Application of the Lempel-Ziv complexity measure to the analysis of biosignals and medical images

Borowska M[1], Oczeretko E[1], Mazurek A[2], Kitlas A[1], Kuć P[2]

[1] Institute of Computer Science, University of Białystok, Poland
[2] Department of Pathophysiology of Pregnancy, Medical University of Białystok, Poland

Abstract

Purpose: **The aim of this study was to apply a Lempel-Ziv complexity measure for quantifying biomedical signals and images.**

Material and methods: **We analyzed angiogenic patterns and the signals (the heart rate, the respiration rate and the blood oxygen concentration). Biomedical signals were obtained by means of Internet. Medical images were from Department of Pathophysiology of Pregnancy Medical University of Białystok.**

Results: **The values of normalized complexity measures for respiratory rate signal are high, what indicates that this time series is close to unstructured randomness. The Lempel-Ziv complexity values for angiogenic patterns were growing with the FIGO stage of disease.**

Conclusions: **Lempel-Ziv complexity may be a very helpful tool in analyzing the signals and images. It can be easily computed from the analysed data.**

Key words: signal and image processing, Lempel-Ziv complexity measure, angiogenic patterns, biomedical time series.

Introduction

Lempel and Ziv proposed a useful complexity measure, which can characterize the degree of order or disorder and development of spatiotemporal patterns [1]. In the first step the signals and images are transformed into binary sequences. Lempel-Ziv algorithm gives the number of distinct patterns contained in the given finite sequence. After normalisation the relative Lempel-Ziv complexity measure (L-Z measure) reflects the rate of new pattern occurrences in the investigated series of the symbols. L-Z values range from near 0 (deterministic equation) to 1 (totally destructured random pattern white noise). That is a new approach for quantifying of medical signals and images [2,3]. The program for Lempel-Ziv complexity analysis was written in C++.

ADDRESS FOR CORRESPONDENCE:
Marta Borowska
Institute of Computer Science, University of Białystok
64 Sosnowa Str., 15-887 Białystok, Poland
Tel/Fax: +48 85 7457662
e-mail: mborowska@ii.uwb.edu.pl

Materials and methods

Biomedical time series

The data were obtained via PhysioNet (http://www.physionet.org/physiobank/santa-fe). PhysioNet offers free Internet access to various kinds of recorded physiological signals and related open-source software [4,5]. PhysioNet is a public service of the Research Resource for Complex Physiological Signals created under the auspices of the National Center for Research Resources of the National Institutes of Health (USA).

These multivariate data were obtained from a sleeping patient and contain three simultaneously recorded signals: heart rate (signal s1), which is defined through the number of heartbeats per minute, respiration rate (or chest volume) – signal s2 and blood oxygen concentration (measured by ear oximetry) – signal s3. The respiration rate and the blood oxygen concentration are given in uncalibrated analog-to-digital converted units. For these three time series the sampling frequency was 2 Hz. The studied patient shows sleep apnea. Sleep apnea is a disorder in which a person stops breathing during the night (sometimes a hundreds of times). He stops breathing for up to 45 seconds.

These signals were divided into 8 segments. Each segment contained 2000 elements.

Angiogenic images

Quantification of angiogenesis is a valuable prognostic tool for tumor progression and metastasis [6,7]. Angiogenic images were from Department of Pathophysiology of Pregnancy Medical University of Białystok. We analyzed 42 angiogenic patterns which were divided into three groups: group A I FIGO stage of disease (n = 19), group B II FIGO stage of disease (n = 13) and group C III FIGO stage of disease (n = 10).

The images were binarized by means of Otsu method [8].

The Lempel-Ziv Complexity Measure – algorithm

The signal to be analyzed is transformed into sequence whose elements are a few symbols. The binary sequence is simple to construct: the data values below or equal the mean have the symbol "1" and the values above the mean have the symbol "0". This algorithm gives the number of distinct patterns contained in the given finite sequence S=s1, s2, …, sn [9]. The calculation of c(n) (Lempel-Ziv complexity) proceeds on diagram (*Fig. 1*).

This method uses comparison and accumulation so the computation of c(n) is easy to calculate.

