ISSN: 2414-1895 DOI: 10.6919/ICJE.202104_7(4).0060

Classification of MEG Signals in Schizophrenia based on EEGNet

Jie Wu[1,a], Xiaoxia Huang[1,b]

[1]Shanghai maritime university, Shanghai 201306, China.

[a]954833020@qq.com, [b]xxhuang@shmtu.edu.cn

Abstract

Schizophrenia is a severe mental disorder, usually occurring in young adults, involving various mental disorders such as sensation, perception, thinking, emotion and behavior, as well as incoordination of mental activities. Clinically, the syndrome generally presents mixed symptoms, which are characterized by a series of positive and negative symptoms such as hearing voices, impaired cognitive function of speech or writing, delusions, hallucinations, apathy and behavioral withdrawal. The purpose of this paper is to classify MEG (Magnetoencephalography) signals of patients with schizophrenia and normal people by EEGNet. I first used the traditional mode: feature extraction with classifier. Under the method of (PE)permutation entropy and KNN (K-NearestNeighbor), the accuracy rate of the test set was 68.7500%. Under the method of permutation entropy and SVM (Support Vector Machine), the accuracy rate of the test set was 68.7500%. After that, the classification accuracy of raw date by EEGNet reached 73.463% in the test set. It can be seen that EEGNet can do more accurate classification on this data set.

Keywords

Schizophrenia; MEG; EEGNet; Permutation Entropy; SVM; KNN.

1. Introduction

Schizophrenia is a group of chronic diseases with unknown etiology. It usually starts slowly or subacute in young adults. Clinically, it is often manifested as a syndrome with different symptoms, involving sensory perception, thinking, emotion, behavior and other disorders as well as the incoordination of mental activities.

MEG is a functional brain imaging technology that studies brain magnetic field signals. It uses the superconducting quantum interferometer SQUID to measure the human brain noninvasively, and gets the weak magnetic field outside the head generated by the neural activity in the brain. MEG can reflect changes in neural activity on a millisecond scale and can be used to study dynamic brain behavior such as brain function to elicit stimulus responses. Cohen in the United States first recorded the magnetic field MEG outside the head in 1968. In the study of schizophrenia based on MEG signal, the time-domain and frequence-domain analysis methods in the traditional signal processing field are the earliest and most widely used. Koudabashi et al. studied the MEG signal cycle synchronization characteristics in response to light stimulation in schizophrenic patients, and found that the power spectrum of schizophrenic patients was lower than that of healthy people, which was consistent with similar studies on EEG signals in schizophrenic patients. Fehr on 28 patients with schizophrenia and 20 healthy controls the resting state of MEG signals combined with MSI (magnetic source imaging) is analyzed, studied the MEG signals of slow wave (in distribution and the position, the relationship between brain areas study found that patients with schizophrenia of slow-wave activity increased significantly in some regions, especially in the rear of the frontal lobe, temporal lobe and occipital lobe is more apparent, according local brain regions in the slow wave may be closely related to the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202104_7(4).0060

pathological characteristic of schizophrenia. However, there are few studies that combine MEG and CNN to classify schizophrenic patients.

EEGNet, a compact convolutional neural network for EEG-based BCI. This paper introduces the use of depth and detachable convolution to construct an EEG-specific model, which encapsulates the concept of EEG feature extraction common in brain-computer interfaces.

The main research of this paper is the comparison between the application of EEGNet in MEG data and the traditional mode of feature extraction + classifier.

And then the structure of the paper,The second part introduces Method and material, and the third part is Results and discussion. The last part is the conclusion. [1-5]

2. Method and material

The experimental data used in this paper are the resting state MEG data of schizophrenics and healthy subjects provided by the National Institute of mental health meg Core Facility in the United States. The subjects included patient group and control group: 12 patients with schizophrenia in the patient group; In the control group, 24 healthy subjects with no history of neurological or psychiatric disease were selected. All subjects were male and female, with an age range of 18 to 35 years, with an average age of 27.6 years. MEG signal data collection was conducted in a magnetic shielding room with Canadian VSMMedTech company CTF275 SQUID all the head type of superconducting quantum interferometer magnetoenc Ephalography does equipment measurement of magnetic field on perpendicular to the direction of the scalp. The MEG data packet collected by each subject contains 275 channels with a sampling frequency of 600Hz and a sampling time of 240s. The MEG signal is preprocessed, including artifact removal and filtering. The method used to remove artifacts is ICA[6].

