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.

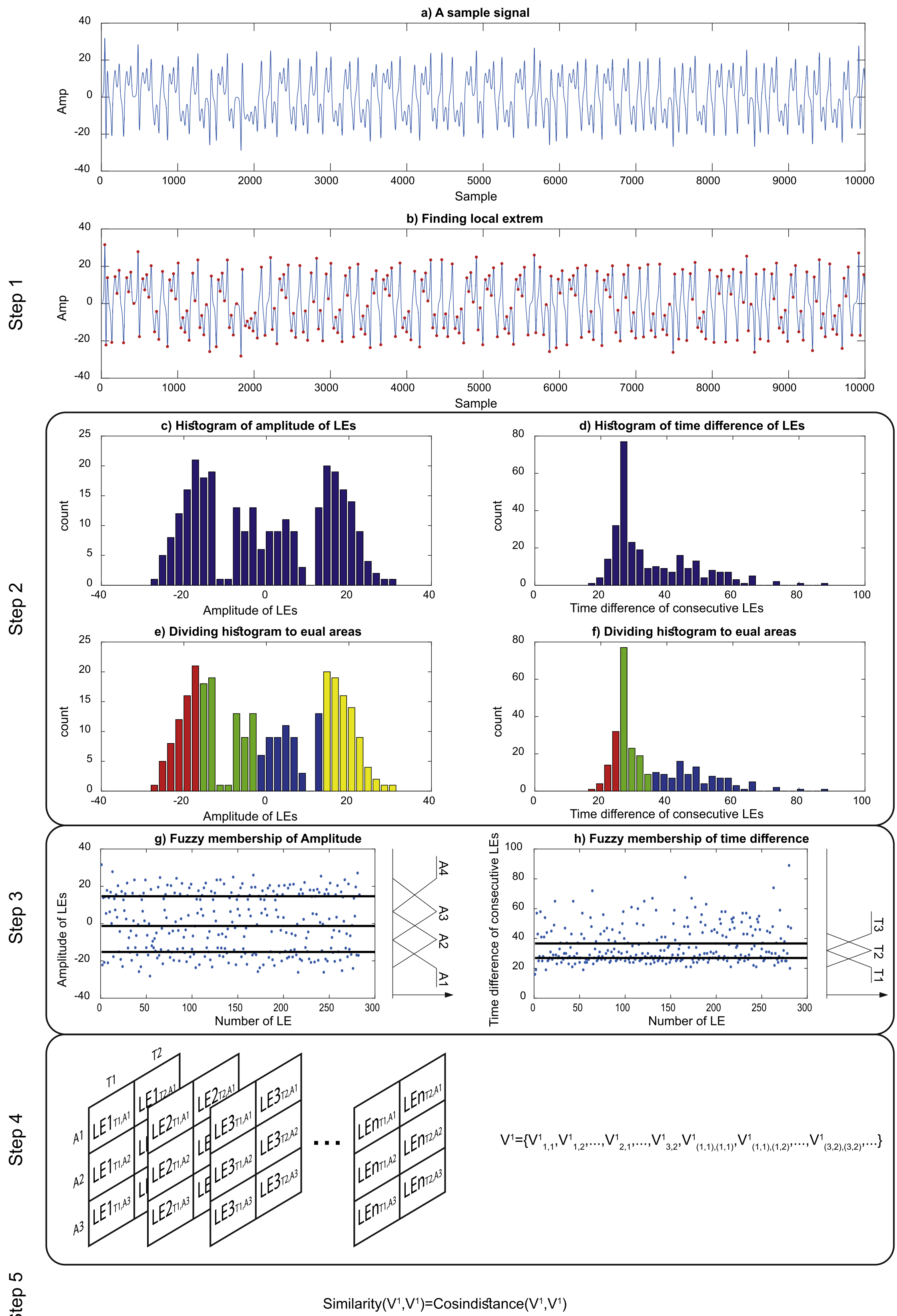

Fig. 1. a) Lorenz system sample signal. b) Step 1, finding local extrema. c), d), e) and f) Step 2, extracting histogram of amplitude of local extrema and difference time of local extrema, dividing the area under the histogram into $M+1$ and $N+1$ segment respectively. g) and h) Step 3, fuzzifying amplitude and time distance of local extrema by using the intervals that are found in step 3. Step 4, constructing V_S^{signal} from one to S sequence of local extrema. Step 5, calculating the similarity of two signals by using cosine distance.

