LEMPEL-ZIV COMPLEXITY FOR ANALYSIS OF NEURAL SPIKES

Mehran Talebinejad, Georgios Tsoulfas, Sam Musallam
Department of Computer and Electrical Engineering, and Department of Physiology, McGill University, Quebec, CANADA

INTRODUCTION

The neural signals recorded from certain parts of the brain using microelectrodes at the single neuron level provide great amount of information about the dynamics of decision making and has been widely researched for development of brain machine interfaces (BMIs) [1-3].

Conventionally, while a subject is performing a reach task according to some specific paradigm microelectrodes record the electrical activity of neuron(s) near the tip of the electrode. The recorded signal contains spikes and low frequency content. The spikes represent the firing of one or more neurons while low frequency content is the sum of all other electrical activity around the electrode.

Analysis of neural spikes is well established [4]. Many methods can be used to uncover distinct patterns of spikes. Consider two spike trains: 10101010 and 11110000, recorded in 1 s (where 1 represents the occurrence of a spike in a bin of length 1000/8 ms). The firing rates of both spike trains are identical (4 Hz), however the patterns are different; that is, the first train shows an oscillatory behavior while the second train resembles a step function. If two separate stimuli elicited these patterns, then the firing rate in this example is inadequate for characterizing the stimulus and does not provide any information about the distinct patterns that might be observed in a spike train. Analyses of distinct spike patterns have revealed underlying nonlinear and chaotic dynamics of firing, and would benefit from complementary metrics for complexity analysis of neural spikes.

Lempel-Ziv (LZ) complexity can be used to quantify distinct patterns in symbolic sequences, especially binary signals [5-11]. LZ gives a measure of complexity based on an estimated number of distinct patterns obtained by parsing a symbolic sequence [5-9].

In this paper, we evaluate requirements of the LZ complexity for ultra-short time series applications similar to neural spikes. We also compare two different parsing approaches and examine which one is more accurate for characterizing ultra-short time series and potentially neural spikes.

METHODS AND MATERIALS

Lempel-Ziv Complexity

The LZ complexity in practice is limited to a finite number of symbols, well suited for binary signals. It is also applicable to three or more symbols [5].

To determine all the distinct patterns in a spike train, all the characters and patterns are parsed and inspected for distinct patterns. There are different approaches for parsing a binary sequence [5-11]. In this work, we compare the original LZ algorithm [5] denoted as ***Algorithm 1***, to a modified algorithm presented in [7] denoted as ***Algorithm 2***.

Algorithm 1 [5]: Let the binary signal P_s, be divided into substrings $P_s(i,j)$ that start at position i and end at position j. That is, when $i < j$, $P_s(i,j)$ = s(i)…s(j) and, when $i > j$, $P_s(i,j)$ = ø, is the null set. Let $V(P_s)$ denote the set of substrings in the signal P_s excluding the null set (e.g., if P_s = 001, then $V(s)$ = {0, 1, 00, 01, 001}). The parsing procedure involves a left-to-right scan of the signal P_s. A substring $P_s(i,j)$ is compared to $V(P_s(1,j\text{-}1))$. If $P_s(i,j)$ is present in $V(P_s(1, j\text{-}1))$, then $P_s(i,j)$ and $V(P_s(i,j\text{-}1))$ are updated to $P_s(i,j+1)$ and $V(P_s(1,j))$, respectively, and the process repeats. If the substring is not present, then $s(j)$ is marked to indicate the end of a new component, and $P_s(i,j)$ and $V(P_s(1,j\text{-}1))$ are updated to $P_s(j+1,j+1)$ and $V(P_s(1,j))$, respectively. The process continues until $j = n$, where the n is the length of the symbolic sequence P_s. For example, 10101010 and 11110000 are parsed as 1.0.10.10.10, and 1.111.0.000 and the number of distinct patterns are 3 and 4 with $V(P_s)$s, {1,0,10} and {1,11,10,0,00} respectively.

Algorithm 2 [7]: Let the binary signal P_s, be rewritten as a concatenation w_1w_2… of words wk in a way that $w_1 = s_1$ and w_{k+1} is the shortest word that has not appeared before; that is, w_{k+1} is the extension of some word w_j in the list, $w_{k+1} = w_j s$, where $0 \le j \le k$, and s is either 0 or 1. For example, 10101010 and 11110000 are parsed as 1.0.10.10.10, and 1.11.10.00.0 and the number of distinct patterns are 3 and 5.

Now let $c(n)$ denote the number of distinct patterns after parsing of the signal P_s. The LZ measure is defined as

$$LZ_n = c(n)[\log_2\{c(n)\}+1]/n \ , \qquad (3)$$

which is an estimate of the total number of samples based on the total number of distinct patterns. Clearly the LZ_n estimate obtained by Eq. 3, is affected by the number of samples, n; this is has been mathematically proven by Hu *et al*., [7].

Normalized Lempel-Ziv Complexity

Following the method of Hu *et al*., [6], we propose a normalized LZ^N measure with the form

$$LZ_n^N = \frac{LZ_n - LZ_n^C}{LZ_n^R - LZ_n^C} \ , \qquad (4)$$

where LZ^C, and LZ^R are the LZ measure of constant (all 1s or 0s) and random (binary white noise) sequences respectively with the same number of samples, n. The LZ^N is mathematically independent of the number of samples [7].

