Using Fuzzy Measure Entropy to Improve the Stability of Traditional Entropy Measures

Chengyu Liu[1], Lina Zhao[2]

[1] School of Control Science and Engineering, Shandong University, Jinan, P.R.China
[2] Shandong Heng-Xin Inspection Technique Exploiture Center, Jinan, P.R.China

Abstract

Traditional entropy measures, such as Approximate Entropy (ApEn) and Sample Entropy (SampEn), are widely used for analyzing heart rate variability (HRV) signals in clinical cardiovascular disease studies. Nevertheless, traditional entropy measures have a poor statistical stability due to the 0-1 judgment of Heaviside function. The objective of this study is to introduce a new entropy measure – Fuzzy Measure Entropy (FuzzyMEn) in order to improve the stability of traditional entropy measures through introducing the concept of fuzzy sets theory. By drawing on Chen et al's research in fuzzy entropy (FuzzyEn), FuzzyMEn uses the membership degree of fuzzy function instead of the 0-1 judgment of Heaviside function as used in the ApEn and SampEn. Simultaneity, FuzzyMEn utilizes the fuzzy local and fuzzy global measure entropy to reflect the whole complexity implied in physiological signals and improves the limitation of FuzzyEn, which only focus on the local complexity. Detailed contrastive analysis and discussion of ApEn, SampEn, FuzzyEn and FuzzyMEn were also given in this study.

1. Introduction

Approximate entropy (ApEn) was introduced as a quantification of the regularity in the time series, initially motivated by the applications to relatively short, noisy time series, based on statistical analyses that appear to be compatible with the general clinical need to distinguish healthy subjects from the abnormal [1]. However, ApEn is a biased estimation for the complexity of the physiological signal what with the self-matching. The ApEn usually shows much dependence on the data length and lacks statistical stability [2,3]. Therefore, Richman et al proposed another statistic, sample entropy (SampEn), to relieve the bias caused by the self-matching. Nevertheless, whatever it is ApEn or SampEn, the calculation requires, initially, selection of the threshold value r. The essential calculation process of ApEn and SampEn is to examine whether or not the vector $X(i)$ is similar to another vector $X(j)$ using the threshold value r. They are two methods to select the r. One is the traditional method to select a fixed r [1,2]. Another one is the recent method to select the r_{max} from the range 0.01 to 1.0 times the standard deviation of the time series which maximizes the ApEn [4,5]. When distinguish two vectors $X(i)$ and $X(j)$, no matter the traditional fixed r or the recent r_{max}, they are constants and are based on Heaviside function of the classical sets. Heaviside function is essentially as a two-state classifier. So it induces the poor statistical stability of ApEn and SampEn. Some clues could be found in the study of Castiglioni et al. [6]. In our recent study, we systemically compared the different threshold value r to the influence of ApEn and discovered that the inherent flaw of the poor statistical stability arise from the constant r [7].

To overcome the poor statistical stability of ApEn and SampEn, Chen et al. introduce the membership degree of Zadeh's fuzzy sets theory to replace the conception of absolute two-state classifier and propose a statistic named fuzzy entropy (FuzzyEn) [8,9]. To explore the inherent vectors' similarity, FuzzyEn focuses on the local vectors' shapes and defines the local shape similarity as the vectors' similarity. Therefore, the global fluctuation of the time series is ignored and the global information is lost. Unlike Chen and his associates' study, we proposed a new statistic of entropy measure named fuzzy measure entropy (FuzzyMEn) in this paper. FuzzyMEn use the fuzzy local and fuzzy global measure entropy to respectively reflect the local complexity and whole complexity implied in physiological signal time series.

2. Fuzzy measure entropy (FuzzyMEn)

The detailed calculation process of FuzzyMEn is described as follows:

1) For an physiological signal sequence $X = \{x_1, x_2, \ldots, x_N\}$, let x_{mean} denotes the mean value of the sequence. The embedding dimension m is set to a constant 2. Then from the beginning of X, choose m consecutive x values to form the vector $X_m(i) = [x_i, x_{i+1}, \ldots, x_{i+m-1}]$, where the i is from 1 to $N-m+1$. So it produces $N-m+1$ vectors in all. Like SampEn, FuzzyMEn excludes self-matches and considers only the first $N-m$ vectors of length m to ensure that $X_m(i)$

and $X_{m+1}(i)$ are defined for all $1 \leqslant i \leqslant N - m$. Let x_{0i} denotes the mean value of vector $X_m(i)$.

