

Analyze EEG Signals with Convolutional Neural Network Based on Power Spectrum Feature Selection

Wenqiang Liu¹

*Brain Cognitive Computing Lab, School of Information Engineering, Minzu University of China
Beijing, 100081, China
Email: liuwqleo@163.com*

Huiping Jiang²³

*Brain Cognitive Computing Lab, School of Information Engineering, Minzu University of China
Beijing, 100081, China
Email: jianghp@muc.edu.cn*

Yao Lu⁴

*Brain Cognitive Computing Lab, School of Information Engineering, Minzu University of China
Beijing, 100081, China
Email: 5206661314@163.com*

Brain Computer Interface refers to the establishment of direct communication and control channels between human brain and computer. The essential part of the BCI system is the feature extraction and classification. In this paper, we proposed an improved architecture based on power spectrum and CNN. Furthermore, two methods of power spectrum feature extraction are discussed. The research result shows that this system can effectively identify the left-right brain signals and have higher classification accuracy with it reaching 82%.

*CENet2017
22-23 July 2017
Shanghai, China*

¹Speaker

²Corresponding Author

³Supported by National Natural Science Foundation of China Youth Project Funding (No: 61503423), the Leading Talent Program of State Ethnic Affairs Commission, and Double First-class Special Funding of MUC.

⁴The Second Author

1. Introduction

EEG (electroencephography) is a modern aided examination of the brain, which is used to record the brain's weak bioelectrical magnification in the form of a graph to help diagnose the disease without any trauma to the subject. Brain Computer Interface refers to establish a direct communication and control channel between human brain and computer. This can effectively enhance the physical disability of patients with the outside world or the ability to control the external environment, thereby improving the quality of life of patients. BCI (Brain Computer Interface) system is the core part of the EEG feature extraction part.

The combination of EEG and BCI highlights the advantages of its good time resolution, ease of use, and portability. But the technology is highly sensitive to noise, making it difficult to filter out the interference components of the data. At the same time the use of EEG as a brain interface requires a lot of training in order to ensure the normal operation of the equipment. So, the classification of EEG signal has been difficult and critical in the Brain Computer Interface development.

In the system designed by Torsten Felzer and Bernd Freislebe, each input signal is given a possible probability by the PEGNC (probability estimating guarded neural classifier) classifier so that if the probability is almost the same, the output will not make a judgment. They evaluated the classifier by experiment: by analyzing a rhythm signal in the EEG signal to determine whether the user is open or closed, and the classification result is as follows. The correct classification rate is 87%, and 12% can't be made to get a result, while only 1% is classification error [1]. In the system of Jose del R. and Josep Mourino, the user has three kinds of psychological work to distinguish. When the classifier can't clearly distinguish the user's current state, nothing could be done unless the clear judgment is out of the user's intention. So far, in their system, the system's correct rate is 70% or more, and the error rate is less than 5%, but the other between the range is unknown [2].

Deep learning is a branch of machine learning that attempts to model high-level abstraction of data through a series of algorithms. Deep learning is different from other machines, which adopt the traditional manual extraction of feature patterns, and makes use of unsupervised or semi-supervised extraction step by step. Therefore, deep learning modeling does not require prior knowledge, Multi-layer neural network can not only represent the nonlinear relationship between the data, but also achieve level extraction among the complex data.

In view of the huge size of the data collected through the BCI brain interface, the data collected in our experiment is 64×3000 . It is very difficult to extract from the actual training system, but it is easy to take up much resources. Considering the characteristics of the EEG signal, we combined the convolution neural network and the power spectrum to estimate the power spectrum of the signal. In the process of data processing and training, the accuracy rate is 82%.

2. Methods

Section 2.1 describes Power spectrum estimation. The Convolutional Neural Network (CNN) and EEG Signal Collection are explained in Section 2.2 and Section 2.3 respectively.