Table 1
Lorenz, Mackey Glass, Tent map and Logistic map system equation and parameters. For each system, a parameter considered as the main parameter (p for Lorenz, a for Mackey Glass, μ for Tent map and r for Logistic map) with six different values, and other parameters are fixed.

Name	Equation	Parameter values
Lorenz [17]	$\frac{dX}{dt} = p(X - Y)$ $\frac{dY}{dt} = XZ + rX - Y$ $\frac{dZ}{dt} = XY - bZ$	$p = \{6, 8.5, 11, 13.5, 16, 18.5\}$ $r = 28$ $b = 8/3$
Mackey Glass [18]	$X(i+1) = X(i) + \frac{aX(i-r)}{1+X(i-r)^c} - bX(i)$	$a = \{5, 5.5, 6, 6.5, 7, 7.5\}$ $b = 0.1$ $c = 10$ $r = 17$
Tent map [19]	$x_{n+1} = \begin{cases} \mu x_n, & \text{if } x_n < \frac{1}{2} \\ \mu(1 - x_n), & \text{if } x_n > \frac{1}{2} \end{cases}$	$\mu = \{0.7, 0.75, 0.8, 0.85, 0.9, 0.95\}$
Logistic map [19]	$x_{n+1} = r(1 - x_n)$	$r = \{3.5, 3.6, 3.7, 3.8, 3.8, 3.9, 4\}$

2.3. Patterns extraction and constructing V_S^{signal} vector

After the step 3, signal is transferred to $n-1$ sequences of mfm matrices, where n is the number of local extrema. Next step is to extract dynamical characteristics by using statistical distribution of the defined patterns. Each pattern quantifies belonging number of s sequential local extrema to possible amplitude and time distance intervals using Eq. (6).

$$V_{(a1,b1)_1,\ldots,(as,bs)_s} = \frac{1}{n-s}\Big[\sum_{i=1}^{n-s} mfm_i(a1, b1) * \ldots * mfm_{i+s}(as, bs)\Big]^{\frac{1}{n-s}} \tag{6}$$

And the numbers of possible patterns for the number of s sequences is:

$$\sharp(s) = ((M+1).(N+1))^s \tag{7}$$

Vector V_S^{signal} is constructed by extracting all patterns values for $s = 1, 2, \ldots, S$ (Eq. (8)).

$$V_S^{signal} = \{V_{(1,1)}, \ldots, V_{(M+1,N+1)}, V_{(1,1),(1,1)}, V_{(1,1),(1,2)}, \ldots, V_{(M+1,N+1),(M+1,N+1)}, \ldots\} \tag{8}$$

This vector has dynamical information of sequential local extrema and will be used in similarity measurement.

2.4. Measuring similarity

Distance between two V_S signals is considered as the similarity measure. Using a bounded distance measure causes more comparability between similarity values in different studies. Therefore, cosine distance that is bounded between zero and one is used as the similarity measure to quantify dynamical similarity of two signals by using their V_S vectors (Eq. (9)).

$$Similarity(V_S^1, V_S^2) = \frac{\langle V_S^1, V_S^2\rangle}{\|V_S^1\|.\|V_S^2\|} \tag{9}$$

where $\|V\|$is norm of V and $\langle V_S^1, V_S^2\rangle$ is inner product of V_S^1 and V_S^2.

3. Method evaluation and discussion

Four approaches are employed to evaluate the efficiency of the proposed similarity index in nonlinear applications. The first approach evaluates the efficiency of the method as an invariant measure practically. The second and third approaches evaluate the similarity index as a clustering and classification tool of nonlinear signals. In the fourth approach, the method is used as a tool for EEG signal processing to detect epileptic seizures. In all of the experiments, the values of M, N and S are selected empirically.

