x_t + w_{hh} x_{t-1} + b_h) \tag{7}$$

$$y_t = w_{hy} h_t + b_y \tag{8}$$

where w is the weight matrices, b denotes the bias vectors and σ is the sigmoid function in hidden layer. LSTMs preserve information from inputs that has already passed through it using the hidden state because the only inputs it has seen are from the past. In order to improve model performance on sequence classification problems, LSTMs can be extended as BiLSTMs. Instead of one direction in LSTMs, BiLSTMs are a modification of the conventional LSTMs to process sequences of signals in both directions, one from past to future and the other from future to past. This can provide additional context to the network and result in fuller learning on the problem.

As shown in Figure 3, combine RNN and LSTM to construct a BiLSTM which uses past and future states to access long-range context in both directions to predict the imaging action of a given EEG signal. As illustrated in Figure 3, the states of forward hidden sequence and the states of backward hidden sequence are calculated by the BiLSTM iteratively. The output sequence y from the backward layer and the forward layer then can be updated. They are derived as following equations.

$$\vec{h}_t = \sigma(w_{x\vec{h}}x_t + w_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}}) \tag{9}$$

$$\overleftarrow{h}_t = \sigma(w_{x\overleftarrow{h}}x_t + w_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}) \tag{10}$$

$$y_t = w_{\vec{h}y}\vec{h}_t + w_{\overleftarrow{h}y}\overleftarrow{h}_t + b_y \tag{11}$$
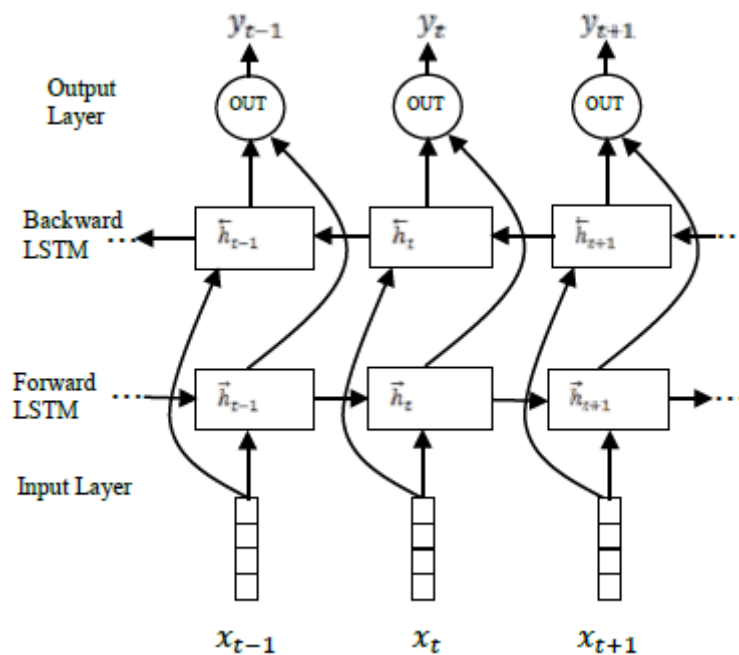

Figure 3. Bidirection Long-Short term Memory

6. Experimental Results
The EEG signals were extracted on locations C3, Cz and C4 in this paper and Emotiv EPOC chip, g.tec dry electrode and Ultracortex helmet are connected to record MI-EEG signals such as imagine left-hand and right-hand movements. It is consumed 9 seconds for each imaginary action to get a data set. The wavelet transform was used to transform EEG signals those were extracted 28 times to obtain their features. Within an interval of two minutes, a data set every 9 seconds was obtained in the experimental data acquisition process. On the first two seconds, the waiting time is set then the testing process is started and a cross sign "+" was displayed for one second after a stimulus signal was sound. Then the right- or left-arrow is shown to hint a subject to imaging right- or left-hand moving. In the acquisition process, the sampling rate is set 128Hz.

Since MI-EEG signals were captured from the electrodes on C3, Cz, and C4 then classified them into several groups. The forward and backward hidden layers in the proposed BiLSTM are all set into seven neurons while three channels are set for the proposed MI-EEG based BCI system. For the purpose of evaluating the classification results and obtain a stable and reliable BCI system, the classification accuracy is estimated by this proposed model executed 500 cross validation. In order to prove a stable and flexible system being obtained, the proposed methods compared to other strategies based on "BCI Competition 2003" [13]. The experimental results are shown in Table 1. From Table I, we can find that the proposed BiLSTM can get better performance than others.

Table 1. The Accuracy Rates of Different Strategies for BCI Competition 2003

Authors	Features	Classifiers	Accuracy rates
Akash Narayana [13]	AR	LDA	84.29%
GAO Xiaorong [13]	ERD	LDA	86.43%
The proposed BiLSTM	DWT	BLSTM	87.14%

7. Conclusions

In this paper, a deep-learning model named BiLSTM was applied to be embedded into a BCI system for MI-EEG signals to identify two imagery movements such as imaging right hand and imaging left hand actions. In the proposed BCI system, the Emotiv EPOC IC with brain waves helmet to capture brainwave signal on C3, Cz, and C4. In this paper, we use the Daubechies wavelet to get feature values on db4 coefficient. The BiLSTM can make each recurrent unit to capture variable-length sequences adaptively. Modified from LSTM, the BiLSTM has both feeding directions, one from past to future and the other from future to past. This can provide additional context to the network and result in fuller learning on the problem. The experimental results show that the BILSTM can get better performance than other strategies.

8. Acknowledgments

9. References

[1]   Decety J and Ingvar D H, 1990 *Acta Psychol* **73**, 13
[2]   Lin J S and Lo C H 2016 *Int. J. of Computer, Consumer and Control*, **5** 18
[3]   Sreeja S R, Joytirmoy R, Nagarjuna K Y, Samanta D, Mitra P and Sarma M 2017 *Proc. Int. Conf. on New Trends in Computing Sciences* (Amman:Jordan) p 61
[4]   Guan S, Zhao K and Yang S 2019 *Computational Intelligence and Neuroscience*, **29** https://doi.org/10.1155/2019/5627156
[5]   Xie X, Yu Z L, Lu H, Gu Z and Li Y 2016 *IEEE Trans. On Neural Systems and Rehabilitation Engineering* **25** 504
[6]   Jois K, Garg R, Singh V, and Darji A 2015 *IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions* (WCI) DOI: 10.1109/WCI.2015.7495507
[7]   Hinton G E and Salakhutdinov R R 2006 *Science* **313** 504
[8]   Bengio Y, Courville A, and Vincent P 2013 *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 1798
[9]   Graves A, Mohamed A -R, and Hinton G 2013 *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*(Vancouver, BC, Canada) p 6645
[10]   Petrosian D, Prokhorov V, Homan R and Wunsch D C 2000 *Neurocomputing* **30** 201.
[11]   Li M, Zhang M, Luo X and Yang J 2016 *Proc. of IEEE International Conference on Mechatronics and Automation* (Harbin, China) p 1971
[12]   Daubechies I 1988 *Communications on Pure and Applied Mathematics*, **41** 909.
[13]   Blankertz B, Müller K R and Curio G 2004 *IEEE Transactions on Biomedical Engineering* **51**1044