A Deep Learning Method for Classification of EEG Data Based on Motor Imagery

Xiu An, Deping Kuang, Xiaojiao Guo, Yilu Zhao, and Lianghua He

The Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, China
Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
anxiuaijia88@sina.cn

Abstract. Effectively extracting EEG data features is the key point in Brain Computer Interface technology. In this paper, aiming at classifying EEG data based on Motor Imagery task, Deep Learning (DL) algorithm was applied. For the classification of left and right hand motor imagery, firstly, based on certain single channel, a weak classifier was trained by deep belief net (DBN); then borrow the idea of Ada-boost algorithm to combine the trained weak classifiers as a more powerful one. During the process of constructing DBN structure, many RBMs (Restrict Boltzmann Machine) are stacked on top of each other by setting the hidden layer of the bottom layer RBM as the visible layer of the next RBM, and Contrastive Divergence (CD) algorithm was also exploited to train multilayered DBN effectively. The performance of the proposed DBN was tested with different combinations of hidden units and hidden layers on multiple subjects, the experimental results showed that the proposed method performs better with 8 hidden layers. The recognition accuracy results were compared with Support vector machine (SVM) and DBN classifier demonstrated better performance in all tested cases. There was an improvement of 4 – 6% for certain cases.

Keywords: Deep Learning, Motor Imagery, EEG, Brain-computer interface, Ada-boost.

1 Introduction

Brain-computer interface is a communication control system without depending on the normal output pathways that composed by brain, peripheral nerve and muscles, that can transfer brain information and realize control by using computer or electrical device to analyze the brain activities under specific task [1].

A lot of studies have done by Shang-Lin Wu and his follows indicates that using the common spatial pattern (CSP) for feature extraction from EEG and the linear discriminate analysis (LDA) for motor imagery classification obtained an average classification accuracy of 80% for two subjects [4]. Additionally, Yohimbe Tom ita et al. proposed bimodal approach that using near infrared spectroscopy (NIRS) simultaneously with EEG to measure the hemodynamic fluctuations in the brain during stimulation with steady-state visual evoked potentials (SSVEP) made the wrong classification for 9 classes for 13 subjects [7,18]. The studies conducted by Like's

D.-S. Huang et al. (Eds.): ICIC 2014, LNBI 8590, pp. 203–210, 2014.

group demonstrated that combining multi-scale filters and Principal Component Analysis (PCA) to enhance the classification performance in identifying EEG signals works and achieve a classification accuracy of 91.13% [2].

Deep Learning is a new field in itself in the machine learning whose motivation is to simulate the human brain's mechanism to explain the data by the composition of multiple non-linear transformations of the data, with the goal of obtain more abstract and ultimately more useful representation [3-5].

Despite the success of DBN, its application in Electroencephalogram (EEG)-based Brain Computer Interaction (BCI) is still rare. The main difficulty is the enormously high feature dimensionality spanning EEG channel, frequency, and time [10]. Deep Learning algorithm has shown superior learning and classification performance in fields such as computer vision, natural language processing and other areas for its excellent feature extraction capabilities except the field of EEG data analysis [15]. In this paper, a classify method was proposed based on deep learning with Ada-boost algorithm. Using this method, the misclassification rate of EEG signals decreased even using fewer channels. The final powerful classifier is combined by several weak classifiers which are trained using single channel data.

The sections below are our learning method and experiments. Section 2 mainly describes the RBM and DBN method and Section 3 presents our experimental results on MI EEG dataset.

2 Method

2.1 Theory of Deep Learning

Deep learning algorithm focus on learning multiple levels of representation of raw data automatically, using a deep architecture which composed of many hidden layers. This algorithm automatically extracts the high-level features necessary for classification which involving more meaningful information that hierarchically depends on other features. Here we use DBN model which formed by a plurality of RBM, each RBM is trained greedily and unsupervised [5].

An RBM has a single layer of hidden layer that are disconnected with the units in the same layer and have undirected, symmetrical connections to the units in the visible layer which makes it easy to compute the conditional probabilities. The key issue of training the RBM is to get generative weights. As shown in Fig. 1, W represents the weights between visible and hidden layers, b, c correspond to the bias of visual and hidden layer respectively [3].

