Physics Letters A 382 (2018) 288–299

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

A new similarity index for nonlinear signal analysis based on local extrema patterns

Hamid Niknazar [a], Ali Motie Nasrabadi [b,*], Mohammad Bagher Shamsollahi [c]

[a] Department of Biomedical Engineering, Science and Research Branch Islamic Azad University, Tehran, Iran
[b] Department of Biomedical Engineering, Shahed University, Tehran, Iran
[c] Department of Biomedical Engineering, Sharif University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 7 August 2017
Received in revised form 16 November 2017
Accepted 18 November 2017
Available online 22 November 2017
Communicated by C.R. Doering

Keywords:
Similarity measure
Nonlinear signal
Symbolic technique
Fuzzifying
Classification

ABSTRACT

Common similarity measures of time domain signals such as cross-correlation and Symbolic Aggregate approximation (SAX) are not appropriate for nonlinear signal analysis. This is because of the high sensitivity of nonlinear systems to initial points. Therefore, a similarity measure for nonlinear signal analysis must be invariant to initial points and quantify the similarity by considering the main dynamics of signals. The statistical behavior of local extrema (SBLE) method was previously proposed to address this problem. The SBLE similarity index uses quantized amplitudes of local extrema to quantify the dynamical similarity of signals by considering patterns of sequential local extrema. By adding time information of local extrema as well as fuzzifying quantized values, this work proposes a new similarity index for nonlinear and long-term signal analysis, which extends the SBLE method. These new features provide more information about signals and reduce noise sensitivity by fuzzifying them. A number of practical tests were performed to demonstrate the ability of the method in nonlinear signal clustering and classification on synthetic data. In addition, epileptic seizure detection based on electroencephalography (EEG) signal processing was done by the proposed similarity to feature the potentials of the method as a real-world application tool.

1. Introduction

Signal classification and time series mining have attracted an increasing interest due to their wide applications in industry, medicine, biology, finance, etc. There are a number of general approaches in signal classification. In a common approach, first, features are extracted from signals or their representations and then a decision-making system such as Artificial Neural Network, Fuzzy system or any classifier classifies the signals. Examples of signal representations include Discrete Fourier Transform (DFT) [1], Discrete Wavelet Transform (DWT) [2], Singular Value Decomposition (SVD) [3] and symbolic techniques like Symbolic Aggregate approXimation (SAX) [4]. Similarity indexes and Distance measures are different approaches for time series classification and mining. In these techniques, signals or their representations are used in a main formula to quantify similarity or dissimilarity between the two signals and classify them. Dynamical similarity index [5], Fuzzy similarity index [6], Statistical behavior of local extrema (SBLE) [7] and SAX are examples of this approach.

SAX has become a major similarity index in time series mining. SAX discretizes signals, reduces dimensionality of data and its distance has a lower bound to the Euclidean distance [8]. SAX has been used in mobile data management [9], financial investment [10] shape discovery [11], biomedical signal processing [12] and many other applications. There are some extensions to the SAX method. The ESAX representation overcomes some of the SAX limitations by tripling the dimensions of the original SAX [13]. ESSVS changes SAX distance to cosine distance [8] and SAX-TD improves SAX by considering trends in segments and changing distance measure [14].

However, there is a major difficulty in using SAX and its extensions in the context of nonlinear signal analysis. The difficulty refers to the nature of the SAX method, which was designed to measure similarity of signals in time domain like cross-correlation. However, in nonlinear and especially chaotic signal analysis, dynamics of systems and signals play the major role. Nonlinear and chaotic systems are highly sensitive to initial points, and their signals with different initial points and the same parameters (same dynamics) may appear uncorrelated. Therefore, a similarity index

* Corresponding author.
E-mail addresses: hd.niknazar@gmail.com (H. Niknazar), nasrabadi@shahed.ac.ir (A. Motie Nasrabadi), mbshams@sharif.edu (M.B. Shamsollahi).

https://doi.org/10.1016/j.physleta.2017.11.022
0375-9601/

for nonlinear signal analysis must compare signals by considering their dynamics, not the appearance of the signals. SBLE similarity index which was introduced by the authors is a symbolic technique to compare dynamics of signals based on statistical analysis of symbolized local extrema and aims to develop a symbolic technique for nonlinear analysis [7], [15]. The SBLE method uses local extrema points amplitude that are discretized into some intervals to make a string of symbols. Then the distribution of some predefined patterns construct a feature vector. The cosine distance calculates the similarity of two feature vectors. However, there are some problems in use of SBLE in real world-applications: 1) The method uses local extrema with crisp boundary of amplitude that cause high sensitivity to noise. 2) The information of main frequency is missed because time distance of local extrema has no role in the method. 3) The boundary positions of amplitude are unknown.

By considering these problems, this paper proposes a similarity index for nonlinear signal mining and classification by extending SBLE similarity index. The method uses local extrema of amplitude values instead of fixed segmentation of time domain and average value (e.g. SAX). Also, fuzzy boundary is used to avoid the problem of crisp boundary, and statistical behavior of sequence of local extrema occurrence in time and amplitude is used to characterize dynamics of signals.

A similarity index for nonlinear signal analysis needs to have two main characteristics: sensitivity to changing dynamics (changing of parameters) and insensitivity to initial points. Therefore, to demonstrate the ability of the method some chaotic systems such as Lorenz and Mackey Glass are used. Furthermore, the clustering and classification task on nonlinear signals are performed by using the proposed similarity index. To evaluate the proposed method as a tool in real-world applications, EEG signal classification is done, and the results are presented.