3.1. Invariant measurement

An initial point invariant measure in nonlinear and chaotic signal analysis has to satisfy two main characteristics [16]:

1. Measure must not change by changing initial points.
2. Measure must change by changing parameters of the system.

An efficient similarity measure for nonlinear signal analysis must have these two properties that means:

1. Similarity value of signals with the same dynamics (the same system parameters) and different initial points must be maximum.
2. Similarity value of signals with different dynamics (different system parameters) must be less than similarity value of signals with the same dynamics.

To evaluate the proposed similarity index as an invariant measure, the signals of some common chaotic systems are used (Table 1). Signals with random initial points in six different presented parameters are generated ten times for each of the systems in Table 1, and the proposed similarity measure for all pairs of signals were obtained for each system. Fig. 2 shows signal of systems with various parameters before normalization (Eq. (10)).

$$x(n)_{Normalized} = \frac{x(n) - \frac{1}{N}\sum_{i=1}^{N} x(n)}{\{\frac{1}{N}\sum_{i=1}^{N}[x(n) - \frac{1}{N}\sum_{i=1}^{N} x(n)]\}^{\frac{1}{2}}} \tag{10}$$

where $x(n)$ is a signal and $x(n)_{Normalized}$ is its normalized version. Fig. 3 shows the box plot of similarity values of signals of the same