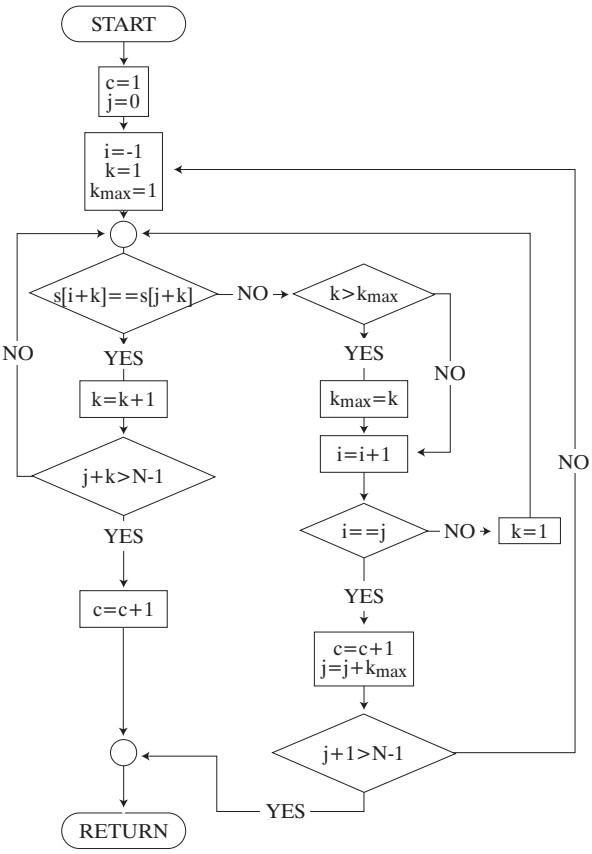
The Lempel-Ziv of the totally random sequence of length n consisting of two different symbols with equal probabilities is

$$b(n)=\frac{n}{log_2(n)}$$

If we divide the complexity of the sequence by the complexity b(n) of the random sequence, we get the normalized Lempel-Ziv C(n), which does not depend on the length of the sequence when n is large

$$C(n)=\frac{c(n)}{b(n)}$$

Figure 1. **Flow chart of the computer program for calculation of Lempel-Ziv complexity (modified from [10])**



Results

The mean values of Lempel-Ziv complexity for biomedical signals are:

1. The high L-Z complexity value for respiratory rate signal indicates that this time series is close to unstructured randomness.
2. The mean values of Lempel-Ziv complexity for angiogenic patterns are.
3. The Lempel-Ziv complexity values were growing with the FIGO stage of disease.

Conclusions

The values of normalized complexity measures for respiratory rate signal are high, what indicates that this time series is close to unstructured randomness. The Lempel-Ziv complexity values for angiogenic patterns were growing with the FIGO stage of disease. We conclude that Lempel-Ziv complexity may be a very helpful tool in analyzing the medical images and biomedical time series. It can be easily computed from the analyzed data.

References

1. Lempel A, Ziv J. On the complexity of finite sequences. IEEE Trans Inform Theory, 1976; 22(1): 75-81.
2. Szczepanski J, Amigo JM, Wajnryb E, Sanches-Vives MV. Characterizing spike trains with Lempel-Ziv complexity. Neurocomputing, 2004; 58-60: 79-84.
3. De Felice C, Bianciardi G, DiLeo L, Latini G, Parrini F. Abnormal oral vascular network geometric complexity in Ehlers-Danlos syndrome. Oral Surg Oral Med Oral Path Oral Radiol Endod, 2004; 98: 429-34.
4. http://www.physionet.org/physiobank/database/santa-fe
5. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng ChK, Stanley HE. Components of a New Research Resource for Complex Physiologic Signals. Circulation, 2000; 101: 215-20.
6. Ahmed A, Perkins J. Angiogenesis and intrauterine growth restriction. Bailliere's Best Pract Res Clin Obstet Gynaecol, 2000; 14: 981-98.
7. Oczeretko E, Juczewska M, Kasacka I. Fractal geometric analysis of lung cancer angiogenic patterns. Folia Histochem Cytobiol, 2001; 39 (Suppl. 2): 75-6.
8. Otsu N. A threshold selection method from gray-level histogram. IEEE Trans Syst Man Cybern, 1979; SMC-9 (1): 62-6.
9. Zhang X-S, Roy RJ, Weber Jensen E. EEG Complexity as a Measure of Depth of Anesthesia for Patients. IEEE Transaction on Biomedical Engineering, 2001; 48(12): 1424-32.
10. Hongxuan Z, Yisheng Z. Qualitative chaos analysis for ventricular tachycardia and fibrillation based on symbolic complexity. Medical Engineering & Physics, 2001; 23: 523-8.