2.1 Power spectrum estimation

The AR model, also known as the autoregressive model, is an all-pole model that can be represented by Formula (2.1).

$$x(n) = \sum_{i=1}^n a(i)x(n-1) + u(n) \quad (2.1)$$

Where $u(n)$ is a white noise sequence with mean zero and variance σ^2 , p is the order of AR model, and a is the parameter of p -order AR model. The transferring function of the AR model system $H(z)$ is shown in Formula (2.2).

$$H(z) = \frac{1}{1 + \sum_{i=1}^p a_i z^{-i}} \quad (2.2)$$

Formula (2.3) for calculating the power spectrum of the AR model is obtained as follows:

$$P_x(k) = \sigma^2 \left| \frac{1}{1 + \sum_{i=1}^p a_i W^{-ki}} \right|^2 \quad (2.3)$$

It can be seen from the above equation that to use the AR model for power spectrum estimation, we must get the model parameters and white-noise sequence variance. The deformation will be described in Formula (2.4).

$$R_x(m) = \begin{cases} -\sum_{i=1}^p a_p(i)r_x(m-i), m \geq 1 \\ -\sum_{i=1}^p a_p(i)r_x(i) + \sigma^2, m = 0 \end{cases} \quad (2.4)$$

The formula AR Model AR Yule-Walker are described in Formula (2.5).

$$\begin{bmatrix} r_x(0) & r_x(1) & r_x(2) & \cdots & r_x(0) \\ r_x(1) & r_x(0) & r_x(1) & \cdots & r_x(1) \\ r_x(2) & r_x(1) & r_x(0) & \cdots & r_x(2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x(p) & r_x(p-1) & r_x(p-2) & \cdots & r_x(p) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (2.5)$$

2.2 CNN

CNN are hierarchical models whose convolutional layers alternate with sub-sampling layers, reminiscent of simple and complex cells in the primary visual cortex [3]. The network of architecture consists of three basic building blocks to be stacked and composed as needed. We have the convolutional layer, the max-pooling layer and the classification layer [4]. CNNs are among the most successful models for supervising image classification and setting the standard in many benchmarks [4, 5].

One feature of the convolution neural networks is the use of local joins and weight sharing to reduce the number of parameters that need to be trained. The convolution neural networks use local connections between adjacent two layers to take advantage of the local characteristics of the data matrix. As is shown in Figure 1, the neurons of each layer are only locally connected to the neurons of the previous layer. For example, the neurons of the layer are only connected to the local area of the neurons of layer L-1, and the width of the receptive field of the first layer is 3. That means, each neuron of the first layer is only within the three phases of the L-1 Adjacent neurons, and the connection between layer L + 1 and layer L has similar rules. It can be seen that the neurons of the L+ 1 layer are 3 with respect to the width of the receiving domain of the first layer, but their acceptance field relative to the L-1 layer is 5, and this structure will be limited to the filter inn the local space (since each neuron does not react to neurons outside of its receptive field), and the number of connections between neurons is reduced, and after multiple layers are stacked together, the filter is made globally, covering a larger area.

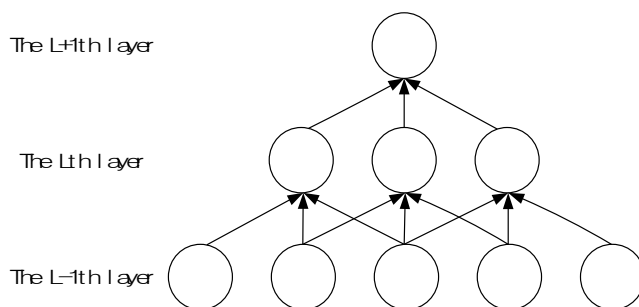


Figure 1 : Diagram of Local Connection

The weight sharing allows the neurons to share the same weights to detect the same feature at different locations, and organize the neurons that share the same weights into a two-dimensional plane to obtain the feature map. As is shown in Figure 2, the same color of the connection represents the same weight, if it is fully connected, the figure below the number of connection weights should be 3×5 , that is, 15, after the use of shared weights, the following figure of the number of connection weights is reduced to 3, reducing by 80%. So the use of weight sharing can greatly reduce the number of parameters that need to be trained. From the point of view of feature extraction, the local sensory field in the two-dimensional space can extract the primary visual features, such as the end point, the corner point and the edge of the specific angle from the two-dimensional image.

