

Extraction and Analysis of Independent Components in Interictal and Preictal Data

Zhangyong Li^{1,a,*}, Xiaoqin Wan^{1,b}, Jing Zuo^{1,c}, Wei Wang^{1,d} and Bin Zheng^{1,e}

¹Research Center of Biomedical Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China

^a2224935165@qq.com, ^b983815492@qq.com, ^c1564458304@qq.com ^d58642660@qq.com, ^e1530868301@qq.com

*Corresponding Author

Keywords: Epilepsy, Time interval, Independent component analysis, Statistical analysis

Abstract. Epileptic seizures can cause changes in the composition of the brain. Therefore, the analysis of independent signals in EEG with epilepsy can help to understand the development of epilepsy. Every interictal and preictal EEG data which has a lasting acquisition time of 10 minutes is divided into segments at the time interval $T=20s$, $30s$, and $60s$ respectively. These segments are analyzed by FastICA after the preprocessing of wavelet transform, then the number of the corresponding data segments' independent components can be counted. Accordingly, the average number of independent components of the whole corresponding data can be obtained. By statistical analysis, it can be known that the number of independent components in preictal data is higher than that in interictal data. And in the preictal data of each period, the closer the data is to the seizure onset, the more the number of independent components will be. Under the condition of the same interictal and preictal data. The difference of the average number of independent components between the preictal and interictal data is the most obvious under the conditions of the same data when $T=30s$, which is $0\sim 0.7$ larger than that when $T=20s$ and $0.2\sim 0.8$ larger than that when $T=60s$. This study can provide a foundation for exploring the epilepsy.

1. Introduction

Epilepsy is a chronic disease of nervous system disorders[1]. According to statistics, about 1% of the world's people are suffering from epilepsy. The mechanism of epilepsy is very complex, which involves genetics, anatomy, pathophysiology, immunization and so on. Although there are some breakthroughs about the treatment of epilepsy, it still remains many problems to be solved[2,3]. Cognitive impairment and emotional disorders caused by epileptic seizures greatly reduce the quality of life of patients[4], epilepsy brings great damage to the patients and their families, and there are many self-injury cases because of epilepsy every year[5]. Currently, the treatment for most epilepsy patients is the use of high doses of antiepileptic drugs, although the symptoms have been certain alleviated, patients still frequently suffer from side effects. For 20-30% of epilepsy patients, antiepileptic drug is invalid[6]. Surgery can have some effect on certain

patients, but it has a irreversible damage to the brain[7]. Although the frequency of epileptic seizures is not very high, patients are always in the anxiety that epilepsy may occur at any time. The changes of the brain function caused by epilepsy can be showed by electroencephalogram, so the analysis and processing of epilepsy EEG signal is an important way to know about epilepsy.

Epilepsy can be divided into four states: interictal (between seizures, or baseline), preictal (prior to seizure), ictal (seizure), and post-ictal (after seizures). If the difference between interictal and preictal data can be recognized, it will help patients have timely treatments to reduce the risk of seizures. Figure 1 shows a period of interictal data and a period of preictal data. At present, there are two ways to distinguish between the interictal and preictal data, one way is extraction of the characteristics of the signal. Although it has a high accuracy, the process of data is much complex and has a long-time consuming[8]. Another way is to classify signals by support vector machines and neural networks[9,10]. However, the classification algorithms usually need a lot of data to train and test in order to ensure the accuracy of the algorithm. Fast independent component analysis (FastICA) is a kind of unmixing method for mixed signals which develops from ICA. Usually the measured EEG data are mixed signals, so FastICA can be used to separate EEG data. According to the study, the results of FastICA about the EEG data meet certain physiology[11].

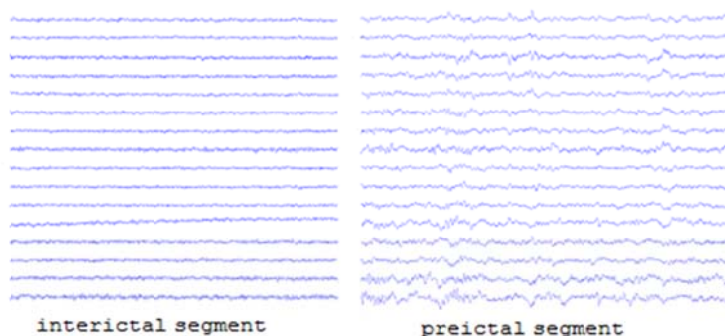


Figure 1 Interictal and preictal data segment.