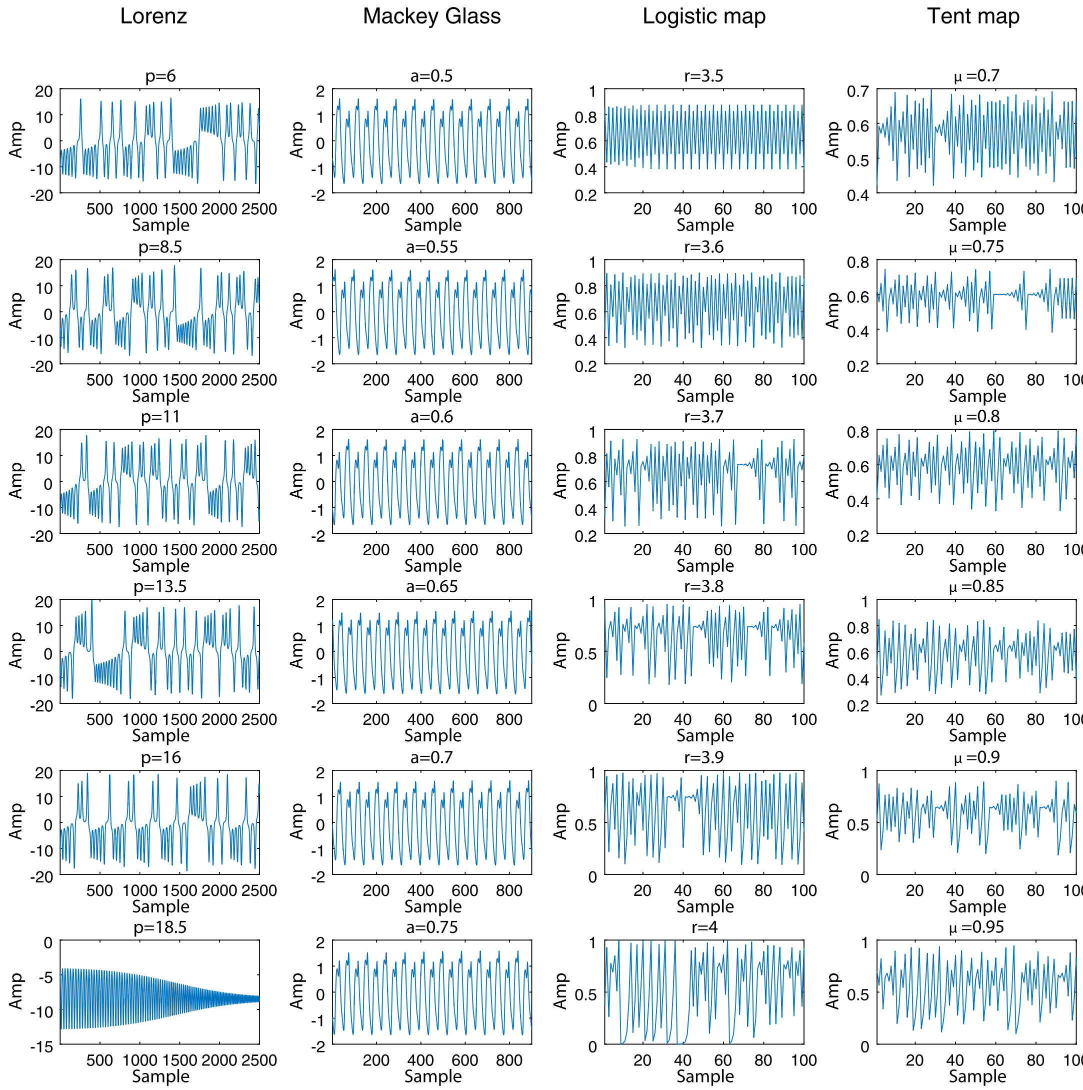

Fig. 2. The sample signals for each system with random initial points and six different parameter values.

system with different parameters and random initial points. As described, similarity measure of signals with the same parameter value must have maximum value (≈ 1) as it can be seen in Fig. 3. Furthermore, all similarity values for signals with different parameters are less than those of the same parameters and by increasing the difference between the two parameters, similarity measures are decreased.

To quantify invariant measuring evaluation, Kruskal–Wallis test [20] is employed. Kruskal–Wallis test is a nonparametric version of classical one-way analysis of variance (ANOVA). Kruskal–Wallis test can be used to determine if there are statistically significant differences between two or more groups of variable. Therefore, it is expected that the p-value of Kruskal–Wallis test of similarity values of signals with the same parameter, and themselves to be one and p-value of Kruskal–Wallis test of similarity values of signals with the same parameter and similarity values of signals with different parameters must be close to zero.

Table 2 reports the p-values of Kruskal–Wallis test. The (i,j) element of each sub-table presents p-values of Kruskal–Wallis test of similarity values of signals with the same parameter (i-th parameter value), and similarity values of signals with i-th parameter value and j-th parameter value. The values of 1 on the diagonal and very small values of the other elements show the ability of the proposed similarity measure as an invariant measure practically.

3.1.1. *Sensitivity to noise*

Using local extrema in a symbolic representation may accentuate the effect of noise and increase computation time. To investigate the effect of noise on the proposed similarity index as an invariant measure, the same four systems are used. White noise with variable power is added to the signals and the similarity between the signals with the same and different parameters, and random initial points are calculated. Fig. 4 shows the variation of these similarity measures with changing the signal to noise ratio (SNR) (Eq. (11)).

$$SNR = 10log\frac{\frac{1}{N}\sum_{i=1}^{N}(S(i))^2}{\frac{1}{N}\sum_{i=1}^{N}(noise(i))^2} \tag{11}$$

where S is clean zero-mean signal and *noise* is zero-mean noise signal.

In Fig. 4, the area between maximum and minimum of similarity values are colored. Narrowness and having no overlap of blue and red areas indicate invariant measurement. By increasing the power of noise that reduces SNR the areas become wider and overlap is increasing. The similarities are separable almost in $SNR > 10$

