ISSN : 0974 - 7435 Volume 10 Issue 18

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(18), 2014 [10224-10229]

A novel method for nonlinear detection of biomedical signal based on fuzzy entropy

Cao Rui, Wang Huiqing, Deng Hongxia, Li Conggai, Chen Junjie*
College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030024, (CHINA)
E-mail : chenjunjie_tyut@sina.com

ABSTRACT

The nonlinearity of biomedical signals time series is detected by surrogate method. However, the traditional statistics in surrogate method, such as correlation dimension (D2) and approximate entropy (ApEn), have some insufficiency in application, especially lower time efficiency. To solve these deficiencies, this study presents the fuzzy entropy (FuzzyEn) as a statistics of the surrogate method to detect the nonlinearity of time series and verify that in two simulation datasets. It was found that, for various lengths of time series, the new method can accurately detect the linearity or nonlinearity of them, and perform much better in time efficiency compared with traditional statistics. The results show that, the method presented in this article is an accurate, effective method to detect the nonlinearity of the biomedical signal.

KEYWORDS

Biomedical signal; Surrogate method; Fuzzy entropy; Nonlinearity; Nonlinear detection.

INTRODUCTION

In recent years, the nonlinear analysis method has become a quite active branch in the field of biomedical signal time series processing. Before making nonlinear analysis of the time series, a necessary prerequisite is to make sure whether the time series contains nonlinearity components, which is nonlinear detection. The surrogate method is a good method to detect whether a time series contains nonlinearity, and it is also the main way of nonlinear detection[1,14]

The basic idea of the surrogate method is to specify the null hypothesis firstly (assuming that the raw time series is linear), then randomly generate a series of linear surrogate data through disrupting order or disrupting phase method. It should be noted that, these surrogate data have the same linear nature with original data (such as mean, standard deviation, or spectrum, etc.). And then to calculate a statistics of the original data and these surrogate data, if the statistics values of the original and surrogate data have no significant difference, which means that the null hypothesis is faithful, then the original data is linear; if there are significant differences in the statistics values, which means that the null hypothesis is not faithful, then the original data is nonlinear.

With the application of the surrogate method, the researchers have studied and improved the generating process of the surrogate data[7,8,13,16,22,23]. For example, the surrogate data is generated by AAFT (amplitude adjusted Fourier transform) instead of FFT (Fast Fourier transform) when the original data obey non-normal distribution, moreover, the data with length of N needs to be made end correction to eliminate the end of the false high-frequency components. There are a lot of statistics to verify the difference of the original data and the surrogate data, the commonly used index is correlation dimension[11,21], approximate entropy[15,20], etc. However, the calculation of the correlation dimension needs longer data, and the time complexity is too high[10,12], there are certain difficulties in the practical application. There are some errors in the approximate entropy algorithm itself and the values of entropy with different parameters relative lack consistency[2,3]. In recent years, the application of the surrogate method has been a lot, but the study and improvement of the algorithm is relatively little. Therefore, in this paper, for the deficiencies of the statistics of the surrogate method, the fuzzy entropy is used as the test statistics to detect the nonlinearity of time series. Then this method is verified on two simulation datasets and one EEG dataset.

METERIAL AND METHODS

Surrogate method

Surrogate method is composed of null hypothesis and test statistics. Null hypothesis (assuming the original data is linear) is realized through the surrogate data generated by certain manner and the surrogate data has the same linearity with the original data. In this paper, we use the AAFT method to generate surrogate data. For the data with length of N, make end point correction (i.e. x (N-1) ≈ x (0)) to eliminate the end of the false high-frequency components. Through above procedure, we generate several surrogate data from each set of original data. Those surrogate data have the same spectrum with the original data, and the surrogate data is linear.

Fuzzy entropy

We calculate the statistics values of the surrogate data and the original data after obtaining several surrogate data. This study uses Fuzzy entropy as the test statistics proposed by Chen[26,27]. In this algorithm, vectors' similarity is defined by fuzzy similarity degree based on fuzzy membership functions, not the Heaviside function, making sure the entropy is continuous. Meanwhile, this algorithm inherits the merits of the approximate entropy, also has the characteristics that the entropy value changes little with different parameters, and strong anti-noise, which is more suitable than the approximate entropy to measure the complexity of the time series[27]. Therefore, this study will use the fuzzy entropy as the test statistics in the surrogate method, to detect whether the values of fuzzy entropy of the original data and the surrogate data have significant differences, and then verify whether the original data include nonlinearity.