2. Method

2.1. Data sources

The data is obtained from four dogs by implanting a 4×4 electrode array in their brains under the situation of spontaneous states (<https://www.kaggle.com/c/seizure-prediction>). It has a sampling frequency of 400 Hz. There are two types of data: interictal data and preictal data. The composition of the data is shown in Table 1 below:

Table 1 The composition of the data

name	I	P1	P2	P3	number of channels
Dog 1	45	4	4	4	16
Dog 2	45	7	7	7	16
Dog 3	45	12	12	12	16
Dog 4	45	5	5	5	15

Where the acquisition time of each data is 10 minutes. I represents the number of the interictal data; P1 represents the number of the preictal data which has a lasting time from 5 minutes to 15

minutes before epilepsy seizures, P2 represents the number of the preictal data which has a lasting time from 15 minutes to 25 minutes before epilepsy seizures; P3 represents the number of the preictal data which has a lasting time from 25 minutes to 35 minutes before epilepsy seizures;

2.2. Data preprocessing

In the acquisition of EEG signal, there will be some unavoidable interference which need to be removed as far as possible to ensure the accuracy of the signal analysis subsequently. Wavelet transform is widely used in signal denoising and feature extraction because of its multi-resolution and self-adaptability[12,13]. In the analysis of the interictal and preictal data, the most important thing is to identify whether there are spike and sharp waves in the data, The duration of spike wave is about 20~80ms. The duration of the sharp wave is about 80~200 ms. Therefore, the frequency range of the epileptic wave is 5~30 Hz. First the wavelet transform is used to decompose the signal into eight layers, then the detail coefficients of the corresponding frequency band are reconstructed. The characteristics of the epileptic wave after denoising become more prominent.

2.3. FastICA algorithm

The human body is a complex multi-source signal field which includes EEG, ECG, EMG and so on. In the process of signal transmission, there will be different degrees of coupling to these signals, and it couples into a part of the external interference at the same time. So the EEG signal is a multi-source mixed signal. However, the mixing way of these signals is usually unknown. Blind source separation is a kind of unmixing technique that arises to solve this problem. FastICA is a much mature algorithm developed from blind source technology. FastICA simplifies the problem and operation which greatly improves the work efficiency, it is widely used in the analysis and processing of EEG[14,15]. The structure between the interictal and preictal data are different which means that the source signal components are also different, so it will be possible to distinguish between the interictal and preictal data by FastICA. FastICA algorithm has many forms like basing on maximum likelihood estimation, negative entropy and so on. It is known that the more chaotic the structure of the signal is, the greater the entropy will be. The structure of the EEG signal will change significantly when the epilepsy occurs, so this study uses FastICA algorithm based on negative entropy. The steps of the algorithm are as follows:

Step one: remove the mean of the mixed signal X to simplify the algorithm.

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i, \quad X'_i = X_i - \bar{X} \quad (1)$$

Where N represents the number of signals, \bar{X} represents the mean of the signal, X'_i represents the variable after removing the mean.

Step two: white the signal to remove the correlation between the signals.

$$X'' = ED^{1/2}E^T X' \quad (2)$$

Where E represents a matrix of eigenvalues, D represents a matrix of eigenvectors, and X'' represents the result after whitening process.





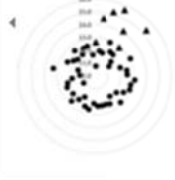




Step three: find the separation matrix W to get the estimated array Y.

3. Data analysis

The interictal and preictal data is divided into segments at the time interval T=20s, 30s, and 60s respectively. These segments are analyzed by FastICA after the preprocessing of wavelet transform, then the number of the corresponding data segments' independent components can be counted.

Accordingly, the average number of independent components of the whole corresponding data can be obtained. So the average number of each data (I, P1, P2, P3) can be known. Compare the difference between the interictal data (I) with every one in the preictal data (P1, P2, P3). The result of the analysis is showed as follows in Table 2, there takes dog 2 as an example:

Table 2 Compare of the average number of the independent components between (I) and (P1, P2, P3)

	T=20S	T=30S	T=60S
(I, P1)			
(I, P2)			
(I, P3)			

Where the triangle represents the average number of independent components of each preictal data in different periods (P1, P2, P3), the circular represents the average number of independent components of each interictal data (I).

