Epileptic EEG information entropy based on different entropy estimation methods

Min Li¹, Yunjie Fang², Yu Zhang³, Yiling Zhang¹, Junwei Chen³, Bingxin Zhu¹, Wei Yan⁴ and Jun Wang¹, ⁵

¹School of Geographic and Biological Information, Nanjing University of Posts and Telecommunications, Nanjing China
²Bell Honors School, Nanjing University of Posts and Telecommunications, Nanjing China
³School of Overseas Education, Nanjing University of Posts and Telecommunications, Nanjing China
⁴Department of Psychiatry, The Affiliated Brain Hospital of Nanjing Medical University, Nanjing China
⁵Email: wangj@njupt.edu.cn

Abstract. We use the information entropy based on linear model, K-nearest neighbour estimation and kernel estimation to study the brain, a complex nonlinear dynamic system, and distinguish the nonlinear dynamic complexity of epileptic and normal EEG signals in Bonn database. The entropy estimation method of linear model and K-nearest neighbour estimation can only distinguish the information entropy of epileptic period from that of other two cases, but it can't distinguish the information entropy of epileptic interval and normal EEG signals. However, kernel estimation can distinguish the information entropy of EEG signals in three states well, and the threshold range is [0.1,3]. With the increase of threshold, the discrimination effect is gradually significant until stable, and the discrimination effect is obvious when the threshold is 0.5. The result of analysis indicated that EEG information entropy was the highest in the epileptic seizure period, followed by the epileptic seizure interval, and the lowest in normal human brain.

1. Introduction

1.1. Background

EEG signal is a weak electrophysiological signal in human brain, which contains a large number of physiological and disease information of individuals. It can objectively reflect the state of the brain and provide diagnostic basis for some brain diseases. Epilepsy is a kind of short-term brain dysfunction syndrome harmful to health, which is characterized by recurrent epilepsy caused by abnormal discharge of brain neurons. In general, epilepsy and seizures are caused by abnormal electrical activity in the brain. Events like wrong wiring during brain development, physical injury or infection, may lead to seizures and epilepsy [1]. At present, electroencephalogram is used in clinical examination and diagnosis of epilepsy.

For the study of nonlinear dynamical systems, (such as correlation dimension [2], Lyapunov exponent [3] and nonlinear prediction method [4][5]), the entropy measurement tools based on
different estimation methods and obtained from information theory are also very popular. Compared with the former, the latter is a less comprehensive but more practical tool, including approximate entropy [6], sample entropy [7], correction condition entropy [8], fuzzy entropy [9], compression entropy [10], permutation entropy [11][12], distribution entropy [13], multiscale entropy [14][15], mutual information and information storage [16][17]. Brain is a typical nonlinear dynamic system. More and more theories and methods of entropy measurement has been applied to the study of EEG signal.

1.2. The basic principle of information entropy
It’s really difficult to analyze the complex dynamics of world systems. Information entropy, as the most basic measure of entropy in information theory, representing the instability of a system and the uncertainty of an event, is a static measure tool of a signal system. When calculating information entropy, only the amount of information that the signal is at that time is calculated. Compared with the entropy tool of dynamic calculation system predictability, information entropy is the simplest and easy to understand measure calculation tool. In this paper, information entropy is selected to extract the complexity of EEG signal in epileptic seizure interval, epileptic seizure period and normal EEG signal under different estimation methods. The difference between different EEG signals and the discrimination effect of different entropy estimation can provide the basis for clinical epilepsy detection and diagnosis.

To deduce information entropy, the concept of self-information put forward by Shannon should be firstly known. The self-information in the specific result \( v \) of the random variable \( V \) is:

\[
h(v) = -\log p(v)
\]

Where \( p(v) \) is the probability of \( V \) to \( v \). The unit of self-information depends on the cardinality of logarithm. If the base is 2, the unit is bit. When the base logarithm is used, the unit will be NAT. For the base 10 logarithm, the unit is hart.