The type of RBM we employed in this work is Gaussian RBMs which use real-valued visible units for training the first layer of the DBNs. Fig. 1 shows the structure of the RBM:

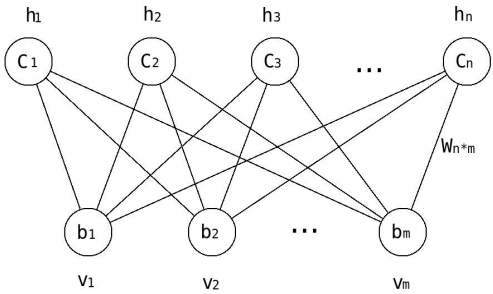

Fig. 1. Structure of RBM

2.2 Classification of MI Based on Deep Belief Net

Now, let v represents the feature vector containing only one channel features. An RBM defines a joint distribution on it, regard as the visible units in DBN and h, the hidden units as follow format [13]:

$$p(v) = \frac{\sum e^{-E(v,h)}}{\sum_u \sum_g e^{-E(u,g)}} \tag{1}$$

Where E is the energy function defined as

$$E(v,h) = \sum_{j \in visible} a_j v_j - \sum_{j \in hidden} b_j h_j - \sum_{i,j} v_j h_j \omega_{ij} \tag{2}$$

Where V_j,h_j are the binary states if visible unit i and hidden unit j, a_i, b_j are their biases and w_{ij} is the weight between them. The network assigns a probability to every possible pair of a visible and a hidden vector via this energy function:

$$p(v,h) = \frac{1}{Z} e^{-E(v,h)} \tag{3}$$

The probability that the network assigns to a training data can be optimized by adjusting the weights and biases to lower the energy of it. The derivative of the log probability of a training vector with respect to a weight calculated as follow:

$$\frac{\partial \log p(v)}{\partial \omega_{ij}} = \frac{\sum_{v \in D} \partial \log p(v)}{\partial \omega_{ij}} = E_{data}\left[\frac{\partial E(v,h)}{\partial \omega_{ij}}\right] - E_{model}\left[\frac{\partial E(u,g)}{\partial \omega_{ij}}\right] \tag{4}$$

Where the first item is the expectation of $\partial E(v,h)/\partial w_{ij}$ responds to the training set D and the hidden variables are sampled according to the conditional distribution of the dataset on $p(h|v)$, given a randomly selected training sample, v, the binary state, h_j, of each hidden unit, j, is set to 1 with probability

$$p(h_j = 1 \mid v) = \sigma(b_j + \sum v_j w_{ij}) \tag{5}$$

Where $\sigma(x)$ is the logistic sigmoid function $1/(1+exp(-x))$. $v_i h_j$ is then an unbiased sample.

For no direct connections between visible units in an RBM, it is also the way to get unbiased sample of the visible unit similar as hidden unit, given a hidden vector

$$p(v_i = 1 \mid h) = \sigma(\alpha_i + \sum_j h_j \omega_{ij}) \tag{6}$$

For training an RBM classifier, the joint distribution of data and class labels, the visible vector is concatenated with binary vector of class labels. The energy function becomes:

$$E(v,l,h) = -\sum_i \alpha_i v_i - \sum_j b_j h_j - \sum_{i,j} \omega_{ij} - \sum_y c_y l_y - \sum_{y,j} \omega_{yj} h_j l_y \tag{7}$$

Where l is the binary class label and w_{ij} is the weights between hidden and label units.

2.3 Boost of the Single Channel Deep Belief Net

Based on the former test performance of the single channel, we adopt the idea of Ada-boost algorithm [8] that combine the weak classifier to one more powerful classifier [19][20]. Here the channel C3, C4, Fc4 were chose as the meta data and the combination tactics to boost each weak classifiers refer to follow [15][16]:

$$M_k(X) = \sum_{k=1}^{k} c_k f_k(X) \tag{8}$$

Where c_k is the estimated coefficient for each DBN model and each DBN model is produces a discrete classification for input data [17].

The whole structure of our model based on DBN shows as follow:

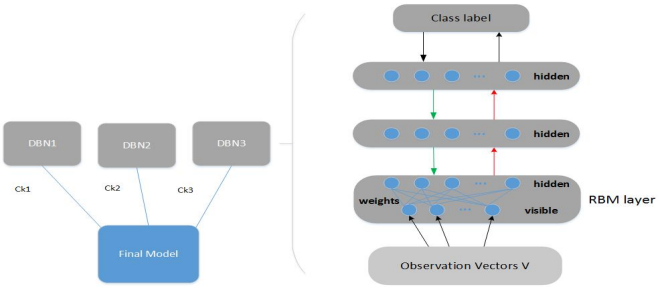

Fig. 2. Structure of the final model

3 Experiment

3.1 Experimental Data

The experimental data was collected from 4 subjects, all of them are students, male and without brain disease history. 30 trials left-hand imagination EEG data and 30 right ones were selected as sample for analysis for each subject. Fig. 3 shows the way to get

the final data which contained 7s data, of which the search adopted the data from 3s to 7s, sample rate is 250HZ/s and each of them contained 4s data, which means that each of them have 1000 sample points.