Data

Following Gao *et al*., [8], we generated ultra-short time series with different complexities using the *Logistic* equation in this form

$$X_{n+1} = \alpha X_n(1 - X_n) \ , \qquad (5)$$

where $X_0 = 0.307$, and α is 2, 3.3, and 3.7, for stable, periodic, and chaotic time series respectively [7]. Five sets with 10, 50, 100, 150, and 200 samples were generated for stable, periodic and chaotic time series resulting in total fifteen sets of data.

RESULTS

Table 1, 2, and 3, summarize the LZ measure of stable, periodic, and chaotic time series parsed with the Algorithms 1 and 2, and computed using original and normalized versions. The OA1 and NA1 denote the measures obtained using the original and the normalized LZ definitions (Eq.3 and Eq.4), respectively parsed using the ***Algorithm 1***. The OA2 and NA2 denote the measures computed using the original and the normalized LZ definitions, respectively parsed using the ***Algorithm 2***.

The LZ measures that are not in the interval [0,1] are not valid [4]. These invalid values arise from an insufficient number of samples.

Table 1: Results for stable time series.

	10	50	100	150	200
OA1	1.4	1.3	0.4	0.3	0.2
OA2	1.0	0.5	0.3	0.2	0.2
NA1	1.1	1.1	0.2	0.2	0.2
NA2	1.0	0.2	0.2	0.2	0.2

Table 2: Results for periodic time series.

	10	50	100	150	200
OA1	1.6	1.4	0.6	0.4	0.4
OA2	1.0	0.6	0.5	0.4	0.3
NA1	1.3	1.1	0.4	0.3	0.3
NA2	1.0	0.3	0.3	0.3	0.3

Table 3: Results for chaotic time series.

	10	50	100	150	200
OA1	1.8	1.6	0.9	0.7	0.6
OA2	1.0	0.9	0.8	0.6	0.6
NA1	1.6	1.3	0.6	0.6	0.6
NA2	1.0	0.6	0.6	0.6	0.6

DISCUSSIONS AND CONCLUSIONS

The results show that the LZ measure is increased for the chaotic time series compared to the stable and periodic time series. This is consistent with greater irregularity and complexity within the chaotic time series compared to the stable and periodic time series.

The results show that the ***Algorithm 1*** requires more samples compared to the ***Algorithm 2*** when chaotic time series are being analyzed, otherwise, the results of ***Algorithm 1*** are not valid, as seen in the case of 10 and 50 samples. This effect is not as severe when analyzing stable and periodic time series. However, with 10 samples it is difficult to characterize the underlying dynamics of each time series.

The effects of the number of samples on the measure computed using the original LZ (Eq.3) reinforce the necessity to use the normalized version when a number of time series are being compared which do not have equal number of samples.

To increase the information transfer rate, it is advantageous to extract information from a minimum number of spikes. The number of spikes in a sample also varies within different trials. This suggests that the combination of the normalized LZ measure (Eq.4) along with the ***Algorithm 2*** is better suited for the analysis of neural spikes compared to the original LZ (Eq.3) and ***Algorithm 1***.

Finally, the results show that the LZ measure could be used as a complementary measure along with firing rate for the analysis of ultra-short neural spikes.

REFERENCES

[1] R.A. Andersen, S. Musallam and B. Pesaran (2004) Selecting the signals for a brain-machine interface, *Current Opinion in Neurobiology*, 14, 720-726.

[2] S. Musallam, B.D. Corneil, B. Greger, H. Scherberger and R.A. Andersen (2004) Cognitive control signals for neural prosthetics, *Science*, 305, 258-262.

[3] J.W. Kable, and P.W. Glimcher (2009) The neurobiology of decision: consensus and controversy, *Neuron*, 6, 733-745.

[4] P. Dayan, and L.F. Abbott (2001) *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, MIT Press.

[5] Lempel and J. Ziv (1976) On the complexity of finite sequences, *IEEE Trans. Inf. Theory*, IT-22, 75–81.

[6] M. Aboy, R. Hornero, D. Abásolo, and D. Álvarez (2006) Interpretation of the lempel-ziv complexity measure in the context of biomedical signal analysis, *IEEE Trans. Biomed. Eng.*, 53, 2282–2288.

[7] J.B. Gao, C. Yinhe, W. Ten, and J. HU (2007) Multiscale analysis of complex time series. New Jersey: Wiley Press.

[8] J. Hu, J. Gao, and C. Principe (2006) Analysis of biomedical signals by the Lempel-Ziv complexity: the effect of finite data size *IEEE Trans. Biomed. Eng.*, 53, 12.

[9] D. Abásolo, R. Hornero, C. Gómez, M. García, and M. López (2006) Analysis of EEG background activity in Alzheimer's disease patients with Lempel-Ziv complexity and central tendency measure, *Med. Eng. Phys.*, 28, 315–322.

[10] R. Nagarajan (2002) Quantifying physiological data with Lempel-Ziv complexity-certain issues, *IEEE Trans. Biomed. Eng.*, 49, 1371–1373.

[11] X. S. Zhang, R. J. Roy, and E. W. Jensen (2001) EEG complexity as a measure of depth of anesthesia for patients," *IEEE Trans. Biomed. Eng.*, 48, 1424–1433.