2.1 ICA

Independent component analysis (ICA) is a signal analysis method based on high order statistics. ICA is a new signal processing method to solve the problem of blind source separation. In the case of unknown source signals, according to the statistical characteristics of the observed mixed signals, the statistically independent source signals can be separated. If X (t) = [x1 (t), x2 (t),..., x3 (t)] t m a observation vector signal, S (t) = [s1 (t) and s2 (t),..., sn (t)] for unknown number n independent non-gaussian source signals. The problem to be solved by ICA is to estimate A and S(t) when only the observed signal X(t) is obtained, but both the source signal S(t) and the mixed matrix A are unknown. The method used by ICA is to set a separation matrix W, making X(t) go through a series of transformations of the separation matrix W, and obtain the n-dimensional output column vector Y(t)=[y1(t), y2(t)... yn(t)],

$$\mathrm{Y(t) = WX(t) = WAS(t)}$$

Y (t) is the estimate of the source signal S(t). Therefore, the purpose of ICA is to seek the separation matrix W, so that after the transformation of the separation matrix W, the output results of the observed signals are as independent as possible, so as to reach the estimated independent source signals.

Both the mixed matrix A and the separated matrix W are unknown. If other unknown parameters are estimated only by the known X(t), there will be multiple solutions of the equation. Therefore, it is necessary to change the equation from multiple solutions to unique solutions under some constraint conditions,

$$\mathrm{Y(t) = WX(t) = WAS(t) = GS(t)}$$

Where, G is the global matrix. If G is (identity matrix), then Y(T)=S(T), and the independent source signal is obtained.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202104_7(4).0060

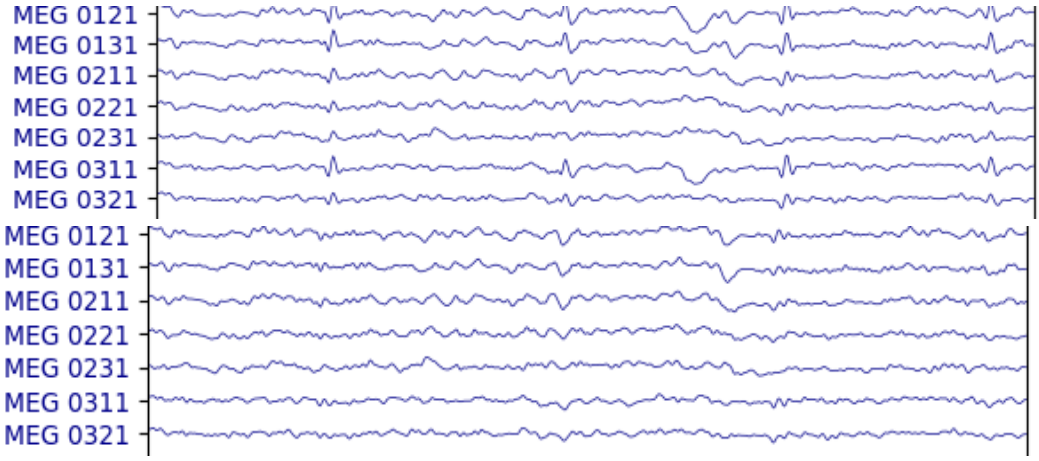

Figure 1. The top is before ICA and the bottom is after ICA

By comparing the ICA time domain waveform before and after the removal of artifacts, it can be seen that the ICA method can effectively achieve the removal of eye movement interference artifacts, ECG interference artifacts and some other artifacts, making the signal more stable.

2.2 Filtering and Z-score normalization

The MEG signal reflecting spontaneous neural activity of human brain in resting state is mainly its low frequency part. Therefore, Butterworth band-pass filter is used to filter the MEG signal with 0.5-30Hz after removing the artifact interference noise component, so as to obtain the required low frequency band.

A standard score, also known as a Z-score, is the difference between the score and the mean divided by the standard deviation.

The formula is: $z=(x-\mu)/\sigma$. Where x is a specific fraction, μ is the mean and σ is the standard deviation.

The z-value represents the distance between the original score and the maternal mean, and is measured in units of standard deviation. Z is negative when the original score is below the mean, and positive when the original score is below the mean.

2.3 Permutation entropy

Set a time sequence x = {x(i) | I = 1, 2,... N} get after the phase space reconstruction of phase space matrix is as follows:

$$\begin{pmatrix} x(1) & x(1+\tau) & \cdots & x(1+(m-1)\tau) \\ \vdots & \vdots & \vdots & \vdots \\ x(j) & x(j+\tau) & \cdots & x(j+(m-1)\tau) \\ \vdots & \vdots & \vdots & \vdots \\ x(K) & x(K+\tau) & \cdots & x(K+(m-1)\tau) \end{pmatrix}$$

m, τ is embedding dimension and delay time; K=n(m-1)τ. Each row in the matrix can be regarded as a reconstruction component, and there are K reconstruction components in total. Take the JTH component of the reconstruction matrix of X (i),(x(j),x(j+τ),…,x[j+(m-1)τ]), rearrange the values in ascending order of size, $j_1, j_2, \ldots, j_m$, represents the index of the column of each element in the refactored component

$$x[j+(j_1-1)\tau] \le x[i+(j_2-1)\tau] \le \ldots \le x[i+(j_m-1)\tau]$$

Therefore, for any time series X (I), a set of symbol sequences can be obtained for each row of the matrix reconstructed

$$S(l) = (j_1, j_2, \cdots, j_m)$$

l=1,2…,k, and k≤m!, The symbol sequence S(L) is one of these permutations. If the probability of the occurrence of each symbol sequence is calculated $P_1, P_2, \ldots, P_k$, According to Shannon's entropy,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202104_7(4).0060

the permutation entropy of k different symbol sequences of time series X (L) can be defined as $H_p(m) = \sum_{j=1}^{k} P_j \ln P_j$