From the Table 2, it can be known that the average number of independent components of the preictal data is generally higher than that of the interictal data. Under the condition of the same time interval, the difference phenomenon of the average number of independent components between (I) and (P1) is much more obvious than that between (I) and (P2). The difference phenomenon of the average number of independent components between (I) and (P2) is much more obvious than that between (I) and (P3). Under the condition of the same interictal and preictal data, when time interval T=20s, the difference phenomenon of the average number of independent components between the interictal and preictal data is much more obvious than that when T=60s. When time interval T=30s, the difference phenomenon of the average number of independent components between interictal and preictal data is much more obvious than that when T=20s.

In order to demonstrate the above results for the remaining three dogs is also applicable. Do the statistical analysis of the average number of the independent component for the 4 dogs. The significance level is taken as 0.01, the result is shown in Table 3:

Table 3 Statistical analysis of the average number of the independent component for the 4 dogs

name	T	P1 (mean±std)	P2 (mean±std)	P3 (mean±std)	I (mean±std)
dog1	20s	14.8±0.3*	14.8±0.4*	14.1±1.2	13.5±0.7
	30s	15.4±0.2*	15.3±0.3*	14.4±1.2	14.0±0.9
	60s	15.8±0.2	16.0±0.1	15.4±0.6	14.9±0.9
dog2	20s	13.9±1.2*	13.7±1.2*	13.1±1.2*	12.2±0.5
	30s	14.2±1.0*	14.4±1.2*	14.0±0.8*	12.3±0.5
	60s	15.1±1.0*	14.6±1.7*	14.3±1.3*	13.3±0.9
dog3	20s	13.2±0.9*	13.2±1.0*	13.3±0.7*	12.4±0.6
	30s	13.9±0.8*	13.9±1.1*	13.8±0.9*	12.5±0.6
	60s	14.6±1.1*	15.0±0.8*	14.6±0.7*	13.4±0.8
dog4	20s	13.0±0.7*	12.9±0.7*	13.1±1.0*	11.6±0.5
	30s	13.7±0.5*	13.4±0.4*	14.0±0.7*	12.0±0.6
	60s	14.5±0.5*	14.5±0.4*	14.4±0.5*	13.0±0.8

Where the ‘*’ indicates the case that the significance of the average number of the independent component between (p1, p2, p3) and (I) is less than 0.01.

It can be seen from table 3 that the average number of independent components of the preictal data (P1, P2, P3) is larger than the average number of independent components of the interval data (I) at any time intervals(20s,30s,60s).it also can be seen that there was no statistical significance between (I) and (P3) when T=60s in dog 1. Therefore, the significance between (I) and (P1) is much stronger than that between (I) and (P3),so do the significance between (I) and (P2) . And the significance when T = 20s and 30s is much stronger than that when T = 60s. In the case of the same T value, it can be seen that the average number of the independent components of P1 is generally greater than or equal to that of P2, and the variance of the independent components of P1 is lower than that of P2.This indicates that the difference phenomenon of the average number of independent components between(I) and (P1) is much more obvious than that between(I) and (P2).Under the condition of the same interictal and preictal data.it shows that the the average number of independent components when T=30s is larger than that when T=20s, and the the variance of independent components when T=30s is smaller than that when T=20s.

In order to better illustrate the degree of difference of the average number of independent components between the interval data and the preictal data, an evaluation index is defined here: the average degree of difference δ . It is shown below:

$$\delta = P_{mean} - I_{mean} \quad (3)$$

So the degree of difference of the average number of independent components between the interval data and the preictal data is shown in Table 4:

Table 4 The degree of difference of the average number of independent components between the interval data and the preictal data

name	T	δ_{p1-i}	δ_{p2-i}	δ_{p3-i}
dog1	20s	1.3	1.3	0.6
	30s	1.4	1.3	0.4
	60s	0.9	0.9	0.5
dog2	20s	1.7	1.5	1.1
	30s	1.9	1.8	1.7
	60s	1.5	1.3	1
dog3	20s	0.8	0.8	0.6
	30s	1.4	1.4	1.3
	60s	0.7	0.6	0.5
dog4	20s	1.4	1.3	1.2
	30s	1.7	1.4	1.3
	60s	1.2	1.2	1.1

In the case of satisfying the significance level less than 0.01, the difference of the $\delta(\delta_{p1-i}, \delta_{p2-i}, \delta_{p3-i})$ under the different T values is analyzed. The results are shown in Table 5 - 7.