Information entropy is the expectation of self-information of random event \( v \). It represents the measurement of uncertainty of an event, which is defined as:

\[
H(V) = E[h(v)] = -E[\log p(v)]
\]

When the value of \( p(v)p(v) \) is continuous, the formula of information entropy becomes:

\[
H(V) = -\int p(v)\log p(v)dv
\]

Where the integral is calculated in the range of continuous value \( A \). When the probability \( p(v) \) is discrete rather than continuous, information entropy is defined as:

\[
H(V) = -\sum_{v \in A} p(v)\log p(v)
\]

Where \( A \) in this case is the set of all values that \( V \) can take.

According to that the value of information entropy is the expectation of random event self-information, the information entropy reaches the maximum value when all symbols appear equality, that is to say, the uncertainty is the highest when all possible events have the same probability. On the contrary, if all observations of variables are the same, there is no uncertainty and the entropy of information is zero.

2. Entropy estimators
Various entropy estimators that follow different methods to calculate probability distribution can be obtained in the literature [18]. The estimators of entropy estimators can be divided into two groups: model-based estimators and modeless estimators. If the probability distribution of data can be
represented by a known parameter distribution (for example, Gauss) by definition, then a model-based estimator can be used as a function of the parameters of the assumed probability distribution to calculate the entropy measure [19]. When the data distribution cannot be assumed, we should follow the modeless method which approximates the probability distribution directly from the data. The most intuitive method is to establish the histogram distribution of quantized time series amplitude. However, this method has been proved to have serious deviation problems, and its estimation strongly depends on the size of quantization level [20][21]. The situation can be improved to some extent by using content-free density estimators, such as kernel estimator [22] or nearest neighbor estimator [23][24]. In this paper, we consider the estimation method based on linear model and two modeless methods using kernel and nearest neighbor entropy estimation.

2.1. Linear estimator

The linear estimator uses the model-based method to estimate the entropy measure, which assumes that the observation variables obey the Gaussian distribution. Under this premise, the Gaussian entropy process is solved as the entropy of the observed object. The assumed Gaussian probability distribution is described as follows:

\[
p(x_n) = (2\pi\sigma_x^2)^{-1/2} \exp\left(-\frac{x_n^2}{2\sigma_x^2}\right)
\]

Where \(\sigma_x^2\) is the variance of variance \(X\). Substituting equation (5) into equation (1), the entropy value of the current state is:

\[
E(X) = H(X_n) = \frac{1}{2} \ln 2\pi\sigma_x^2
\]

It should be noted that the information entropy of a static Gaussian process is only a function of its variance. When linear estimation is used, the assumption of random process obeys Gaussian distribution, which reduces the complexity of calculation process.

2.2. K-nearest neighbour estimator

K-nearest neighbor(KNN) estimator is a modeless method, the basic principle of which is: according to the distance function, calculate the distance between the past time state set and the present time state set of random process \(X\), then select the appropriate \(k\) as the neighborhood size of distance selection, and then calculate the number of neighborhood points according to the neighborhood size.

The KNN estimator approximates the probability distribution according to the statistics of the distance between adjacent points in the multidimensional space spanned by the observation variables. The average Shannon information content of general \(d\) - dimensional random variable \(V\) can be estimated from a set of implementation \(\{v_1, v_2, \ldots, v_N\}\). The formula is:

\[
-\mathbb{E}[\ln p(v_n)] = \psi(N) - \psi(k) + d\mathbb{E}[\ln \varepsilon_n]
\]

Where \(\psi\) is the digamma function derived from the gamma function, and \(\varepsilon_n\) is twice the distance(taking the maximum distance of the scalar component) between the result \(v_n\) calculated according to the maximum norm and its \(k\) nearest neighbor.

Combined with equation (1), it is easy to derive the expression of KNN estimation of the information entropy of the current state of the stochastic process \(X\) calculated for time series \(\{x_1, x_2, \ldots, x_N\}\):

\[
E(X) = H(X_n) = \psi(N) - \psi(K) + \langle \ln \varepsilon_n \rangle
\]
In the aspect of implementation, KNN program is easy to realize quickly, and has low error in fault tolerance and evaluation algorithm.