2) Form the local vector $XL_m(i) = [x_i - x_{0i}, x_{i+1} - x_{0i}, \ldots, x_{i+m-1} - x_{0i}]$ and global vector $XF_m(i) = [x_i - x_{mean}, x_{i+1} - x_{mean}, \ldots, x_{i+m-1} - x_{mean}]$, both $XL_m(i)$ and $XF_m(i)$ has length m and $1 \leqslant i \leqslant N - m$. The $dL_m(i, j)$ and $dF_m(i, j)$ are respectively removed a local baseline and global baseline. They are defined to respectively describe the distance of the local vectors between $XL_m(i)$ and $XL_m(j)$, and global vectors between $XF_m(i)$ and $XF_m(j)$. The calculated formula for $dL_m(i, j)$ and $dF_m(i, j)$ are described as follows:

$$\begin{cases} dL_m(i,j) = d[XL_m(i), XL_m(j)] \\ \qquad = \max\left|(x_{i+k} - x_{0i}) - (x_{j+k} - x_{0j})\right| \\ dF_m(i,j) = d[XF_m(i), XF_m(j)] \\ \qquad = \max\left|(x_{i+k} - x_{mean}) - (x_{j+k} - x_{mean})\right| \end{cases} \quad (1)$$

3) Then two fuzzy functions, the local fuzzy function $\mu_L(dL_m(i,j), n_L, r_L)$ and global fuzzy function $\mu_F(dF_m(i,j), n_F, r_F)$ can be calculated as follows:

$$\begin{cases} \mu_L(dL_m(i,j), n_L, r_L) = \exp(-(dL_m(i,j)/r_L)^{n_L}) \\ \mu_F(dF_m(i,j), n_F, r_F) = \exp(-(dF_m(i,j)/r_F)^{n_F}) \end{cases}, \quad (2)$$

where both the local and global fuzzy function similarly use the typical exponential function, r_L and r_F respectively denote the threshold value r for $\mu_L(dL_m(i,j), n_L, r_L)$ and $\mu_F(dF_m(i,j), n_F, r_F)$, n_L and n_F denote the weight of vectors' similarity for $\mu_L(dL_m(i,j), n_L, r_L)$ and $\mu_F(dF_m(i,j), n_F, r_F)$. In this study the r_L and r_F also set to be 0.2 times the standard deviation of the time series. A value of n_L or n_F more than 1 will weights the similarity of the close vectors and unweights that of the far vectors in $XL_m(i)$ or $XF_m(i)$, while a value of n_L or n_F less than 1 has the contrary effect. The larger n_L or n_F is, the more the closer vectors and less the further ones are weighted. When n_L or n_F becomes enough large close to infinity, the exponential function in equation (2) is reduced to Heaviside function. So a large n_L or n_F will lose the detailed information. In this study the n_L and n_F respectively set to be 3 and 2.

So the local similarity degree $DL_m(i,j)$ between vector $XL_m(i)$ and $XL_m(j)$ and the global similarity degree $DF_m(i,j)$ between vector $XF_m(i)$ and $XF_m(j)$ can be calculated as follows:

$$\begin{cases} DL_m(i,j) = \mu_L(dL_m(i,j), n_L, r_L) = \exp(-(dL_m(i,j)/r_L)^{n_L}) \\ \mathrm{DF}_m(i,j) = \mu_\mathrm{F}(d\mathrm{F}_m(i,j), n_\mathrm{F}, r_\mathrm{F}) = \exp(-(d\mathrm{F}_m(i,j)/r_\mathrm{F})^{n_\mathrm{F}}) \end{cases} \quad (3)$$

For all $1 \leqslant i, j \leqslant N-m$, the mean values of $DL_m(i,j)$ and $DF_m(i,j)$ respectively describe as $\phi L_m(n_L, r_L)$ and $\phi F_m(n_F, r_F)$. The calculated formula for $\phi L_m(n_L, r_L)$ and $\phi F_m(n_F, r_F)$ are described as follows:

$$\begin{cases} \phi L_m(n_L, r_L) = \dfrac{1}{N-m} \sum_{i=1}^{N-m} \left(\dfrac{1}{N-m-1} \sum_{j=1, j \neq i}^{N-m} DL_m(i,j) \right) \\ \phi F_m(n_F, r_F) = \dfrac{1}{N-m} \sum_{i=1}^{N-m} \left(\dfrac{1}{N-m-1} \sum_{j=1, j \neq i}^{N-m} DF_m(i,j) \right) \end{cases} \quad (4)$$

4) Then the local vector $XL_m(i)$ and global vector $XF_m(i)$ with length m are extended to $XL_{m+1}(i)$ and $XF_{m+1}(i)$, which has the length $m+1$. The $\phi L_{m+1}(n_L, r_L)$ and $\phi F_{m+1}(n_F, r_F)$ are calculated repeating the similar step (1) to (3).