The criterion of significant difference degree

To test whether the fuzzy entropy values of the original and the surrogate have significant difference, a reasonable approach is to take the mean value of fuzzy entropy values of a number of surrogate to compare with the fuzzy entropy value of the original data, and through the method of statistical test to determine whether there is a significant difference between the two.

Let E_o be the FuzzyEn of the original data, and let E_s and σ_s be the mean and SD values of the FuzzyEn values for the surrogate, respectively. A measure of statistical significance S was computed as follows:

$$S = \frac{| E_o - E_s |}{\sigma_s} \quad (1)$$

In this paper, 20 groups of surrogate were generated to match each original data. This statistic follows a Student's t-test distribution with 19 degrees of freedom. At the 0.95 level of significance, the critical value of t is 2.093. Accordingly,

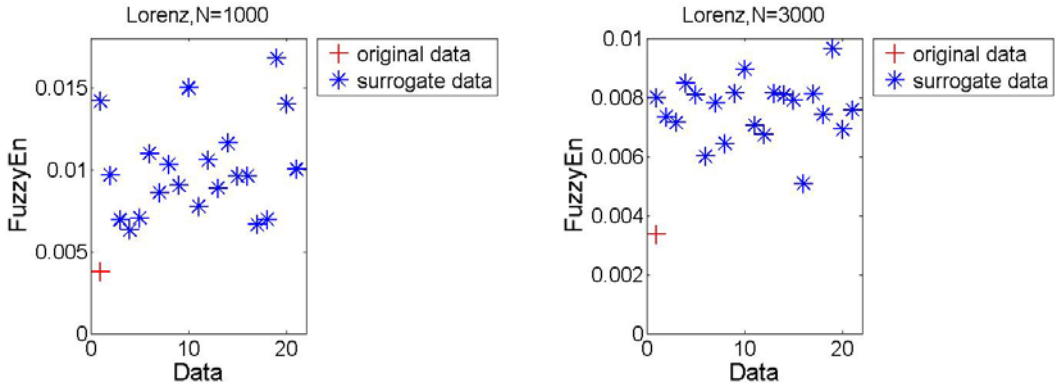
when S is larger than 2.093, the null hypothesis is rejected at the 95% probability level, and the original data are considered to contain non-linear dynamical features, and vice versa.

SIMULATION EXPERIMENTS RESULTS

The equations of Lorenz is used to generate nonlinear time series, while linear AR model is used to generate linear time series, and both of data are used to verify whether the fuzzy entropy can be used as a statistics to correctly identify the nonlinearity or linearity. Referring to the study of Chen[26], in this study, when calculating the fuzzy entropy, the parameters follow: *m=2, n=2, r=0.3SD* (SD is the standard deviation of time series), the data length N is from 100 to 3000.

Nonlinear simulation data and results

Nonlinear simulation data is generated by Lorenz equations when the parameters are σ = 10, r = 28, b = 8/3. To insure randomness of the data in the calculation process, generate a set of data with the length of 5000 first, then randomly select 10 times as the original data with each the length of 100, 200, 300, 500, 800, 1000, 2000, 3000, finally generate 20 surrogate data sets for each original data and calculate FuzzyEn of all data. Figure 1 shows the distribution of the FuzzyEn of the first original data and the corresponding 20 surrogate data, when the data length N=1000, 3000. Other results are similar. From the Figure we can see that, entropy of the original data and the surrogate data have significant difference for this time, but whether the result achieve the significant differences, it needs the results of subsequent statistical tests to prove.



(a) N=1000 (b) N=3000

Figure 1 : The fuzzy entropy of the original data and the surrogate data

S (Significant difference degree) between the FuzzyEn of original data and surrogate data also can be calculated by formula (10) in section 2.3. For example, when N=100, there are 10 original data with this length, so S_1,S_2…S_{10} can be acquired. S_i is the significant difference degree of the i*th* original data and the corresponding 20 surrogate data. Then calculate the mean value of S_1, S_2.. S_{10}, and get the average significance difference degree $\overline{S}$ when N is 100. Repeat the calculation process above, $\overline{S}$ of each length, i.e. 100,200, 300, 500, 800, 1000, 2000, 3000, can be acquired and shown in TABLE 1.

TABLE 1 : The results of nonlinear simulation data

Data length(N)	average significant difference ($\overline{S}$)	Data length (N)	average significant difference ($\overline{S}$)
100	2.1224	800	2.2647
200	2.1905	1000	2.2954
300	2.1110	2000	3.0960
500	2.6248	3000	4.0473

TABLE 1 shows that, to various length of Lorenz data, when the fuzzy entropy used as the statistics, significant difference degree of the original data and the surrogate data is greater than 2.093, means a rejection of the null hypothesis. So the original data is identified as nonlinear, which is consistent with the known characteristic (nonlinear) of Lorenz data. It means this algorithm can stably and accurately detect the nonlinearity of the data.