2.3. Kernel estimator
The kernel entropy estimator reconstructs the probability distribution of the observed variable by centering the kernel function at each result of the variable, and then uses the estimated probability to derive the relevant entropy measure. It is also a modeless approach. The kernel is used to weight the distance from each point to the reference point in the time series according to the kernel function. For example, if the random process \( X \) is determined, starting from the realization of the length \( N \) obtained in the form of time series \( \{x_1, x_2, ..., x_N\} \), the kernel estimation of the probability distribution can be calculated according to the following formula:

\[
p(X_n) = \frac{1}{N} \sum_{i=1}^{N} K(\|X_n - X_i\|)
\]

(9)

Where \( K \) is one of the kernel functions and \( \| \cdot \| \) in the formula is an approximate norm.

Substituting equation (9) into equation (1), the kernel estimation formula of information entropy can be obtained:

\[
E(X) = H(X_n) = -\ln \left( p(X_n) \right)
\]

(10)

In this paper, the kernel estimator is used to calculate the distance. For example, \( \|x_n - x_i\| \) is the most common measure, that is the so-called Chebyshev distance or the maximum norm. The formula is \( \|x_n - x_i\| = \max_{1 \leq k \leq m} |x_{n,k} - x_{i,k}| \).

In this paper, we select the most commonly used core function Heaviside, whose formula is:

\[
K = \Theta(\|X_n - X_i\|) = \begin{cases} 
1, & \|X_n - X_i\| \leq r \\
0, & \|X_n - X_i\| > r
\end{cases}
\]

(11)

The threshold \( r \) is the width of the Heaviside kernel function, which controls the accuracy of density estimation. Smaller value of \( r \) gives more detailed estimates, but more data points are required to be accurate. That is to say, if some points have large errors, the results will be affected; while too large value of \( r \) produces very rough probability estimates, which will reduce the computational complexity. At the moment, the accuracy of the estimates is low, because there are too many points in the neighborhood of the reference points.

3. Data process and analysis
The data used in this paper is from the Bonn database [25], which can be publicly available on the Internet. These data are widely used in the research of epilepsy. Data were collected by the Bonn Institute in Germany, which collected 500 single channel EEG records. 500 single channel EEG signals were divided into five groups (Z, O, N, F, S) with 100 EEG records in each group. Set Z and O include EEG records of cerebral cortex collected from five healthy volunteers. These data record two states of healthy volunteers, in which set Z is awake state and set O is asleep state. Set N, F and S are the EEG signals of five epileptics, of which set N and F are the EEG signals in the interictal and set S is the signals in the ictal. The duration of each record is 23.6 seconds. The sampling rate is 173.61 samples per second, so the length of each record is 173.61×23.6≈4097 samples [26], which is enough to realize the robust estimation of entropy. The waveforms of the five sets of EEG signals are shown in Figure 1.
When using information entropy for data processing, three different entropy estimation methods are used to quantify EEG signals, and then the results are averaged to eliminate accidental errors. The same operation is performed on the five sets respectively. Compare the information entropy of five groups of data and make a graph comparison, at the same time, change the variables in the estimation method to measure the stability of the estimation method.

3.1. Result of linear estimator

Using linear estimation to calculate the information entropy of data, draw the scatter diagram of three categories as shown in Figure 3, where the abscissa is the serial number (100 groups of data are arranged in chronological order). It can be seen from Figure 2 that the information entropy of epileptic patients in the seizure period (the asterisk in Figure 2) is the largest, and the information entropy of epileptic patients in the seizure-free period (the circle in Figure 2) and normal EEG (the square in Figure 2) coincides with each other.

![Figure 1. EEG time series of the five sets.](image)

![Figure 2. The results of the linear estimator.](image)
In order to make the effect of discrimination more intuitive, this paper compares three kinds of EEG signals, and the results are shown in Figure 3.

![Comparison results of three groups.](image)

**Figure 3.** Comparison results of three groups.

In distinguishing EEG signals in different states, the information entropy of linear estimation can correctly distinguish EEG signals in the seizure period and EEG signals of other two states, but it is not good in distinguishing between epileptic interval and normal EEG signals.

