



METHODS IN COMPLEXITY ANALYSIS OF DISCRETE TIME SIGNALS

Zoltán GERMÁN-SALLÓ¹, Zoltán GERMÁN-SALLÓ, jr.²

¹ George Emil Palade University of Medicine, Pharmacy, Sciences and Technology of Tirgu Mures,

38 Gheorghe Marinescu street, Tirgu Mures, Romania

¹zoltan.german-sallo@umfst.ro

²Technical University of Cluj-Napoca,

28 Memorandumului street, Cluj-Napoca, Romania

²zoltan.germansallo@protonmail.com

Abstract

Discrete time signals carry information about systems and their internal functional mechanisms which characterize their complexity. Complexity measures are strongly related to information content and evaluations have been made on various signals in many ways in last few years. This paper uses information theory estimates of complexity as different types of entropies in order to estimate the complexity of various time discrete synthesized signals. Results show that this kind of indices can be a useful tool in diagnostic, fault detection and further development.

Key words: entropy, complexity, regularity, randomness

1. Introduction

Quantifying the complexity of a discrete signal is an important task because usually it is strongly related to the information content of the signal. This paper introduces different entropies as information theory-based parameters in order to evaluate the complexity of discrete signals. Both of clear and noisy signals are analyzed, the evaluations are performed on synthesized signals. This paper is organized as follows: the second chapter offers a review of the most used entropy metrics and presents their characteristics through the corresponding implementing algorithms. In chapter three are presented the test signals and the proposed procedures, chapter four brings the experimental results. Finally, concluding remarks and possible future work ideas are presented.

2. An overview of complexity indices

This paragraph presents the most applied entropy forms in discrete time series analysis. Fundamentally, the entropy is strongly related to the information content of a discrete signal due to its probability density function [1]. Since Shannon's first definition of entropy as average information content (1948), many other definitions of entropy appeared in scientific researches. The main idea of Shannon entropy is a measure of the average information content of a discrete events-based information source, defined by the equation

$$S(x) = -\sum_{i=1}^n p(x_i) \log_2[p(x_i)] \quad (1)$$

where $X(x_i)$ is a discrete random variable, $p(x_i)$ is the probability of a discrete event x_i [1].

The Fourier or Spectral entropy is another complexity index defined in frequency domain. To obtain this entropy, a power-spectral-density (PSD) of a time series is computed from the signals Fourier transform as:

$$X(\omega_i) = DFT(x_i) \quad (2)$$

The Power Spectral Density of the signal is performed through squaring its amplitude and normalizing by the number of bins:

$$PSD(\omega_i) = \frac{1}{N} |X(\omega_i)|^2 \quad (3)$$

The calculated PSD must be normalized so that it can be viewed as a probability density function:

$$p_i = \frac{PSD(\omega_i)}{\sum_i PSD(\omega_i)} \quad (4)$$

The corresponding Shannon entropy is called the Fourier or Spectral entropy:

$$FE = -\sum_S p_i \cdot \log p_i \quad (5)$$

This entropy seems to be the easiest approximation in case of large data sets, often is a reference in comparison with other methods [2]. One of most important of lately introduced entropies is the Sample Entropy (SE) which estimate the conditional probability that two sequences of given number successive samples (embedding dimension m) in a signal that are similar to each other (within a given tolerance r) will remain similar when one more successive sample is included [3], [4].

The sample entropy (SE) was introduced by Richman and Moorman [5] and is the negative natural logarithm of the conditional probability that two sequences similar for a given number of points remain similar when one more point is added, where self-matches are eliminated in calculating the probability. The probability that two sequences match for m points

$$P^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} C_i^m(r) \quad (6)$$

the number of matching templates for each template within a time series [6].

$$C_i^m(r) = \frac{1}{N-m} \sum_{j=1, x \neq 1}^{N-m+1} H(r - d_{ij}) \quad (7)$$

H is the Heaviside function, defined as

$$H(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases} \quad (8)$$

For $m + 1$, $C_i^m(r)$ and $\Phi^m(r)$ are calculated [7]. Finally, sample entropy is:

$$SE(n, r) = -\log \frac{P_{m+1}(r)}{P_m(r)} \quad (9)$$

The permutation entropy introduced by C. Bandt and B. Pompe in 2002 describes complexity of a time series or discrete signals through phase space reconstruction [6], [7]. The implementing algorithm partitions the one-dimensional discrete signal into a matrix of overlapping column vectors using two parameters, the embedding time delay τ which controls the number of time periods between elements of each of the new column vectors and the embedding dimension D which controls the length of each of the new column vector. In the new matrix each column vector has a number of elements equal to the embedding dimension, D , the duration between each element in the vectors is equal to the embedded time delay τ . The relative frequencies of appearance of the possible patterns of symbol sequences are named permutations. These relative frequencies are obtained as appearance probabilities in the created patterns.