Table 5 The difference of the δ_{p1-i} under the different T values

name	$\delta_{p1-i(T=30s)} - \delta_{p1-i(T=20s)}$	$\delta_{p1-i(T=20s)} - \delta_{p1-i(T=60s)}$	$\delta_{p1-i(T=30s)} - \delta_{p1-i(T=60s)}$
dog1	0.1	----	----
dog2	0.2	0.2	0.4
dog3	0.6	0.1	0.7
dog4	0.3	0.2	0.5

Table 6 The difference of the δ_{p2-i} under the different T values

name	$\delta_{p2-i(T=30s)} - \delta_{p2-i(T=20s)}$	$\delta_{p2-i(T=20s)} - \delta_{p2-i(T=60s)}$	$\delta_{p2-i(T=30s)} - \delta_{p2-i(T=60s)}$
dog1	0	----	----
dog2	0.3	0.2	0.5
dog3	0.6	0.2	0.8
dog4	0.1	0.1	0.2

Table 7 The difference of the δ_{p3-i} under the different T values

name	$\delta_{p3-i(T=30s)} - \delta_{p3-i(T=20s)}$	$\delta_{p3-i(T=20s)} - \delta_{p3-i(T=60s)}$	$\delta_{p3-i(T=30s)} - \delta_{p3-i(T=60s)}$
dog1	----	----	-----
dog2	0.6	0.1	0.6
dog3	0.7	0.1	0.8
dog4	0.1	0.1	0.2

Where "----" means that it does not meet the significance of the conditions of less than 0.01 at the T value, then the average difference between each other become meaningless.

From Table 5 to Table 7, it can be seen that .The difference of the average number of independent components between the preictal and interictal data when T=30s is 0~0.7 larger than that when T=20s.The difference of the average number of independent components between the preictal and interictal data when T=20s is 0.1~0.2 larger than that when T=60s.The difference of the average number of independent components between the preictal and interictal data when T=30s is 0.2~0.8 larger than that when T=60s.So the T division relationship which makes significant difference of the average number of independent components between the preictal and interictal data from large to small is 30s> 20s> 60s.

Next, the difference of the average number of independent components between each preictal data in different periods (p1, p2, p3) is also analyzed. The results are shown in Table 8:

Table 8 The difference of the average number of independent components between each preictal data in different periods (p1, p2, p3)

name	T	δ_{p1-p2}	δ_{p1-p3}	δ_{p2-p3}
dog1	20s	0	0.7	0.7
	30s	0.1	1	0.9
	60s	0	0.4	0.4
dog2	20s	0.2	0.6	0.4
	30s	0.1	0.2	0.1
	60s	0.2	0.5	0.3
dog3	20s	0	0.2	0.2
	30s	0	0.1	0.1
	60s	0.1	0.2	0.1
dog4	20s	0.1	0.2	0.1
	30s	0.3	0.4	0.1
	60s	0	0.1	0.1

As it can be seen from Table 8, $\delta_{p1-p3} = \delta_{p1-p2} + \delta_{p2-p3}$. When $T = 20s$, the average number of independent components of (p1) is 0 ~ 0.2 larger than that of (p2), and 0.2 ~ 0.7 larger than that of (p3). the average number of independent components of (p2) is 0.1~0.7 larger than that of (p3). When $T = 30s$, the average number of independent components of (p1) is 0~0.3 larger than that of (p2), and 0.1~1 larger than that of (p3). the average number of independent components of (p2) is 0.1~0.9 larger than that of (p3).When $T = 30s$, the average number of independent components of (p1) is 0~0.2 larger than that of (p2), and 0.1~0.5 larger than that of (p3). the average number of independent components of (p2) is 0.1~0.4 larger than that of (p3).So the order relationship of each preictal data in different periods whose the average number of independent components from large to small is $p1 > p2 > p3$.

4. Discussion

The higher the frequency of the EEG signal is , the smaller the phase synchronization index is [16]. The interictal data belongs to normal EEG data, and the preictal data contains epileptic waves, so the frequency of preictal data is higher than that of interictal data. and the phase synchronization index of the preictal data is smaller than that of interictal data. The closer to the epilepsy seizure, the more epileptic waves it will have, the phase synchronization will be closer to the lowest value. When the brain is in preictal state, the original balance of the brain is broken, the original synchronization law also has been broken, and the brain is in a state of confusion, This situation is equivalent to add some signal source which are constituted by epileptic wave into the normal EEG signal,so the number of independent components will increase. Meanwhile the closer to the epilepsy seizures, the more epileptic waves will contain in the data, so under the condition of the same time interval,the difference of the average number of independent components between (p1) and (I) is larger than that between (p2) and (I).the difference of the average number of independent components between (p2) and (I) is larger than that between (p3) and (I).In the analysis of FastICA , the number of observed signals is greater than or equal to the source signal. On the one hand,the longer the length of EEG data is ,the closer the number of independent components will be to the number of channels collected,so when the epilepsy data are divided into segments at the time interval of 60s,the number of independent components of both interictal and preictal data close to the number of channels.so the difference of the average number of independent components between interictal and preictal data when $T=60s$ is not good. On the other hand, the smaller the T value is,the number of independent components is not enough stable, and the operation takes a long time.So in summary, the difference of the average number of independent components between interictal and preictal data when time interval $T=30s$ is better than $T=20s$ or $60s$.