For the EEG data, de-noising processing and filtering were applied such as Elec-tro-Oculogram (EOG) as well as separated the data according to the channel. The EOG is removed by the Neuroscan software. And for the filtering work, this paper mainly analyzes the frequency band of 8-30Hz. So an elliptic filter was designed, with band-pass from 8 to 30Hz. And then converted the time domain data to frequency domain data via FFT (Fast Fourier Transformation) algorithm.

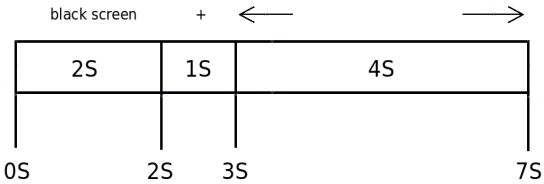

Fig. 3. Experiment time distribution for one trial

3.2 Test on Combine of the Single Channel

For each subject, 20 trials were selected as the train set from the total data with the remaining data to be the test samples from each channel. The weights were randomly initialized and the turning parameters were set as: learning rate for weight and biases = 0.07, momentum=0.5 and weight decay = 0.002. Four to sixteen layers were trained and tested for each channel of every subject. The results for eight layers worked better than others and the performance of nine and ten or more layers were very similar. In the paper we have not shown results for all of the layers due to space limitations. Table 1 shows the result of the classification performance with 7,8, 9, 10 layers respectively, using a fixed layer size of 2048 under the same condition and the result shows that the average recognition rate of DBN with 8 hidden layers is 0.81 which is higher than others.

The result for every subject with different hidden layers lists as Table 1 and the number represent the recognition accuracy rate:

Table 1. Performance of DBN with different layers

Subjects	DBN			
	7 hidden layers	8hidden layers	9 hidden layers	10 hidden layers
SHY	73%	85%	83%	80%
XB	56%	65%	59%	58%
ZJH	44%	77%	78%	74%
WDM	82%	95%	94%	96%

According to the results above which shows that the DBN classifier outperform others with 8 hidden layers, then we did the test for the different combination of hidden units under this condition, the result shows that there's no obvious effect on the performance. Table 2 shows the final performance of the recognition rate of all cases.

Table 2. Performance of DBN with different hidden nodes

Subjects	DBN_8 layers					
	2000-800-700-600-500-300-200-900	3000-1800-1700-1600-1500-1300-1200-900	4000-1100-1200-1300-1400-1500-1600-900	5000-2100-2200-2300-2400-2500-2600-900	6000-3100-3200-3300-3400-3500-3600-1900	8000-2100-2200-2300-2400-2500-2600-1900
SHY	83%	84%	84%	85%	84%	85%
XB	65%	66%	65%	65%	65%	64%
ZJH	73%	75%	77%	77%	75%	74%
WDM	96%	94%	93%	95%	95%	95%

The performances of SVM based on the same input features was further investigated.Fig.4 shows the contradistinction of recognition accuracy for DBN and SVM, the performance of SVM is inferior to that of DBN and the discrepancy is particularly evident for subject ZJH and WDM.

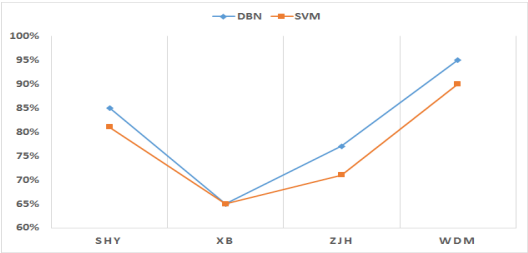

Fig. 4. The recognition accuracy of DBN vs. SVM

3.3 Experiment on Time Series

The experimental data was by time segments, and each section contains 1s as data to be classified. Fig.4 presents the performance of classification with different subjects. From the figure we can see that the average recognition rate of first 2 seconds can reach 83%, while the last 2 seconds is lower, we can explain that at the beginning of the experiment the subjects can preferably focus on the motor imagery experiment, but with the passage of time, the subjects may get absent-minded which would affect the validity of the experimental data, and finally leads to the low recognition accuracy.