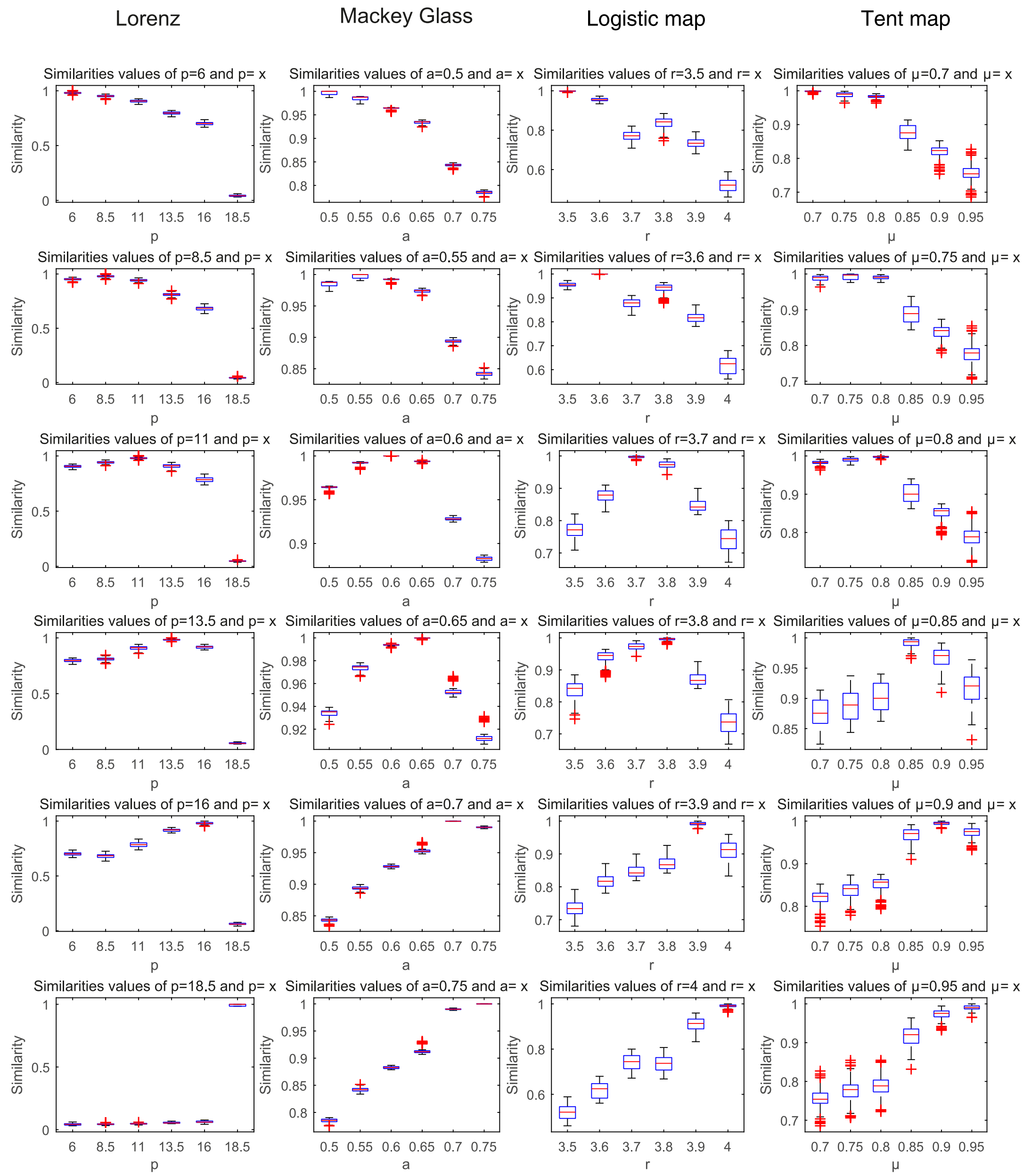

Fig. 3. The box plot of values of similarity between the same and different parameters. i-th row shows similarity of 10 signals with i-th parameter and signals with all six parameters. For each figure x is the six different parameter values of the horizontal axis.

for continuous systems and $SNR > 5$ for maps. Furthermore, the similarity in Logistic map is separable even in $SNR = 0$.

In continuous systems (Lorenz and Mackey Glass) with high SNR the extracted information by the method is related to local extrema that are small pieces of signal and other points have no information. However, in low SNR values, local extrema that are affected from the noise impact on the extracted information. Because of the bigger number of these local extrema than the number of main local extrema, the main information is lost in the noise information, and all signals will be similar and dynamics will be noise dynamics. Nevertheless, in discrete signals such as Tent and Logistic maps, the number of main local extrema is almost equal to the number of samples, so, the effect of noise is less than that of continuous systems.

Adding noise to signals, especially continuous signals, such as the Lorenz causes an increased number of local extrema and subsequently computation time (CT). Fig. 5 shows the computation time for extracting V_s^{signal} from 20 signals with the same parameter and random initial points in different SNR.