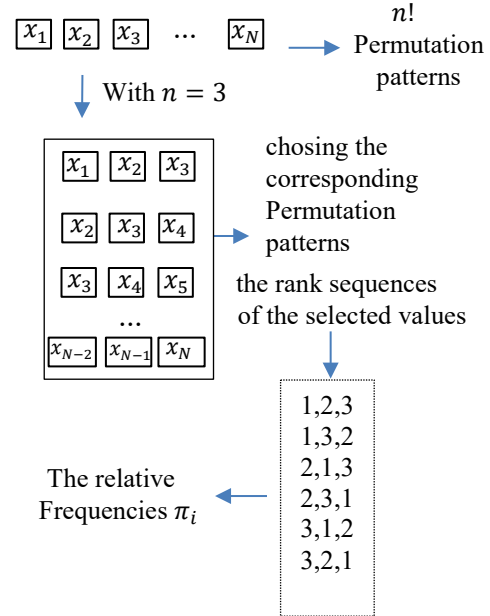


Fig. 1. Exemplification for pattern length 3

The permutation entropy which reflects the rank order of successive x_i in sequences of length n is defined by

$$PE_n = \sum_{i=1}^n \pi_i \cdot \log_2 \pi_i \quad (10)$$

Figure 1 presents the case of embedding length of 3 and time unitary time delay.

3. Materials and method

The proposed method will analyze the results obtained on various discrete time signals with different lengths presenting different levels of complexity. The first two are taken from Matlab Signal processing Toolbox ("vonkoch"-V, "leleccum"-L) and the third is a synthesis of different test signals from the same

database, noted as T. These signals have different lengths and different complexity degrees. The proposed test signals are presented on figure 2.

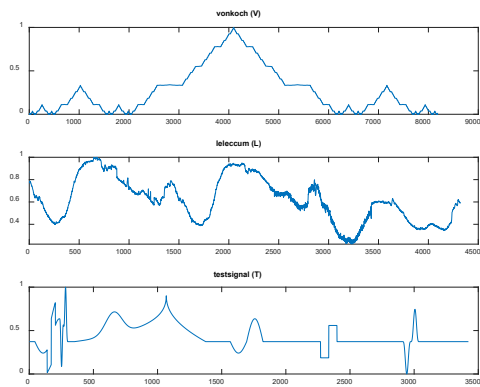


Fig. 2. The used test signals

These signals also have a very different amplitude distribution as it can be seen on figure 3

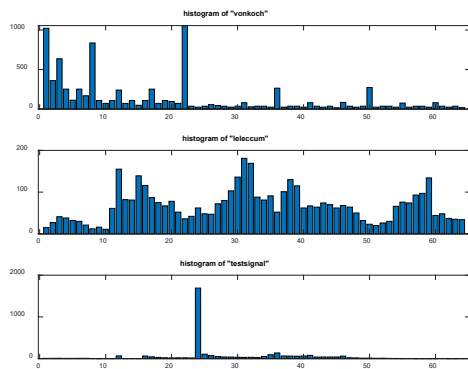


Fig. 3. The histograms of the used signals

The procedure of complexity evaluation using different entropy-based metrics is presented in a general form on figure 4. In the preprocessing phase, the signal are normed and zero valued samples were replaced with a small offset value to make possible logarithmic computation [8], [9], [10].

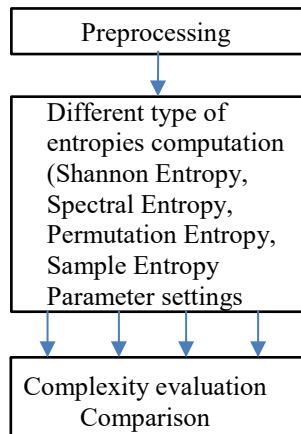


Fig. 4: The proposed procedure

4. Experimental results

At first the Shannon entropies were computed for the test signals, in this case there are no significant differences as it can be observed on figure 5.

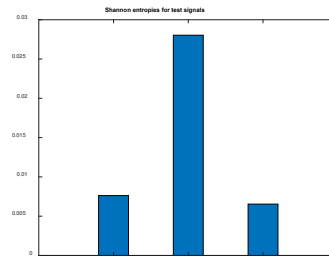


Fig. 5. The Shannon entropies for test signals

The spectral entropies are computed through Fourier transform based power spectral density estimation. The three signals have the following PSDs.

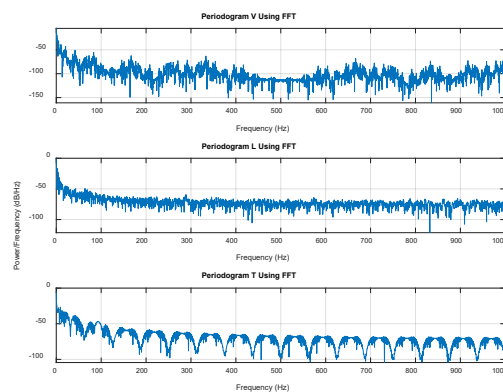


Fig. 6. The periodograms of used signals

Sample entropies were computed for a 32 embedding length and the threshold was set as usual on 20% of standard deviation. In case of permutation entropy, the time delay was unitary and the embedding length was chosen 3. After that all proposed entropies are computed, the results are shown on figure 7.