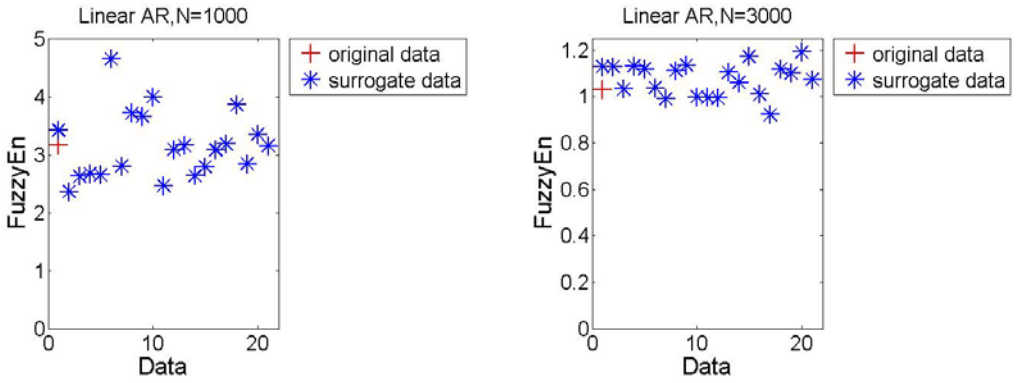
Linear simulation data and experimental results

In this study, we use the autoregressive AR model to generate linear data, which is calculated by formula (2)

$$y(t) = ky(t-1) + e(t) \quad (2)$$

We take the k=0.2 in this study. e (t) represents the white noise and the distribution of e(t) is that: the mean is 0, the variance is 1.Where $y(1) = 0 + e(1)$.

The data processing is the same as the previous one. Figure 2 (a) and Figure 2 (b) show the distribution of the fuzzy entropy of the first original data and the corresponding 20 surrogate data when the data length N=1000, 3000.



(a) N=1000 (b) N=3000

Figure 2 : The fuzzy entropy of the original data and the surrogate data

The results of the linear AR data are shown in TABLE 2. From TABLE 2 we can see that, when the fuzzy entropy used as the statistics, the linear AR data with different length, the significant difference of the original data and the surrogate data is less than 2.093, that the null hypothesis is accepted. So the original data is identified as linear, which is consistent with the known characteristic (nonlinear) of linear AR data. It means this algorithm can stably and accurately detect the linearity of the data.

TABLE 2 : The results of linear simulation data

Data length(N)	average significant difference ($\overline{S}$)	Data length (N)	average significant difference ($\overline{S}$)
100	1.3361	800	1.3361
200	0.1498	1000	0.1498
300	1.5224	2000	1.5224
500	1.1529	3000	1.1529

Comparison of the algorithm running time

The nonlinear verification algorithm based on fuzzy entropy is stable and accurate, and the running time also has higher efficiency than correlation dimension and approximate entropy. This paper selects Lorenz data at different length, run and record the running time of three nonlinear verification algorithms based on correlation dimension, approximate entropy and fuzzy entropy respectively. Every length of data is run 10 times, getting the average running time showed in TABLE 3.

TABLE 3 : Average running time of algorithms (second)

Data Length(N)	100	200	500	1000	2000	3000
D_2	75.76	193.62	1628.59	7452.81	33887.17	80858.38
ApEn	21.28	80.10	517.25	1962.26	8298.24	18613.66
FuzzyEn	2.29	4.00	16.19	55.65	200.31	434.79

This result shows that the algorithm raised in this paper substantially improve the time efficiency compared to the usual algorithm.

EEG data and experiment

The results of the simulation experiments above show that the algorithm raised in this paper can accurately and efficiently detect the nonlinearity of the time series. This study applies this algorithm to a group of open-source EEG data.

The database was accessed online at http://kdd.ics.uci. edu/ in January2009. The study randomly selects the EEG data of 10 control subjects which are numbered as co2c0000339、co2c0000340…. The result of the nonlinear detection is showed in Figure 3.