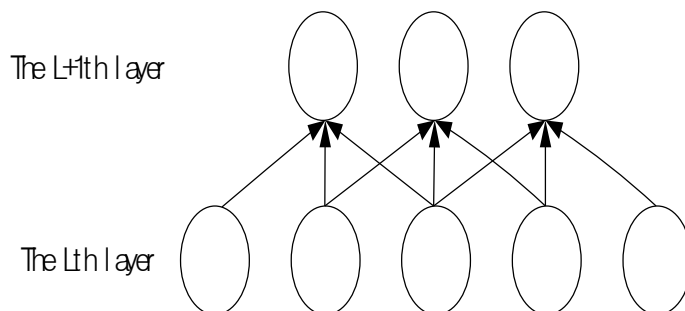


Figure 2 : Diagram of Sharing Weights

The subsequent layers can be obtained by combining these primary features High level and more abstract features. And the weight sharing makes the translation change in the input, in the feature map, and appear in the same direction and distance, so the use of weight sharing makes the convolution neural network have a translation without deformation.

The convolution neural network uses local connection and weight sharing to extract the local features of the input at each position of the input, and simulates the simple cells in the primary visual cortex of the primate.

2.3 EEG Signal Collection

The data used in this paper are collected from the EEG acquisition system shown in Figure 3, in which there include a stimulus system (STIM PC), a wet electrode cap, an electroencephalographic synchronous amplifier (Synamps), and EEG Receiving system (SCAN PC). The subjects wear an electrode cap, and the STM system generates an animated picture stimulus sequence, which is amplified by a Synamps amplifier connected to the electrode cap and converted to a digital signal to a PC that stores EEG data on a PC Acquire software as cnt-formatted files.

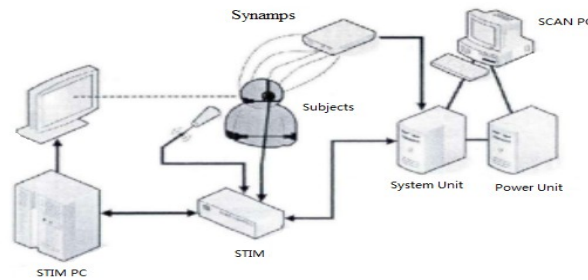


Figure 3 : Neuroscan EEG Acquisition Subsystem

In the EEG information collection process, we selected a total of 10 subjects, aged between 20 and 26 years old, normal vision, good health, without any neurological damage and mental history and sleep in good condition. Before the experiment, the subjects were informed of experimental purposes, the experimental process and experimental considerations. For example, they need to close the phone because other equipment may cause interference. Also, the subjects were told that the experiment will not cause any harm. Subjects sat in a better soundproof room with the animation reality in front of 50 cm on the LCD monitor. A total of 64 was obtained in the data collection.

The experiment to collect the EEG data is relatively weak and vulnerable to the interference of other signals. Because the signal noise caused by some factors such as eye movement during the experiment is not completely controlled, it is necessary to pretreat the EEG signal and remove the artifacts contained in the original signal so as to obtain a relatively reliable EEG signal.

3. Simulation

In EEG signal analysis, for a common random state signal with various states, it can't be described by a clear mathematical relationship, but the power spectral density of a stationary

random signal can be estimated by using a given N sample data, also known as power spectrum estimation. In fact, the length of the obtained random signal is always limited, with a finite length of the signal calculated power spectrum knowledge really an estimate of the power spectrum. The power spectrum estimation is divided into classical power spectrum estimation and modern power spectrum estimation. The classical power spectrum estimation is based on the Fourier transform, and the modern power spectrum estimation is based on the parametric model of stochastic process.

In this paper, the same test environment, the same data set, and different experimental process are set to compare the experiment. Because the EEG signal collected through the brain and brain interface is huge data and direct processing can not only increase the cost of the computer, but also be very time-consuming. Given the analysis of the experimental results are very unfavorable, we firstly use the power spectrum estimation data to extract the data Down dimension, and then follow-up the signal processing. In order to compare the output results, we use the correlation function method in the classical power spectrum estimation and the AR model in the modern power spectrum estimation to extract the power spectrum of the signal respectively. The original data 64×3000 is reduced to 64×15 , Network system, and finally through the Softmax classification to extract the results calculate the accuracy rate.

The original date 64×3000 are processed based on the standard dataset $64 \times 256 \times 15$. The filters found in the previous section are not only interesting but also biologically plausible. We now initialize a CNN with the same topology to fine-tuned for classification tasks This has already alleviated common problems with training Convolutional Neural Network [6].