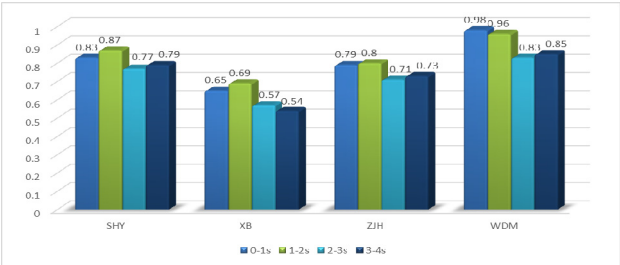

Fig. 5. Classification accuracy of time series

4 Conclusion

In this paper, a DBN classifier model for the classification of MI pattern is proposed. In the research, results showed consistent improvements for all tested cases over SVM through multiple cross-validation experiments. The test on different combination of hidden units was conducted, and it is found that the number of nodes had no obvious effect on the performance of classification. The experimental results showed that Deep Learning algorithm performs effectively on the task of classification with MI data. And the experimental results of time series show that the performance of classification depends on concentration of the subjects, for the accuracy rate is affected greatly by the status of the subject. Deep learning plays an important role in the process of classification because it can learn the advanced abstract representation from numerous unlabeled data. Our study suggests that DBN has great potential to be a powerful tool for the BCI research.

For the next stage, we'll try to employ this algorithm into classification of Multi-class based on EEG data, and merge more channels in order to take full use of the EEG data information to achieve better recognition results.

References

1. Reza, K., Chai, Q.: A Brain-Computer Interface for classifying EEG correlates of chronic mental stress. In: Proceedings of International Joint Conference on Neural Networks, pp. 757–762 (2011)
2. Li, K., Rui, Q.: Classification of EEG Signals by Multi-Scale Filtering and PCA. In: IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS 2009, pp. 362–366 (2009)
3. Gorge, H.S.O.: A fast learning algorithm for deep belief nets. Neural Computation, 1527–1554 (2006)
4. Wu, S., Wu, W.: Common Spatial Pattern and Linear Discriminant Analysis for Motor Imagery Classification. In: 2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), pp. 146–151 (2013)
5. Yoshua, B., Pascal, L.: Greedy layer-wise training of deep networks. NIPS (2006)
6. Jarrett, K., Kavukcuoglu, K.: What is the best – stage architecture for object recognition. In: ICCV (2009)

7. Yohei, Y., Mitsukura, Y.: Hemodynamic characteristics for improvement of EEG-BCI performance. In: 2013 The 6th International Conference on Human System Interaction (HSI), pp. 495–500 (2013)
8. Yohei, T., Yasue, M.: Boosted Network Classifiers for Local Feature Selection. IEEE Transactions on Neural Networks and Learning Systems, 1767–1778 (2012)
9. Plamen, D., Jesse, S.: Classification of Imagined Motor Tasks for BCI. In: 2008 IEEE Region 5 Conference, pp. 1–6 (2008)
10. Karl, J.: Characterizing Functional Asymmetries with Brain Mapping, pp. 161–186. The MIT Press (2003)
11. Guger, C., Schlogl, C., Neuper, D., Walterspacher, T., Strein, G.: Rapid prototyping of an EEG-based brain computer interface (BCI). IEEE Trans. on Neural Systems and Rehabilitation Engineering, 49–58 (2001)
12. Wolpaw, J., Birbaumer, N.: Brian-computer interfaces for communication and control. Clinical Neurophysiology, 767–791 (2002)
13. Bengio, Y., Lecun, Y.: Scaling learning algorithms towards AI. Large-Scale Kernel Machines, 1–34 (2007)
14. Jonathan, R., Wolpaw, N., Birbaumer, D.: Brain-computer interfaces for communication and control. Clin. Neurophysiol., 767–791 (2002)
15. Kirkup, L., Searle, A.: EEG-based system for rapid on-off switching without prior learning. Medical and Biological Engineering and Computing, 504–509 (2007)
16. Hochberg, L., Serruya, M.: Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature, 164–170 (2006)
17. Cheng, M., Gao, X.: Design and implementation of a brain-computer interface with high transfer rates. IEEE Transactions on Biomedical Engineering, 1181–1186 (2002)
18. Shoker, L., Sanei, S.: Distinguishing between left and right finger movement from EEG using SVM. In: Engineering in Medicine and Biology 27th Annual Conference, pp. 5420–5423 (2005)
19. Ohkawa, Y., Suryanto, C.: Image set-based hand shape recognition using camera selection driven by multi-class Ada Boosting. Advances in Visual Computing, 555–566 (2011)
20. Shen, C., Li, H.: Boosting through optimization of margin distributions. IEEE Trans. Neural Netw. Learn. Syst., 659–666 (2010)