5) The fuzzy local measure entropy $FuzzyLMEn$ and fuzzy global measure entropy $FuzzyFMEn$ are defined as follows:

$$\begin{cases} FuzzyLMEn(m, n_L, r_L, N) = \ln \phi L_m(n_L, r_L) - \ln \phi L_{m+1}(n_L, r_L) \\ FuzzyFMEn(m, n_F, r_F, N) = \ln \phi F_m(n_F, r_F) - \ln \phi F_{m+1}(n_F, r_F) \end{cases} \quad (5)$$

Therefore, FuzzyMEn for a RR sequence is calculated as follows:

$$\begin{aligned} & FuzzyMEn(m, n_L, r_L, n_F, r_F, N) \\ & = FuzzyLMEn(m, n_L, r_\mathrm{L}, N) + FuzzyFMEn(m, n_F, r_F, N) \end{aligned} \quad (6)$$

3. Comparison of ApEn, SampEn, FuzzyEn and FuzzyMEn

In this study, four types of entropy measures are discussed and compared: ApEn, SampEn, FuzzyEn and FuzzyMEn proposed in this paper. The detailed description of the former three entropy measures can be referred to these literatures: ApEn [1,10], SampEn [3] and FuzzyEn [8,9]. To illuminate the relations and differences of four entropy measures, we take an in-depth analysis for the mechanism of different entropy measure before the emulation experiment.

3.1. Vector similarity judgment standard

When distinguishing whether a vector $X(i)$ is similar to another vector $X(j)$ using a threshold r, either in ApEn or in SampEn, the judging standard is based on Heaviside function, with the formula expressed as:

$$\theta(d_{ij}, r) = \begin{cases} 1 & d_{ij} \leq r \bullet \sigma_X \\ 0 & d_{ij} > r \bullet \sigma_X \end{cases}. \quad (7)$$

Where d_{ij} denotes the distance between the two vectors $X(i)$ and $X(j)$. Heaviside function is a two-state classifier essentially and its judging standard is very rigid: the vectors with the d_{ij} inside the boundary are treated equally, while the vectors just outside the boundary are abnegated. To weaken the rigid judging standard of the conventional two-state classifier, FuzzyEn and FuzzyMEn introduce the "membership degree" with a fuzzy function μ_X to the entropy calculation process. Fuzzy sets theory provides a mechanism for measuring the membership degree to which a vector belongs to a given class: the nearer the value of μ_X to 1, the higher the membership degree of the vector to the given class. Through the fuzzy function μ_X, the 0-1 judgment in the classical sets theory is instead by the continuous membership degree within 0 and 1. This judgment method for two vectors $X(i)$ and $X(j)$ can be summarized as follows:

$$X(i), X(j) \xrightarrow{\mu_X(X(i),X(j))} [0,1]. \qquad (8)$$

FuzzyEn and FuzzyMEn employ the exponential function $\exp(-(d_{ij}/r)^n)$ as the fuzzy function with a view to its fine characteristic of smoothness, continuity, convexity, symmetry, and so on [8]. In the exponential function there is no rigid boundary. This phenomenon is clearly shown in Figure 1. Take the similarity of the two vectors $X(m)$ and $X(n)$ to the vector $X(i)$ for example. The vector $X(i)$ set to be the center vector. The d_{ij} between $X(i)$ and $X(m)$ and the d_{ij} between $X(i)$ and $X(n)$ respectively set to be 0.195 and 0.205 times the standard deviation of the signal. When the threshold r is 0.2 times the standard deviation of the signal, the $X(m)$ is regard as the similar vector for $X(i)$, while the $X(n)$ is rejected as the similar vector for $X(i)$. The judging standard is so rigid that the two original similar vectors $X(m)$ and $X(n)$ are regard as absolutely different. But when increase the threshold r to 0.21 times the standard deviation, both $X(m)$ and $X(n)$ are regard as the similar vectors for $X(i)$, while when decrease the threshold value r to 0.19 times the standard deviation, both $X(m)$ and $X(n)$ are regard as absolutely different vectors for $X(i)$. It is hardly to choose an appropriate r for all situations, so the statistical results for vector similarity will exhibits a great instability. This is the reason of the poor statistical stability for traditional entropy measures ApEn and SampEn.