3.2. **Result of K-nearest neighbour estimator**

There are two variables in KNN estimation, namely the number of neighbors K and dimension. We fix the dimension value first to estimate the information entropy of five sets of signals. Under the fixed dimension value, change the number of neighbors to observe the change of discrimination effect. And then, change different dimension values, and compare the results under different dimension values to get the variable range of the best differentiation effect. About setting the variable range, the dimension value range is set to [2,5], the number of neighbors’ value range is set to [3,7]. The calculation results are shown in Figure 4.
According to the curve trend in Figure 4, it can be seen that the KNN estimation entropy of EEG signal in epileptic seizure period is the largest, which is consistent with the result of linear estimation. The difference of EEG signals between epileptic patients for interictal and normal individuals can't be distinguished. However, the information entropy of EEG signal in epilepsy interval is greater than that of normal EEG. At the same time, as the number of neighbor points increases, the information entropy of KNN estimation first decreases and then tends to be stable.

When changing the value of dimension, the KNN estimation information entropy has no significant improvement on the differentiation effect of EEG signals. We calculated the values of dimension 2, 3, 4 and 5 in order to observe the effect of differentiation. When the dimension is changed, the information entropy of KNN estimation does not change in the effect of distinguishing different EEG signals, and the trend is the same with the increase of the number of neighbor points. The computational complexity increases with the change of dimension. Therefore, when using, the dimension value should be the minimum value.

3.3. Result of kernel estimator

When kernel estimation calculates information entropy, there is a variable $r$ called threshold in kernel function. In the selection of threshold, too large threshold will lead to the decrease of accuracy in calculation, and too small threshold will lead to the error in estimating the probability of a certain time in the signal. In this paper, the range of threshold is selected in [0.1,3] after several trials, and calculate the points within the range of threshold to get the best threshold.

Figure 4. The results of the KNN estimator.
As shown in Figure 5, kernel estimation can distinguish the information entropy of EEG signals in three states well. Compared with linear estimation and KNN estimation, which is impossible to distinguish the information entropy of epileptic interval and normal EEG signals, kernel estimation shows superior performance. The information entropy of the kernel estimation of EEG signal in epileptic attack period is the largest, and the information entropy of EEG signal in interictal state is larger than that of normal EEG signal, though the difference between the latter two is not as obvious as that between the former two. In different states of healthy EEG, the value of information entropy in awake state is greater than that in asleep state. In the relationship between threshold and information entropy, the information entropy of kernel estimation first decreases and then tends to be stable with the increase of threshold, and the change trend of three EEG signals is same.

4. Conclusions
In this paper, we mainly study the information entropy of EEG signals based on linear estimation, KNN estimation and kernel estimation. In the analysis of the above results, it can be seen that linear estimation and KNN estimation can only distinguish the epileptic period from the other two states. And the result of the above two methods is that the information entropy of the attack period is greater than that of the other two states. Kernel estimation is better than the above two methods in the discrimination of EEG information entropy, which can distinguish the three states well. Under the method of kernel estimation, the information entropy of the kernel estimation of EEG signal in epileptic attack period is the largest, and the information entropy of EEG signal in interictal state is larger than that of normal EEG signal. It can be seen that the EEG information entropy can distinguish the different brain states and it can assist clinical diagnosis.

Acknowledgments
The project is supported by the National Natural Science Foundation of China (Grant Nos.31671006, 61771251), Jiangsu Provincial Key R&D Program (Social Development) (Grant No.BE2015700, BE2016773), Natural Science Research Major Program in Universities of Jiangsu Province (Grant No.16KJA310002), Innovative training program for College Students (Grant No. SZDG2019032).