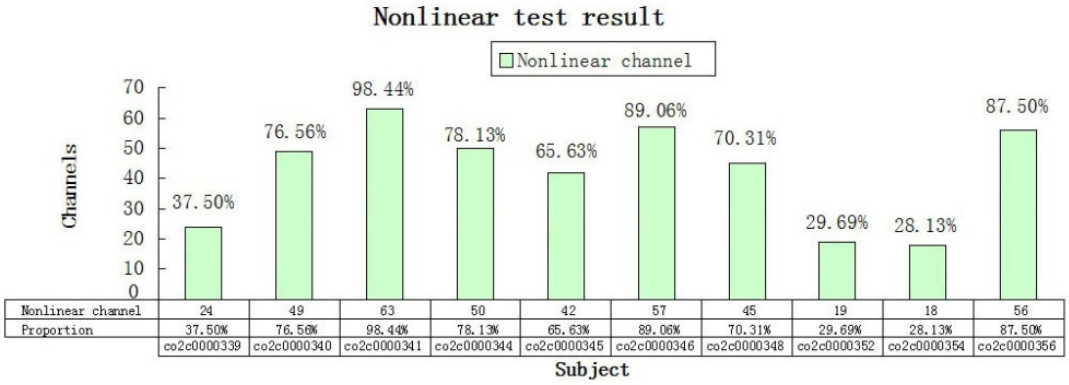


Figure 3 : Nonlinear detection result of EEG data

We found that in the 64 channel EEG of each subject, some channels show out the nonlinearity, while some show out the linearity. It give evidence of that there're nonlinear components in EEG. Therefore, the traditional way that merely used linear method or nonlinear method to study the EEG may lead to some lose of valuable information. The better way to study EEG is making a combination using of the linear and nonlinear algorithms[19,25].

CONCLUSION

Surrogate method has been much applied in the study of human physiological signals such as EEG[28], EMG[4 ,5], ECG[17,24], HRV[6,9] etc, as well as other data. The fuzzy entropy used as the test statistic in this paper, has overcome the shortcomings and deficiencies of the correlation dimension and approximate entropy. The experimental results of two simulation datasets show that, for the time series with different lengths (even for the time series with shorter length), the method raised in this study can be accurately and effectively detect whether the data contains nonlinearity, and can substantially improve the time efficiency. At last, this paper applies this algorithm to the EEG data, finding the nonlinear components in EEG. This study can be concluded: the surrogate method based on fuzzy entropy is a stable and effective nonlinear detection method. The conclusion can provide a better significance for the processing of biomedical signals such as ECG, EEG and heart rate, etc.

ACKNOWLEDGEMENT

This research is supported financially by the National Natural Science Foundation of China (NO. 61170136, NO.61373101) and the Research Fund for the Doctoral Program of Higher Education of China (NO.20131402120009).

REFERENCES

[1] A.K.Charakopoulos, T.E.Karakasidis, P.N.Papanicolaou, A.Liakopoulos; Nonlinear time series analysis and clustering for jet axis identification in vertical turbulent heated jets. *Physical Review E*. **89(3)**, 032913 **(2014)**.

[2] Cao Rui, Li Li, Chen Junjie; Comparative Study of Approximate Entropy and Sample Entropy in EEG Data Analysis. *BioTechnology: An Indian Journal.* **7(11)**, 493-498 **(2013)**.

[3] Chengyu Liu, Ke Li, Lina Zhao, Feng Liu, Dingchang Zheng, Changchun Liu, Shutang Liu; Analysis of heart rate variability using fuzzy measure entropy. *Computers in Biology and Medicine*. **43(2)**, 100-108 **(2013)**.

[4] A.Diab, M.Hassan, C.Marque, B.Karlssonb; Quantitative performance analysis of four methods of evaluating signal nonlinearity: Application to uterine EMG signals. *33rd Annual International Conference of the IEEE EMBC*. 1045-1048 **(2012)**.

[5] A.Diab, M.Hassan, C.Marque, B.Karlssonb; Performance analysis of four nonlinearity analysis methods using a model with variable complexity and application to uterine EMG signals. *Medical Engineering and Physics*. 3 **(2014)**. (In Press)

[6] Diego Guarin, Edilson Delgado, Alvaro Orozco; Testing for nonlinearity in non-stationary physiological time series. *33rd Annual International Conference of the IEEE EMBC*. **(2011)**.

[7] J.H.Lucio, R.Valdes, L.R.Rodriguez; Improvements to surrogate data methods for nonstationary time series. *Physica Review E*. **85(5)**, 056202 **(2012)**.

[8] C.J.Keylock; A wavelet-based method for surrogate data generation. *Physica D: Nonlinear Phenomena*. **225(2)**, 219-228 **(2007)**..

[9] Ko Keun Kim, Hyun Jae Baek, Yong Gyu Lim, Kwang Suk Park; Effect of missing RR-interval data on nonlinear heart rate variability analysis. *Computer Methods and Programs in Biomedicine*. **106(3)**, 210-218 **(2012)**.