In two continuous systems (Mackay Glass and Lorenz), adding noise leads to increase the number of local extrema and CT. Nevertheless, in signals of discrete maps (Logistic and Tent), the number of local extrema is close to the number of samples, so, adding the noise does not effect on CT.

3.2. Nonlinear signal classification and clustering

The proposed similarity measure can be used in nonlinear signal clustering and classification applications. In this subsection, the similarity index is used to classify and cluster 15 chaotic system signals that are described in Appendix A. For each system and parameter, signals are generated ten times with random initial points and are used to evaluate the method.

Table 2
p-Values of Kruskal–Wallis test between the same parameter similarity values and not the same parameter similarity values. Each row i value is the p-values of Kruskal–Wallis test between the same parameter similarity values (i-th parameter value), and similarity values of signals with i-th parameter value and column number parameter value.

Lorenz

1	6.90E−34	2.49E−34	2.49E−34	2.49E−34	1.51E−28
3.02E−31	1	1.58E−33	2.49E−34	2.49E−34	1.51E−28
2.49E−34	2.98E−34	1	2.49E−34	2.49E−34	1.51E−28
2.49E−34	2.49E−34	2.49E−34	1	2.49E−34	1.51E−28
2.49E−34	2.49E−34	2.49E−34	2.49E−34	1	1.51E−28
2.014E−20	2.01E−20	2.014E−20	2.01E−20	2.01E−20	1

Mackey Glass

1	1.11E−32	2.49E−34	2.49E−34	2.49E−34	2.49E−34
2.49E−34	1	1.13E−18	2.49E−34	2.49E−34	2.49E−34
2.49E−34	2.49E−34	1	2.49E−34	2.49E−34	2.49E−34
2.49E−34	2.49E−34	2.49E−34	1	2.49E−34	2.49E−34
2.49E−34	2.49E−34	2.49E−34	2.49E−34	1	2.49E−34
2.49E−34	2.49E−34	2.49E−34	2.49E−34	2.49E−34	1

Logistic map

1	2.49E−34	2.49E−34	2.49E−34	2.49E−34	2.49E−34
2.49E−34	1	2.49E−34	2.49E−34	2.49E−34	2.49E−34
2.49E−34	2.49E−34	1	2.01E−33	2.49E−34	2.49E−34
2.49E−34	2.49E−34	1.90E−31	1	2.49E−34	2.49E−34
2.49E−34	2.49E−34	2.49E−34	2.49E−34	1	2.49E−34
2.49E−34	2.49E−34	2.49E−34	2.49E−34	2.49E−34	1

Tent map

1	1.80E−27	3.57E−34	2.49E−34	2.49E−34	2.49E−34
2.52E−7	1	4.05E−6	2.49E−34	2.49E−34	2.49E−34
2.81E−34	6.81E−27	1	2.49E−34	2.49E−34	2.49E−34
2.49E−34	2.49E−34	2.49E−34	1	2.54E−26	2.49E−34
2.49E−34	2.49E−34	2.49E−34	3.23E−33	1	1.01E−31
2.49E−34	2.49E−34	2.49E−34	2.49E−34	4.91E−24	1

3.2.1. Nonlinear signal clustering

According to Appendix A, for each system there are four clusters with 10 samples. Clustering with K-means algorithm is done for each system by the following steps:

Table 3
The result of using the proposed similarity measure for unsupervised clustering of nonlinear signals.

Name	The number of clusters, samples	Clean accuracy	Noisy, SNR = 10 accuracy
Autonomous 4D Circ	4, 40	0.818 ± 0.0723	0.817 ± 0.0337
Chua Circ	4, 40	0.898 ± 0.0612	0.754 ± 0.0617
Chua Circ CN	4, 40	0.886 ± 0.0598	0.85 ± 0
Colpitts Osc	4, 40	0.714 ± 0 0.0232	0.85 ± 0
RC Colpitts	4, 40	0.975 ± 0.0343	0.975 ± 0
RC Hysteresis	4, 40	0.707 ± 0.0332	0.675 ± 0.0357
RC Nonlin Circ	4, 40	0.705 ± 0.0737	0.889 ± 0.0676
Simple Chaotic Circ	4, 40	0.516 ± 0	0.8 ± 0
Lorenz	4, 40	1 ± 0	0.975 ± 0
Genhaos	4, 40	1 ± 0	1 ± 0
Henon map	4, 40	1 ± 0	1 ± 0
Logistic map	4, 40	1 ± 0	1 ± 0
PWAM map	4, 40	1 ± 0	1 ± 0
Tent map	4, 40	1 ± 0	1 ± 0
Bernoulli map	4, 40	1 ± 0	1 ± 0