References
[1] Dojnow P 2007 Multifractal analysis of correlation properties of electroencephalograms (EEG) C. R. Acad. Bulg. Sci. 60(6) 607-12
[2] Farmer J and Sidorowich J 1987 Predicting chaotic time series Phys. Rev. Lett. 59(8) 845
[3] Casdagli M 1989 Nonlinear prediction of chaotic time series Phys. D. 35(3) 335-56
[4] Porta A, Baselli G, Liberati D, Cogliati C, Gnegchi-Ruscone T, Malliani A and Cerutti S 1998 Measuring regularity by means of a corrected conditional entropy in sympathetic outflow Biol. Cybern. 78(1) 71-8
[5] Kurths J, Voss A, Saparin P, Witt A, Kleiner H J, Wessel N 1995 Quantitative analysis of heart rate variability Chaos: An Interdisciplinary Journal of Nonlinear Science. 5(1) 88-94
[6] Pincus S M 1991 Approximate entropy as a measure of system complexity Proc. Natl. Acad. Sci. 88(6) 2297-301
[7] Richman J S and Moorman J R 2000 Physiological time-series analysis using approximate entropy and sample entropy AM. J. PHYSIOL-HEART. C. 278(6) H2039-49
[8] Porta A, Baselli G, Liberati D, Montano N, Cogliati C, Gnegchi-Ruscone T, Malliani A and Cerutti S 1998 Measuring regularity by means of a corrected conditional entropy in sympathetic outflow Biol. Cybern. 78(1) 71-8
[9] Chen W, Wang Z, Xie H and Yu W 2007 Characterization of Surface EMG Signal Based on Fuzzy Entropy IEEE. T. NEUR. SYS. REH. 15(2) 266
[10] Truebner S, Cygankiewicz I, Schroeder R, Baumert M, Vallverdu M, Caminal P, Vazquez R, de Luna A B and Voss A 2006 Compression entropy contributes to risk stratification in patients with cardiomyopathy/Kompressionsentropie zur verbesserten Risikostratifizierung bei Patienten mit DCM BIOMED. ENG-BIOMED. TE.. 51(2) 77-82
[11] Bandt C and Pompe B 2002 Permutation entropy: a natural complexity measure for time series Phys. Rev. Lett. 88(17) 174102
[12] Riedl M, Müller A and Wessel N 2013 Practical considerations of permutation entropy EUR. PHYS. J-SPEC. TOP. 222(2) 249-62
[13] Li P, Liu C, Li K, Zheng D, Liu C and Hou Y 2015 Assessing the complexity of short-term heartbeat interval series by distribution entropy MED. BIOL. ENG. COMPUT. 53(1) 77-87
[14] Costa M, Goldberger A L and Peng C K 2007 Multiscale Entropy Analysis of Complex Physiologic Time Series Physical Review Letters. 89(6) 705-8
[15] Angelini L, Maestri R, Marinazzo D, Nitti L, Pellicoro M, Pinna G D, Stramaglia S and Tupputi S A 2007 Multiscale analysis of short term heart beat interval, arterial blood pressure, and instantaneous lung volume time series ARTIF INTELL MED. 41(3) 237-50
[16] Costa M, Goldberger A L and Peng C K 2005 Multiscale entropy analysis of biological signals Phys. Rev. E. 71(2) 021906
[17] Faes L, Porta A and Nollo G 2015 Information decomposition in bivariate systems: theory and application to cardiorespiratory dynamics ENTROPY-SWITZ. 17(1) 277-303
[18] Grassberger P and Procaccia I 1984 Physica. D. 13(1-2) 34-54
[19] Barnett L, Barrett AB and Seth AK 2009 Phys. Rev. Lett. 103(23) 238701
[20] Victor JD 2002 Nov Phys. Rev. E 66(1) 051903
[21] Panzeri S, Senatore R, Montemurro M A and Petersen R S 2007 J. Neurophysiol. 98(3) 1064-72
[22] Kaiser A and Schreiber T 2002 Physica. D. 166(1-2) 43-62
[23] Kozachenko L F and Leonenko N N 1987 Sample estimate of the entropy of a random vector Probi Peredachi. Inf. 23(23) 9-16
[24] Kraskov A, Stögbauer H and Grassberger P 2004 Estimating mutual information Phys. Rev. E. 69(6) 066138
[25] EEG time series from Bonn database[DB/OL]. http://epileptologie-bonn.de
[26] Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P and Elger CE 2001 Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state Phys. Rev. E. 64(6) 061907