

A novel motor imagery EEG recognition method based on deep learning

Ming-ai Li ^a, Meng Zhang ^b, Yan-jun Sun

College of Electronic Information & Control Engineering, Beijing University of Technology, Beijing, China

^alimingai@bjut.edu.cn, ^bvzhangmeng726@163.com

Keywords: Motor Imagery EEG. Deep Belief Networks. Wavelet Packet Transform. Softmax. Brain Computer Interface. Deep Learning.

Abstract. The Motor Imagery electroencephalogram (MI-EEG) is time varying and subject-specific, its recognition needs the perfect adaptability and combination of feature extraction method and classifier. In this paper, Deep Belief Networks (DBN) is integrated with Wavelet Packet Transform (WPT) to yield a novel recognition method, denoted as WPT-DBN. Firstly, the MI-EEG is transformed into power signal and analyze the effective time domain. Then, WPT is applied to each channel of MI-EEG to obtain the effective time-frequency information. Finally, DBN is used for the identification and classification simultaneously. Experiments are conducted on a publicly available dataset, and the 5-fold cross validation experimental results show that WPT-DBN yields relatively higher classification accuracies compared to the existing approaches.

Introduction

Brain computer interface (BCI) is a communication system that does not depend on the brain's normal output pathways of peripheral nerves and muscles. It is an alternative and a novel interface between human and computers [1]. With the aging of population, the incidence of stroke is increasing. According to the plasticity of the brain, with active participation in sports training is more conducive to the rehabilitation of the patients [2]. Brain computer interface based on motor imagery EEG (MI-EEG) can improve the rehabilitation effect of patients with motor dysfunction, and this method has obvious advantages in improving the recovery of motor function [3]. BCI has become a research hotspot in the field of rehabilitation engineering.

Because of the nonlinear, non-stationary, time-varying sensitivity and individual difference of the MI-EEG, the key problem of BCI system is how to identify it quickly and accurately. Deep learning (DL) discover characteristic of distributed presentation of data by simulating brain information processing mechanism and combine low-level features to form a more abstract high-level characteristics[4]. Deep belief network (DBN) is one of the most commonly used models in DL, by introducing a new method of training, it solves the problem of deep neural network optimization. It is widely used in the images and voices processing [4], but has not been widely used in field of MI-EEG. DBN can extract essential feature of MI-EEG signals which come from the brain, thus it is more conducive to the classification of MI-EEG. However, the MI-EEG signal contains a large amount of time and frequency information. If the information in the time domain and the frequency domain is not fully utilized, the DBN model is very difficult to obtain a good classification effect.

This paper presented a novel MI-EEG recognition algorithm (WPT-DBN) which combined WPT and DBN method. Firstly, the MI-EEG is transformed into power signal and analyze the effective time domain. Secondly, WPT is applied to each channel of MI-EEG to obtain the effective time-frequency information. Thirdly, unsupervised training methods was adopted to preliminary training of each layer, then use supervised training method to fine-tuning of the whole network. Finally, pattern classification was realized by softmax classifier. The experimental results showed that improved DBN can make full use of MI-EEG information in time and frequency domain and can achieve better classification performance.

Method

Deep learning. DL achieves signal recognition by constructing multi-layer neural network structure and describing signal characterization in phases [4]. Commonly used DL model include DBN, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN). Among them, DBN is one of the most applied model. It is a deep neural network which composed of a plurality of restricted Boltzmann machine (RBM), has excellent feature learning ability, can learn more essential characteristic of the data, thereby DBN is favorable for data classification.

DBN obtaining the initial values of network using unsupervised method by initialized training of each layer, and then fine-tuning the entire network through supervised Error Back Propagation (BP) algorithm, solve the problem of training multilayer neural network. DBN shown in Fig.2. When one layer of RBM training is completed, the output of the previous layer will be treated as the next layer input, like this, initialized training of the network will carry on by layers. DBN is stacked up by multiple RBM, and it is formed by introducing softmax classifier in the last layer. The network makes sense of softmax labeling information, and then use BP algorithm to fine-tune the network.

RBM mode. RBM is an undirected graph model based on energy. Every RBM can be treated as a feature extractor [5]. As shown in Fig.1,

Energy of joint configuration of visible unit v and hidden units h given by:

$$E(v, h/q) = -\sum_{i=0}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i w_{ij} h_j \quad (1)$$

Where n, m are the number of visible unit and hidden unit, v_i, h_j are the binary states of i^{th} visible unit and j^{th} hidden unit, $q = \{w_{ij}, a_i, b_j\}$ are the parameters of the RBM, a_i, b_j are their biases, and w_{ij} is the weight between a_i and b_j .