3.2. Local similarity or global similarity?

There is a distinct difference between FuzzyEn and FuzzyMEn. FuzzyEn just focuses on the similarity of local waveform shape but ignores the fluctuation of global waveform shape. Figure 2 shows a section of R-R intervals signal. There are four vectors $X(i)$, $X(m)$, $X(m)$ and $X(j)$. They all have the same local waveform shape. When we take into account the similarity between the reference vector $X(i)$ and the other three vectors, FuzzyEn will make the similar results because the slight difference in the local waveform shape of above four vectors. Whereas, the vector $X(j)$ is very close to the reference vector $X(i)$ not only in local waveform shape but also in global waveform shape, while vectors $X(m)$ and $X(n)$ is similar to the reference vector $X(i)$ only in local waveform shape but have a great difference in global waveform shape. FuzzyMEn integrates both local and global waveform shape to educe two aspect fuzzy measure entropies: fuzzy local measure entropy and fuzzy global measure entropy, to respectively reflect the local and global complexity implied in physiological signals. It improves the limitation of FuzzyEn, which only focus on the local complexity.

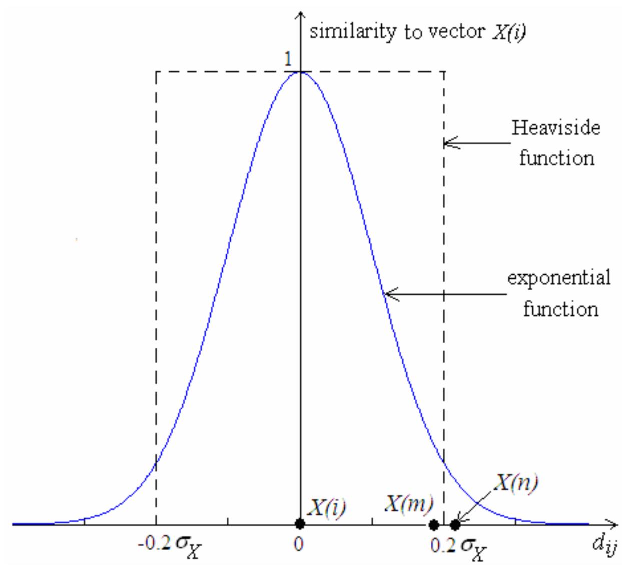

Figure 1. The similarity judgment of vectors $X(m)$ and $X(n)$ to the vector $X(i)$ using the Heaviside and exponential functions based on the d_{ij} of vectors.

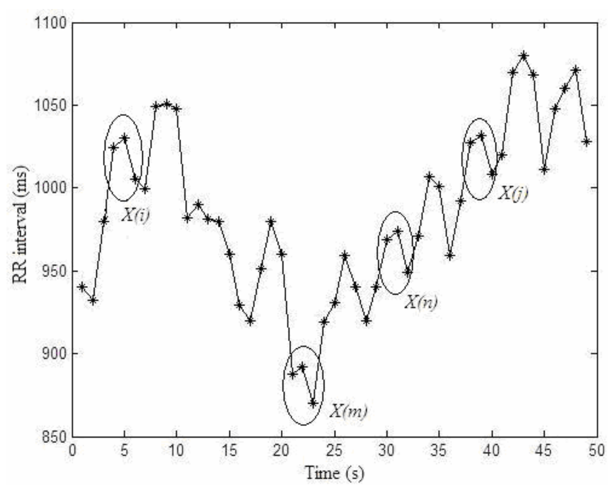

Figure 2. Both local and global waveform shape similarity implied in a section of RR interval signal.