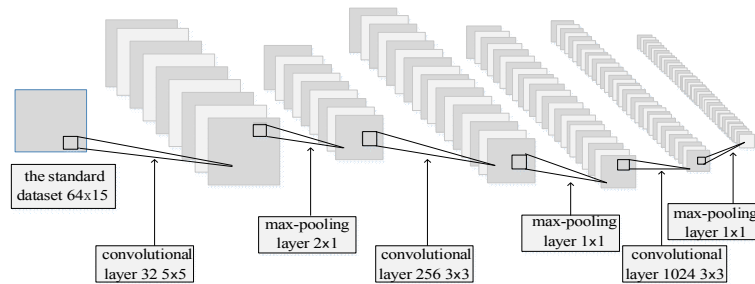


Figure 4 : Architecture of Convolutional Neural Network

We use the well labeled samples benchmark to show the effect of pre-training for subsets of defined sizes. The network has 8 hidden layers: 1) convolutional layer with 32 5×5 filters per input channel; 2) max-pooling layer of 2×1 ; 3) convolutional layer with 256 3×3 filters per map; 4) max-pooling layer of 1×1 ; 5) convolutional layer of 1024 maps of size 3×3 ; 6) max-pooling layer of 1×1 ; 7) a fully-connected layer of 4 hidden neurons; 8) a dropout layer. The output layer has a Softmax activation function with one neuron in two classes. The learning rete is annealed during training.

We use the two kinds databases, which are from Auto-relativity Function Method and Auto Regressive Model to input the convolutional neural network system, classification results on Auto-relativity Function Method by using defined subsets of the full data whose accuracy rate is 82%, and classification results on Auto Regressive Model whose accuracy rate is 42.8%.

The basis of classical spectral estimation is the Fourier transform. Because the Fourier transform domain is infinite, and the observed data are finite and the undetected data is actually forced to be treated as zero, which is equivalent to an infinite length of the sample with a rectangular window to be truncated. As there is still a strong correlation between the data

outside the window, the power spectrum estimated from the finite sample sequence is bound to produce a large deviation. The AR model is characterized by extrapolation of the correlation function, and the data is predicted from the observed data. AR parameter model method has the advantage of high frequency resolution, especially for short data processing, and smooth spectrum, which is conducive to automatic extraction and quantitative analysis of parameters. In the EEG signal spectrum estimation, it is more suitable for the segmentation spectrum estimation, and the processed signal linearity, smoothness and signal to noise ratio requirements are higher, so it is not suitable for long data EEG signal analysis and processing.

The comparison of these two methods of power spectrum, although the modern power spectrum estimation is not limited by the window, has the advantage of high resolution. However, in this experiment, the accuracy of the test is not classical power spectrum estimation method measured power spectrum, input to the convolution of the neural network to obtain the high accuracy.

4. Conclusion

In this paper, we have successfully presented the features of the EEG signal extracted by the connection of statistical process and CNN can be used to discriminate different kinds of EEG signals. At the first input EEG signal to power spectrum extraction, in the same test environment and the same data set, extraction method by using Autocorrelation Method and Burg Model, then the data input to the convolution neural network, and then the data processing and training. And it finally reached an accuracy rate up to 82%.

The waveform of EEG, as a kind of energy spectrum based on the analysis of power spectrum before extracting signal, can not only improve the experimental classification accuracy but boost the efficiency of experiments. The power spectrum and the combination of feature classification play more important roles in the EEG signal processing. Furthermore, the EEG signal analysis methods proposed in this paper may be used as auxiliary treatment of diseases and more researches will be conducted in this regard.

References

- [1] Hubel D H, Wiesel T N. *Receptive fields and functional architecture of monkey striate cortex*[J]. The Journal of physiology, 1968, 195(1): 215-243.
- [2] Millan J R, Mouriño J. *Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project*[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2003, 11(2): 159-161.
- [3] Hubel D H, Wiesel T N. *Receptive fields and functional architecture of monkey striate cortex*[J]. The Journal of physiology, 1968, 195(1): 215-243.
- [4] LeCun Y, Bottou L, Bengio Y, et al. *Gradient-based learning applied to document recognition*[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [5] Krizhevsky A, Hinton G. *Learning multiple layers of features from tiny images*[J]. 2009.
- [6] Erhan D, Bengio Y, Courville A, et al. *Why does unsupervised pre-training help deep learning?*[J]. Journal of Machine Learning Research, 2010, 11(Feb): 625-660.