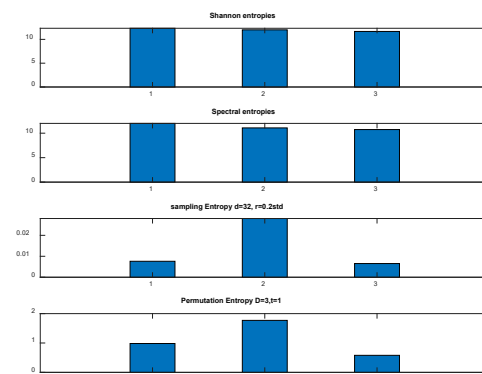


Fig. 7. The computed entropies for the test signals

Adding a gaussian white noise of different levels the results concerning the sample entropy are presented on figure 8.

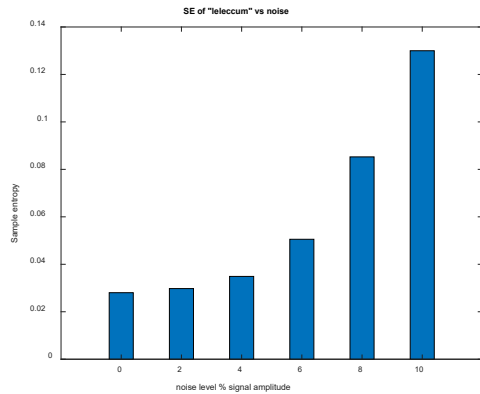


Fig. 8. The sample entropy vs. added gaussian white noise
The permutation entropy seems to be not really influenced by the added white noise, the truth is that after a relative low level, the algorithm can fail if it becomes comparable with the difference between two consecutive samples. The results are shown on figure 9.

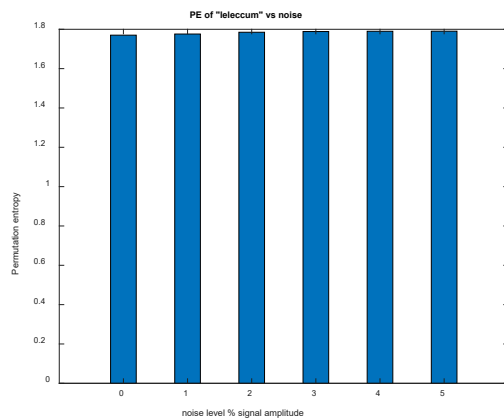


Fig. 9. The permutation entropy vs. added gaussian white noise

5. Discussion

This paper focuses on information theory-based complexity indices which are different types of entropies. Three types of discrete signals with very different properties were analyzed through four types of entropies in order to evaluate their complexity. Usual, complexity means more information content and greater entropy value. The most accurate and useful results were given by sample and permutation entropies which are more complex as procedure and

computational algorithm. Sampling and permutation entropies gave a more accurate difference concerning signal complexity as the Shannon or Fourier entropies at their level of computation. The effect of added noise is that the measured complexity indices will be augmented which means that for an accurate evaluation the analyzed signal must be filtered previously.

References

- [1] Shannon, C. E., *A mathematical theory of communication*, Bell Syst. Tech. J. Nr. 27 (1948) pp. 379-423., 623-656.
- [2] Fefferman, R., A. *A theory of entropy in Fourier analysis*. Advances in Mathematics Volume 30, Issue 3, December 1978, Pages 171-201
- [3] Jiang, Y., Mao, D. and Xu., Y., *A fast algorithm for computing sample entropy*, Advances in Adaptive Data Analysis, Vol. 03, No. 01, (2011), pp. 167-186.
- [4] Pincus, S. M., *Approximate entropy as a measure of system complexity*. Proceeding of National Academy of Sciences USA, Mathematics, Vol. 88 (1991), pp. 2297-2301.
- [5] Lake, D. E., Richman, J. S , Griffin, M. P., Moorman, J. R. *Sample entropy analysis of neonatal heart rate variability*. Am J Physiol 2002; 283(3):R789-R797
- [6] Bandt, C.; Pompe, B. *Permutation entropy: A natural complexity measure for time series*. Phys. Rev. Lett. 2002, 88, 174102:1-174102:4
- [7] Molina-Picó, A., Cuesta-Frau,D., Aboy, M., Crespo, C., Miró-Martínez, P., Oltra-Crespo, S. *Comparative study of approximate entropy and sample entropy robustness to spikes*, Artificial Intelligence in Medicine, vol. 53, no. 2, 2011, pp. 97-106.
- [8] Lempel, A., Ziv, J., *On the complexity of finite sequences*, IEEE Transactions on Information Theory, vol. IT-22, no. 1, 1976, pp. 75-81.
- [9] Y. Bar-Yam, *Dynamics of complex systems*, vol. 213. Addison-Wesley Reading, MA, 1997.
- [10] Ziv, J., Merhav, J., Z., *A measure of relative entropy between in-dividual sequences with application to universal classification*, IEEETransaction on Information Theory, vol. 39, no. 4, pp. 1270–1279, Jul. 1993.