4. Discussion and conclusions

Recent years have seen rapid development of various traditional entropy measures such as ApEn [1,2,10], SampEm [3], mode entropy (ModEn) [11], multiscale entropy(MSEn) [12], and so on. However, the traditional entropy measures exhibit a poor statistical stability for analysis of physiological signals such as HRV signal. In this study, we investigate the essential reason of the inherent limitation of ApEn and SampEm and educe the conclusion that the method by setting a certain threshold value r to judge a vector' belongingness to a given class is the essential reason of the poor stability in traditional entropy measures. So we introduce the concept of fuzzy sets theory and use the membership degree of fuzzy function to describe the relation whether a vector' belongingness to a given class. This judgment standard exhibits the gentle boundary effect while the traditional 0-1 judgment standard is rigid in the boundary of threshold value r. Based on these analysis, we proposed a new entropy measure named FuzzyMEn.

Subsequently, we studied the potential problem in FuzzyEn and compared the difference between FuzzyEn and FuzzyMEn. It was found that FuzzyEn just focused on the similarity of local waveform shape but ignored the fluctuation of global waveform shape. So we took into account the similarity not only of local waveform shape but also of global waveform shape. Therefore, FuzzyMEn was defined by integrating both fuzzy local measure entropy and fuzzy global measure entropy, to respectively reflect the local and global complexity implied in HRV signals. It improved the limitation of FuzzyEn, which only focused on the local complexity. The detailed differences between ApEn, SampEn, FuzzyEn and FuzzyMEn mentioned above were summarized in Table 1. The future work will design the systemic experiment to evaluate the algorithm consistency and sensitivity for FuzzyMEn.

Table 1 The detailed differences between ApEn, SampEn, FuzzyEn and FuzzyMEn

Difference aspects	ApEn	SampEn	FuzzyEn	FuzzyMEn
Sets theory	classical	classical	fuzzy	fuzzy
Judgment a vector to given class	0-1 judgment	0-1 judgment	membership degree	membership degree
Judgment function	Heaviside function	Heaviside function	fuzzy function	fuzzy function
Is self-matching calculated?	yes	no	no	no
Is a bias estimation?	yes	no	no	no
Description complexity	global	global	local	local and global

Acknowledgements

This work was supported by the China Postdoctoral Science Foundation funded project (No. 20110491593).

References

[1] Pincus S M. Approximate entropy as a measure of system complexity. Proc. Nat. Acad. Sci. USA 1991; 88: 2297-2301.

[2] Pincus S M and Goldberger A L. Physiological time-series analysis: what does regularity quantify? American Journal of Physiology: Heart and Circulatory Physiology 1994; 266: 1643-1656.

[3] Richman J S and Moorman J R. Physiological time-series analysis using approximate and sample entropy. American Journal of Physiology: Heart and Circulatory Physiology 2000; 278: 2039-2049.

[4] Lu S, Chen X, Kanters J K, Solomon I C and Chon K H. Automatic selection of the threshold value R for approximate entropy. IEEE Trans. Biomed. Eng. 2008; 55: 1966-1972.

[5] Chon K H, Scully C G and Lu S. Approximate entropy for all signals. IEEE Engineering in Medicine and Biology Magazine 2009; 28: 18-23.

[6] Castiglioni P and Rienzo M D. How the threshold "R" influences approximate entropy analysis of heart-rate variability. Computers in Cardiology 2008; 35: 561-564.

[7] Liu C Y, Liu C C, Shao P, Li L P, Sun X, Wang X P and Liu F. Comparison of different threshold values r for approximate entropy Application to investigate the heart rate variability between heart failure and healthy control groups. Physiological Measurement 2011; 32: 167-180.

[8] Chen W T, Wang Z Z, Xie H and Yu W X. Characterization of surface EMG signal based on fuzzy entropy. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2007; 15: 266-272.

[9] Chen W T, Zhuang J, Yu W X and Wang Z Z. Measuring complexity using FuzzyEn, ApEn, and SamEn. Medical Engineering and Physics 2009; 31: 61-68.

[10] Pincus S M and Singer B. Randomness and degrees of irregularity. Proc. Nat. Acad. Sci. USA 1996; 93: 2083-2088.

[11] Ning X B, Xu Y L, Wang J and Ma X F. Approximate entropy analysis of short-term HFECG based on wave mode. Physica A. 2005; 346: 475-483.

[12] Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. Phy. Rev. Lett. 2002; 89: 68-102.

Address for correspondence.

Chengyu Liu
School of Control Science and Engineering
Shandong University, 17923 Jingshi Road
Jinan, P. R. China 250061
Tel: +86-159-53148364
E-mail: bestlcy@sdu.edu.cn.