Parameters w, b and v are obtained by contrastive divergence (CD) algorithm to minimize the reconstruction error training. CD algorithm is shown below:

$$\frac{\partial \log P(x)}{\partial W_{ij}} \approx \langle x_i h_j \rangle_{\text{data}} - \langle x_i h_j \rangle_{\text{recon}} \quad (2)$$

RBM does not work well with data of continuous values, while Gauss restricted Boltzmann machine (GRBM) can be modelled on continuous values [6]. The energy equation of GRBM is defined as follows:

$$E(v, h) = -\sum_{i=\text{vis}} \frac{(v_i - b_i)^2}{2s_i^2} - \sum_{j=\text{hid}} b_j h_j - \sum_{i,j} \frac{v_i}{s_i} h_j w_{ij} \quad (3)$$

Where the parameters are $q = \{w, b_i, b_j, s_i, s_j\}$, these parameters are also obtained by CD algorithm.

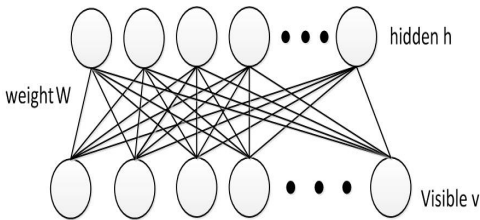


Figure 1. The structure of RBM.

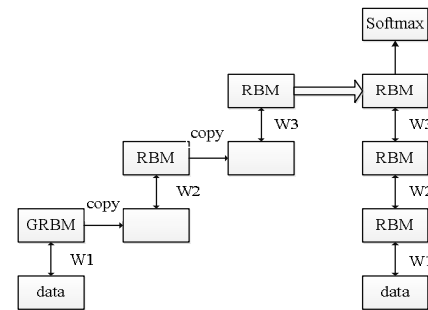


Figure 2. The structure of DBN network.

The method of wavelet packet. WPT method can not only decompose the low-frequency and the high-frequency portion of data, but can also select the corresponding bands according to the signal characteristic, it is a more elaborate decomposition method than the wavelet. WPT retains the effective time-frequency information by filtering and reconstructing MI-EEG to ensure valid signal will not be lost [7]. In this paper, the WPT method was adopted to decompose MI-EEG by time-frequency, and then frequency band of Mu rhythm and Bate rhythm were selected to be

reconstructed. Wavelet packet improved the frequency resolution by adopting binary to subdivide frequency of wavelet subspace.

Assuming U_l^r represents the r^{th} ($r = 1, 2, 3, \dots, 2^l - 1$) wavelet packet subspace in the l^{th} layer. The orthogonal group corresponding to the space is $u_{l,k}^r(t) = 2^{-l/2} u^r(2^{-l}t - k)$, k is the translation factor which satisfies the following Two-Scale equation:

$$u_{l,0}^r = \sum_k g_0(k) u_{l-1,k}^r \quad (r \text{ is even})$$

$$u_{l,0}^r = \sum_k g_1(k) u_{l-1,k}^r \quad (r \text{ is odd})$$

Where $l, k \in \mathbb{Z}$, $r = 1, 2, 3, \dots, 2^l - 1$, $g_0(k)$ a low-pass filter and $g_1(k)$ a high pass filter are a set of mutually orthogonal filter, which meets the following equation $g_1(k) = (-1)^{1-k} g_0(1-k)$.

Using orthogonal wavelet packet to decompose transform signal, the coefficient of wavelet packet of k^{th} point in the l^{th} layer follows the following equation:

$$d_l^{2r}(k) = \sum_m g_0(m-2k) d_{l-1}^r(m) \quad (6)$$

$$d_l^{2r+1}(k) = \sum_m g_1(m-2k) d_{l-1}^r(m) \quad (7)$$

Where, after decomposition of wavelet packet, the original signal is divided into a number of wavelet packet subspaces according to frequency band, each band of subspace is:

$$\{[0, \frac{f_s}{2^{l+1}}]; [\frac{f_s}{2^{l+1}}, \frac{2f_s}{2^{l+1}}]; [\frac{2f_s}{2^{l+1}}, \frac{3f_s}{2^{l+1}}]; \dots; \mathbb{L}; [\frac{(2l-1)f_s}{2^{l+1}}, \frac{f_s}{2}]\} \text{ Where } f_s \text{ is the signal sampling rate.}$$