[10] K.P.Harikrishnan, G.Ambika, R.Misra; Characterizing chaos-hyperchaos transition using correlation dimension. *The European Physical Journal Special Topics*. **222(3-4)**, 839-846 **(2013)**.

[11] R.Kunhimangalam, P.K.Joseph, O.K.Sujith; Nonlinear analysis of EEG signals: Surrogate data analysis. *IRBM*. **29(4)**, 239-244 **(2008)**.

[12] Lucas Lacasa, Jesus Gomez-Gardenes; Correlation Dimension of Complex Networks. *Physical Review Letters*. **110**, 168703 **(2013)**.

[13] M.McCullough, A.Kareem; Testing Stationarity with Wavelet-Based Surrogates. *Journal of Engineering Mechanics-ASCE*. **139(2)**, 200-209 **(2013)**.

[14] Min Gan, Yun-Zhi Huang, Ming Ding, Xue-ping Dong, Jiang-Bei Peng; Testing for nonlinearity in solar radiation time series by a fast surrogate data test method. *Solar Energy*. **86(9)**, 2893-2896 **(2012)**.

[15] H.Ocak; Automatic detection of epileptic seizures in EEG using discrete wavelet transforms and approximates entropy. *Expert Systems with Applications*. **36(2)**, 2027-2036 **(2009)**.

[16] Ralf Engbert; Testing for nonlinearity: the role of surrogate data. *Chaos, Solitons and Fractals*. **13(1)**, 79-84 **(2002)**.

[17] Raul Alcaraz, J.Jose Rieta; Nonlinear synchronization assessment between atrial and ventricular activations series from the surface ECG in atrial fibrillation. *Biomedical Signal Processing and Control*. **8(6)**, 1000-1007 **(2013)**.

[18] R.G.Abeysuriyaa, C.J.Rennie, P.A.Robinson, J.W.Kim; Experimental observation of a theoretically predicted nonlinear sleep spindle harmonic in human EEG. *Clinical Neurophysiology*. **(2014)**. (In Press)

[19] Shayan Motamedi-Fakhr, Mohamed Moshrefi-Torbati, Martyn Hill, M.Catherine Hill, R.Paul White; Signal processing techniques applied to human sleep EEG signals—A review. *Biomedical Signal Processing and Control*. **10**, 21-33 **(2014)**.

[20] H.Sohn, I.Kim, W.Lee, et al; Linear and non-linear EEG analysis of adolescents with attention-deficit/hyperactivity disorder during a cognitive task. *Clinical Neurophysiology*. **121(11)**, 1863-1870 **(2010)**.

[21] Spasic Sladana; Surrogate data test for nonlinearity of the rat cerebellar electrocorticogram in the model of brain injury. Signal Processing, **90(12)**, 3015-3025 **(2010)**.

[22] T.Suzuki, T.,Ikeguchi, M.Suzuki; Algorithms for generating surrogate data for sparsely quantized time series. *Physica D: Nonlinear Phenomena*. **231(2)**, 108-115 **(2007)**.

[23] Thomas Schreiber, Andreas Schmitz; Surrogate time series. *Physica D: Nonlinear Phenomena*. **142(3-4)**, 346-382 **(2000)**.

[24] U.Rajendra Acharya, Eric Chern-Pin Chua, Oliver Faust, Teik-Cheng Lim, Liang Feng Benjamin Lim; Automated detection of sleep apnea from electrocardiogram signals using nonlinear parameters. *Physiological Measurement*. **32(3)**, 287 **(2011)**.

[25] V.Sakkalis; Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Computers in Biology and Medicine*. **41(12)**,1110-1117 **(2011)**.

[26] Weiting Chen, Zhizhong Wang, Hongbo Xie, Wangxin Yu; Characterization of Surface EMG Signal Based on Fuzzy Entropy. *IEEE transactions on neural systems and rehabilitation engineering.* **15(2)**, 266-272 **(2007)**.

[27] Weiting Chen, Jun Zhuang, Wangxin Yu, Zhizhong Wang; Measuring complexity using FuzzyEn, ApEn, and SampEn. *Medical Engineering & Physics*. **31(1)**, 61-68 **(2009)**.

[28] Yifan Zhao, S.A.Billings, Hualiang Wei, P.G.Sarrigiannis; A Parametric Method to Measure Time-Varying Linear and Nonlinear Causality with Applications to EEG Data. *IEEE Transactions on Biomedical Engineering*. **60(11)**, 3141-3148 **(2013)**.