Epilepsy is a kind of chronic disease, if the treatment are delayed ,patients may suffer serious brain damage[17-18].the analysis of the epilepsy data can help people to know about the development of epilepsy in order to control and reduce the number of the epilepsy seizures,so that the patients and their families can live a more normal life.

Acknowledgments

I would like to thank the Natural Science Foundation of China (61571070) and the Chongqing Municipal Science and Technology Commission (No.cstc2015shms-zttx0098, No. cstc2015zdcy-zttx120002 and No.KJZH14208) for the support to this study.

References

- [1] Z Anwar. Epilepsy: Neurological Disorder-a Review[J]. Asian Journal of Medical Sciences, 2014, 6(3):25-29.
- [2] Sadowski K, Jóźwiak S. Recent advances in pathophysiology studies and treatment of epilepsy in neurocutaneous disorders[J]. Journal of Epileptology, 2015, 22(2):99-108.
- [3] Ponnusamy A. Recent Advances in Predicting and Preventing Epileptic Seizures - Seizure - European Journal of Epilepsy[J]. Seizure, 2014, 23(4):324.
- [4] Rahman Z, Wong C H, Dexter M, et al. Epilepsy in patients with primary brain tumors: The impact on mood, cognition, and HRQOL.[J]. Epilepsy & Behavior, 2015, 48:88-95.
- [5] Virta M, Pessi T M, Seiskari T, et al. Self-harm in people with epilepsy: a retrospective cohort study.[J]. Epilepsia, 2014, 55(9):1355–1365.
- [6] Schmidt D, Schachter S C. Drug treatment of epilepsy in adults[J]. Bmj, 2014, 348(9):254-254.
- [7] Schad A, Schindler K, Schelter B, et al. A pplication of a muhivariate seizure detection and prediction method to noninvasive and intracranial long-term EEG recordings[J]. Clin Neurophysiol, 2008, 119 (1): 197 -211.
- [8] Wen-Feng H U, Song J L, Zhang R, et al. Method of Feature Extraction of EEG[J]. Journal of Xian University, 2016.
- [9] Li X, Chen X, Yan Y, et al. Classification of EEG Signals Using a Multiple Kernel Learning Support Vector Machine[J]. Sensors, 2014, 14(7):12784-12802.
- [10] Djemili R, Bourouba H, Korba M C A. Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals[J]. Biocybernetics & Biomedical Engineering, 2015, 36(1):285-291.
- [11] Martín-Clemente R. Exploratory Analysis of Brain with ICA[M]// Blind Source Separation. Springer Berlin Heidelberg, 2014:435-463.
- [12] Cui H, Zhao R, Hou Y. Improved Threshold Denoising Method Based on Wavelet Transform[C]// International Conference on Modelling, Identification and Control. IEEE, 2015:1354–1359.
- [13] Chang H Y, Yang S C, Lan S H, et al. Epileptic seizure detection in grouped multi-channel EEG signal using ICA and wavelet transform[C]// IEEE International Symposium on Circuits and Systems. IEEE, 2010:1388 - 1391.
- [14] Zhao D, Jiang J, Wang C, et al. FPGA Implementation of FastICA Algorithm for On-line EEG Signal Separation[J]. Communications in Computer & Information Science, 2015, 491:59-68.
- [15] Zakeri Z, Asseondi S, Bagshaw A P, et al. Influence of Signal Preprocessing on ICA-Based EEG Decomposition[J]. IFMBE proceedings, 2014, 41:734-737.
- [16] Osorio I, Lai Y C. A phase-synchronization and random-matrix based approach to multichannel time-series analysis with application to epilepsy.[J]. Chaos, 2011, 21(3):1111-1117.
- [17] Grillo E. The circular dilemma of seizure-induced brain injury.[J]. Brain A Journal of Neurology, 2014, 137(11):305-305.
- [18] Grillo E. Postictal MRI abnormalities and seizure-induced brain injury: Notions to be challenged[J]. Epilepsy & Behavior, 2015, 44(3):195-199.