The reconstruction formula of the wavelet packet coefficient of the k^{th} point in the $(l+1)^{\text{th}}$ layer is

$$d_{l+1}^r(k) = \sum_m g_0(m-2k) d_l^{2r}(m) + \sum_m g_1(m-2k) d_l^{2r+1}(m) \quad (8)$$

Through decomposition and reconstruction of wavelet packet, Mu rhythms and Beta rhythms energy of which is different clearly are gotten. After getting effective frequency band, the classical method directly process the two bands of MI-EEG signals, however the dimension of feature space is still high, the method does not work effectively. The classical methods requires some criterion to extract further feature, however DL which has a strong presentation skills of characteristics can directly recognize the two frequency signal pattern and classify. DL with strong versatility is more conducive to eliminate the individual differences.

Experiment and analysis

Experimental data. The experiment data were from Data set III of the “BCI Competition 2003” contest database. Database was collected from 140 motor imagery experiments. Three bipolar MI-EEG channels were measured over C3, Cz and C4. The MI-EEG was sampled with 128Hz.

MI-EEG analysis and preprocessing. MI-EEG signal collected through the electrode cap were stored as the voltage amplitude, so the instantaneous power was calculated as follows:

$$P_j^i = (x_j^i)^2 \quad (8)$$

Where x_j^i was the j^{th} MI-EEG data in the i^{th} trial, P_j^i represented the average of the instantaneous power of MI-EEG.

Assuming the average power of the j^{th} data by N trials, the average power calculated in (17):

$$\bar{P}_{(j)} = \frac{1}{N} \sum_{i=1}^N (x_j^i)^2 \quad (9)$$

When N=140, the average power during 0~9s was obtained for left-hand on electrode C3 or C4 respectively, namely PLC3, PLC4, and for right-hand on electrode C3 or C4 , namely PRC3, PRC4. From Fig.3 and Fig.4, there was an obvious difference between left-imagery and right-imagery in 3.5~8s.

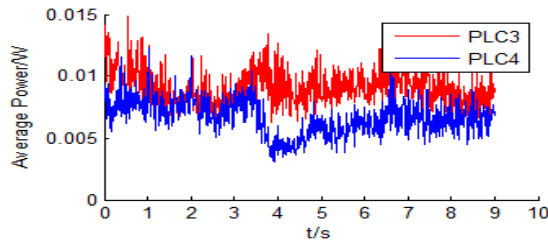


Figure 3. Variation of the average power of C3 and C4 with time for imagery left-hand movement.

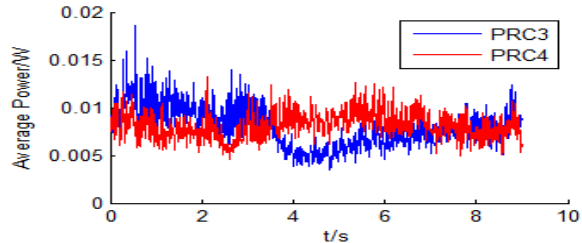


Figure 4. Variation of the average power of C3 and C4 with time for imagery right-hand movement.

Determination of DBN structure. Under certain conditions, with the increase of neural network layers, the ability to hold information of model would enhance, and the model could achieve better classification results. The experiment treated the power of MI-EEG signal as input data, using WPT-DBN as a model. The results obtained were as follows, the set of network parameter was in table 1, the identification results of different layers were shown in Fig.5.

As shown in the Fig.5, the accuracy of network classification increased with the increase of the number of layer, but up to a certain number of layers, accuracy rate of classification began to reduce. This is because the different layers of DBN has different capacity. When the same amount of training data, the DBN network layer has increased to a certain value, and then the results are getting worse. So the network layer was determined to 5 in the end.

Table 1. Main network parameter setting table.

Unsupervised process	
Learning Rate	0.01
Iterations	150
Momentum a	0.5
Cost function weights	0.002
Supervised process	
Learning Rate	0.05
Iterations	100

Table2. Methods and results of three methods in BCI contest

Author	Feature extraction	Classification	Recognition rate
Christin Schafer	Wavelet	Bayes	89.29%
GAO Xiaorong	ERD	LDA	86.43%
Akash Narayana	AR	LDA	84.29%
Proposed method	WPT+DBN	softmax	93.57%

Comparison with other algorithms. Based on the same set III Data, this paper identifies the method and results of the first three of the "Competition BCI". The results are shown in Table 2. As can be seen from the table, the method of this paper is to improve the recognition rate of 4.28%.