2.4 The CNN Model

The CNN network EEGnet, which is suitable for EEG data, wants to make an experiment on MEG data to confirm whether it is feasible. EEGNnet is a three-layer convolutional small network suitable for small data sets. The original 240s data can be cut into 4s segments to ensure that it has some features. At the same time, the data volume is increased to transform EEGnet into a network suitable for MEG data, as shown in Table 1

Table 1. CNN structure, F1 is number of temporal filters, F2 is number of pointwise filters

	Input	Sizes of Convolution kernels	Output
Conv2D	273*2400	F1*1*64	F1*273*2400
Depthwise Convolution	F1*273*2400	(D*F1)*273*1	(D*F1)*1*2400
AveragePool	(D*F1)*1*2400	1*4	(D*F1)*1*600
Dropout		P=0.25	
SeparableConv2D	(D*F1)*1*600	F2*1*16	F2*1*600
AveragePool2D	F2*1*600	1*8	F2*1*75

This network uses the activation function of ELU, which has the positive characteristic like RELU, so it can also alleviate the gradient disappearance problem, while ELU also has a negative value, which can make the mean value of the output of the activation unit closer to 0, so as to achieve the regularization effect. And the negative value of ELU is an exponential function, which does not mutate, so it is more robust to input changes or noise. Depthwise Convolution significantly reduces the number of parameters in the convolution layer, but this method cannot extend the Feature map and fails to effectively utilize the Feature information of different channels in the same spatial location. Pointwise Convolution can expand the number of feature maps and combine the feature maps generated in the previous step in spatial dimension, which just makes up for two shortcomings of Depthwise Convolution.

3. Results and discussion

For the experiment on Raw Date, the data of 240s for a single person was cut into 4s and 60 for a single person. There were 25 normal people and 13 patients. It is divided into training set and data set. 22 normal subjects and 10 patients as the training set and the others as the test set. Therefore, the data volume is 1920 training sets and 360 test sets. The calculation formula of accuracy is

$$\text{Acc} = \frac{acc}{y}$$

ACC is the correctly predicted data, and Y is the number of test sets.

Table 2. Performance of CNN and SVM on the raw data

Acc	CNN	SVM
Training set	75.520%	56.000%
Test set	73.463%	50.000%

It can be seen from the above table that CNN established on RAW Data can well undertake the task of feature extraction and classification. In the test set of patients with schizophrenia, the accuracy rate is 73%, while SVM does not well distinguish between normal people and patients, only 50%, with almost no classification function.

The classification performance of data with permutation entropy in SVM and KNN is as follows,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202104_7(4).0060

Table 3. Performance of SVM and KNN after PE

Acc	PE+SVM	PE+KNN
Training set	83.3335%	78.3350%
Test set	56.2500%	68.7500%

It can be seen from the above table that combining PE as feature extraction, SVM and KNN as different classifiers are slightly inferior to the performance of CNN directly used in the original data in terms of classification results. After the two-fold cross validation, PE+SVM has more than 83% accuracy in the training set and 56.25% accuracy in the test set, while PE+KNN has 78% accuracy in the training set and 68.5% accuracy in the test set.

4. Conclusion

In terms of the original data, CNN has the function of feature extraction and classification, while the traditional classifier is unable to classify it, and the accuracy rate of 73% is also achieved. Compared with the traditional feature extraction + classifier classification effect is slightly better than them. At the same time, it also provides a research idea for novices who have not been exposed to feature extraction, such as researchers in the medical field. It is feasible to use CNN for direct classification on MEG data.

References

[1] Insel T R. Rethinking schizophrenia.[J]. Nature, 2010, 468(7321):187-193.

[2] Lawhern V J, Solon A J, Waytowich N R, et al. EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces[J]. Journal of Neural Engineering, 2016, 15(5):056013.1-056013.17.

[3] Koudabashi A, Fujimoto T, Takeuchi K, et al. Spatiotemporal characteristics of MEG and EEG entrainment with photic stimulation in schizophrenia.[C]// Conference: International Conference of the IEEE Engineering in Medicine & Biology Society IEEE Engineering in Medicine & Biology Society Conference. 2004:4465-8.

[4] Guo G, Wang H, Bell D, et al. KNN Model-Based Approach in Classification[C]// OTM Confederated International Conferences "On the Move to Meaningful Internet Systems". Springer, Berlin, Heidelberg, 2003.

[5] Burges C J C . A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2):121-167.

[6] Bingham E, Hyv?Rinen A. A Fast Fixed-Point Algorithm for Independent Component Analysis[J]. International Journal of Neural Systems, 2000, 10(01):1-8.