The rest of this paper is organized as follows: Section 2 introduces the proposed similarity index. Section 3 presents the results of the method in analyzing and classifying nonlinear signals. Finally, section 4 concludes the paper.

2. Material and method

The proposed method uses time and amplitude information of local extrema to characterize dynamics of signals. The method consists of five steps to measure similarity of signals: finding local extrema, finding optimum amplitude and difference time intervals, fuzzifying values by using membership functions, extracting sequential information, and measuring similarity. Fig. 1 shows these five steps.

Finding local extrema is the first step of calculating the proposed similarity index (Fig. 1-b). Local extrema have information about amplitude and global frequency of signals and are considered as down-sampled versions of signals. Finding local extrema is highly sensitive to noise, especially to high-frequency noises. Thereby, an efficient noise removal approach is needed in practice to decrease computation time. After finding local extrema amplitudes and time distances (time difference of each two sequential local extrema), these points must be divided into some intervals.

2.1. Amplitude and time distance segmentation

The method proposes an approach to find optimum time and amplitude intervals to maximize accessible information from local extrema. To maximize accessible information of amplitude and time of local extrema, the entropy of these values must be maximized. Equation (1) is used as an entropy measure and must be maximized.

$$\text{Entropy} = \frac{1}{n}\sum_{i=1}^{n} p_i log(p_i) \tag{1}$$

where p_i is the probability of occurrence of i-th local extremum in one of the amplitude and time distance intervals. To maximize entropy value, probability of each local extrema must be the same. If the distribution probability of values considered as uniform distribution, the number of local extrema in each interval must be the same. Therefore, histograms of amplitude and time distance of local extrema are divided into $M+1$ and $N+1$ segments, respectively, as all of those have the same area (Fig. 1-c to f). M and N boundaries of these intervals that maximize extractable information are used in fuzzifying amplitude and time distance values. In this study, M and N are selected empirically.

2.2. Fuzzifying amplitude and time distance

The SBLE, SAX and its extension methods divide amplitude into intervals by crisp values. Using crisp boundaries causes a high sensitivity to noise and small changes also may affect the efficiency of similar approaches. Therefore, fuzzy boundaries of amplitude and time distance of local extrema are used in the proposed method. M and N boundaries of $M+1$ and $N+1$ intervals of amplitude and time distance that are selected from previous step are used to define membership functions (mf) of fuzzifier. Types of membership functions can be selected by considering the distribution of values. This means that to maximize extracted information, membership functions can be selected as histogram of values similar to estimation of probability distribution. Fig. 1-g and h show an example of fuzzifying values by using triangular membership functions by Eqs. (2)–(4).

$$mf_{Fi}(x) = \begin{cases} 0, & \text{if } \frac{L_{i-2}+L_{i-1}}{2} > x, \text{ where } L_0 = L_1 - \frac{L_2-L_1}{2} \\ \frac{2x-L_{i-1}-L_{i-2}}{L_i-L_{i-2}}, & \text{if } \frac{L_{i-2}+L_{i-1}}{2} \le x < \frac{L_i+L_{i-1}}{2} \\ 1-\frac{2x-L_i-L_{i-1}}{L_{i+1}-L_{i-1}}, & \text{if } \frac{L_i+L_{i-1}}{2} \le x < \frac{L_{i+1}+L_i}{2} \\ 0, & \text{if } \frac{L_i+L_{i+1}}{2} \le x, \text{ where } L_{l+1} = L_l + \frac{L_l-L_{l-1}}{2} \end{cases} \tag{2}$$

$$mf_{F1}(x) = \begin{cases} 1, & \text{if } x < L_1 - \frac{L_2-L_1}{2} \\ 1-\frac{2x-3L_1+L_2}{2L_2-2L_1}, & \text{if } L_1 - \frac{L_2-L_1}{2} \le x \le \frac{L_1+L_2}{2} \end{cases} \tag{3}$$

$$mf_{Fl+1}(x) = \begin{cases} \frac{2x-L_{l-1}-L_l}{2L_l-2L_{l-1}}, & \text{if } \frac{L_l-L_{l-1}}{2} \le x < L_l + \frac{L_l+L_{l-1}}{2} \\ 1, & \text{if } x \ge L_l + \frac{L_l-L_{l-1}}{2} \end{cases} \tag{4}$$

$mf_{F1}(x)$ is the first membership function, $mf_{Fl+1}(x)$ is the last membership function and $mf_{Fi}(x)$ is the i-th membership function where $i = \{2, 3, ..., l\}$. F refers to amplitude (A) or time distance (T), L_i refers to i-th boundary of M or N respect to A or T, l is M or N and x is amplitude or time distance values of local extrema.

After defining the membership function, all local extrema are fuzzified and construct a membership matrix mfm_i for i-th local extremum. Element (o, p) of mfm_i refers to belonging i-th local extremum to mf_{Ao} and mf_{Tp}. mfm_i is defined by Eq. (5).

$$mfm_i = \begin{bmatrix} LEi_{A1,T1} & \dots & LEi_{A1,T(N+1)} \\ \vdots & \ddots & \vdots \\ LEi_{A(M+1),T1} & \dots & LEi_{A(M+1),T(N+1)} \end{bmatrix},$$
$$LEi(Ao, Tp) = mf_{Ao}(Amp(LEi)) * mf_{Tp}(TD(LEi)) \tag{5}$$

where $Amp(LEi)$ is amplitude of i-th local extremum and $TD(LEi)$ is its time distance to the next local extremum. In addition, in the entire proposed method, T-norm product is used as T-norm fuzzy logic.