Based on the same data set, the method of this paper is compared with the classical methods such as WP + LDA[8], WT + LDA [9], WPD + Fisher [10], WPD + SVM [11]. The results are shown in Figure 6:

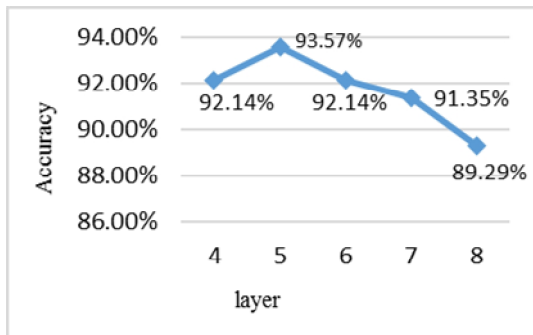


Figure 5. The effect of neurons number in hidden layer to the classification accuracy.

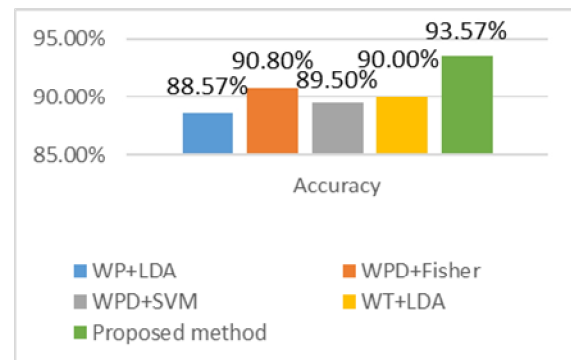


Figure 6. The comparison of six methods recognition result.

As shown in the Fig.6, WPT-DBN method is better than the other five methods of recognition results are improved, the method can effectively improve the classification accuracy of MI-EEG.

Conclusions

DL in the areas such as computer vision has been a huge success, but in MI-EEG its effect is not very good, and it has not been in common used. Using DL model to process MI-EEG, if the MI-EEG time series data directly used as static data such as image, will lost time-frequency domain information. WPT can make full use of MI-EEG time-frequency information and DBN has strong ability of feature representation. This paper was a specific application of DL in processing time series data, offering a novel thought for applications of DL in handling time series data.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (No.81471770) and the Natural Science Foundation of Beijing (No.7132021).

References

- [1] Wolpaw J R, Birbaumer N, et al. Brain-computer interface technology: a review of the first international meeting[J]. IEEE transactions on rehabilitation engineering, 2000, 8(2): 164-173.
- [2] Ungerleider L G, Doyon J, et al. Imaging brain plasticity during motor skill learning[J]. Neurobiology of learning and memory, 2002, 78(3): 553-564.
- [3] Fu M J, Daly J J, et al. Assessment of EEG event-related desynchronization in stroke survivors performing shoulder-elbow movements[C]//Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on. IEEE, 2006: 3158-3164.
- [4] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [5] Hinton G, Osindero S, et al. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.
- [6] Bengio Y, Lamblin P, et al. Greedy layer-wise training of deep networks[J]. Advances in neural information processing systems, 2007, 19: 153.
- [7] Zandi A S, Javidan M, et al. Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform[J]. Biomedical Engineering, IEEE Transactions on, 2010, 57(7): 1639-1651.
- [8] Baoguo X, Aiguo S. EEG signal recognition method based on wavelet packet transform and clustering analysis [J][J]. Chinese Journal of Scientific Instrument, 2009, 1: 007.

- [9] Xu B, Song A, et al. Algorithm of imagined left-right hand movement classification based on wavelet transform and AR parameter model[C]//Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on. IEEE, 2007: 539-542.
- [10] Ting W, Guozheng Y, et al. EEG feature extraction in brain computer interface based on wavelet packet decomposition[J]. Chinese Journal of Scientific Instrument, 2007, 28(12): 2230.
- [11] Yanjing W, Xiaoyan Q, et al. Classification of motor imagery task based on wavelet packet entropy and support vector machines [J]. Chinese Journal of Scientific Instrument, 2010, 12: 2729-2735.