1. Finding the amplitude and time distance intervals by using one randomly selected signal.
2. Extracting V_S^{signal} vector for all 40 signals.
3. Selecting four V_S^{signal} vectors as cluster indicators randomly.
4. Measuring the similarities of all other V_S^{signal} vectors to cluster indicators.
5. Assigning each V_S^{signal} vector to a cluster where V_S^{signal} is more similar to its cluster indicator.
6. Updating cluster indicators by averaging on V_A^{signal} vectors of clusters.
7. Repeating step 3 to 5 until cluster indicators do not change.

These steps are performed with $M = 3$, $N = 3$ and $S = 3$ on 40 signals of each system. Furthermore, to evaluate the sensitivity to noise, these steps are repeated for noisy signals where white noise with standard deviation of $\delta = 0.1$ is added to the normalized signals ($SNR = 20$). Table 3 reports the accuracy of this unsupervised clustering.

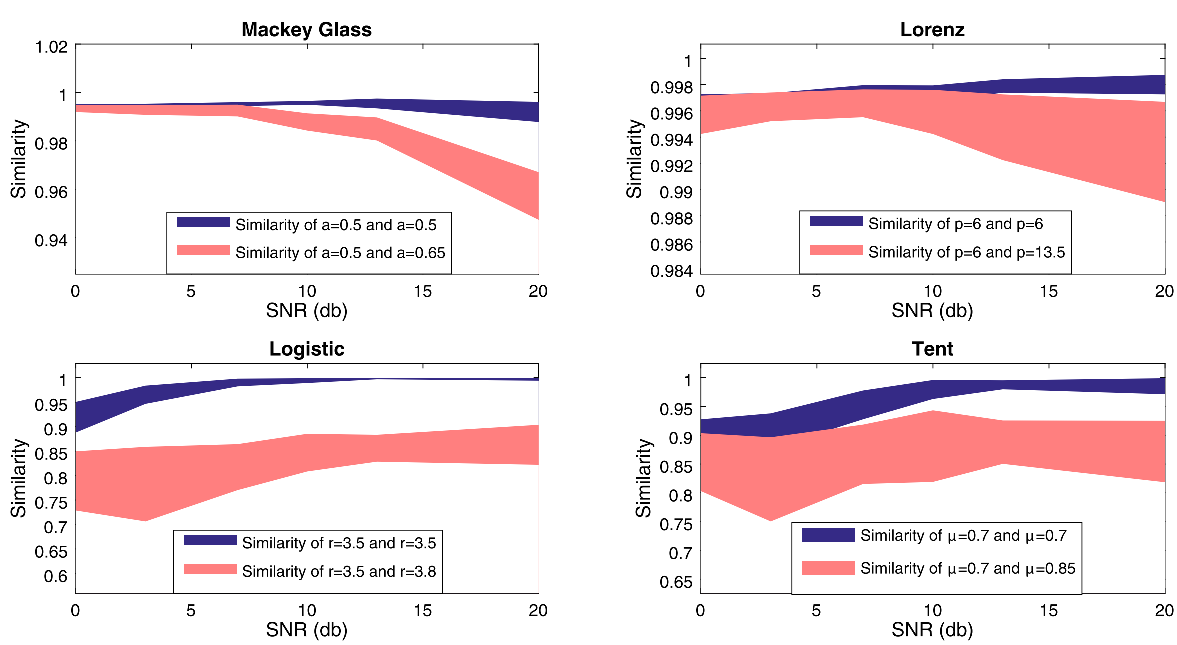

Fig. 4. The effect of adding noise with different SNR to similarity values. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

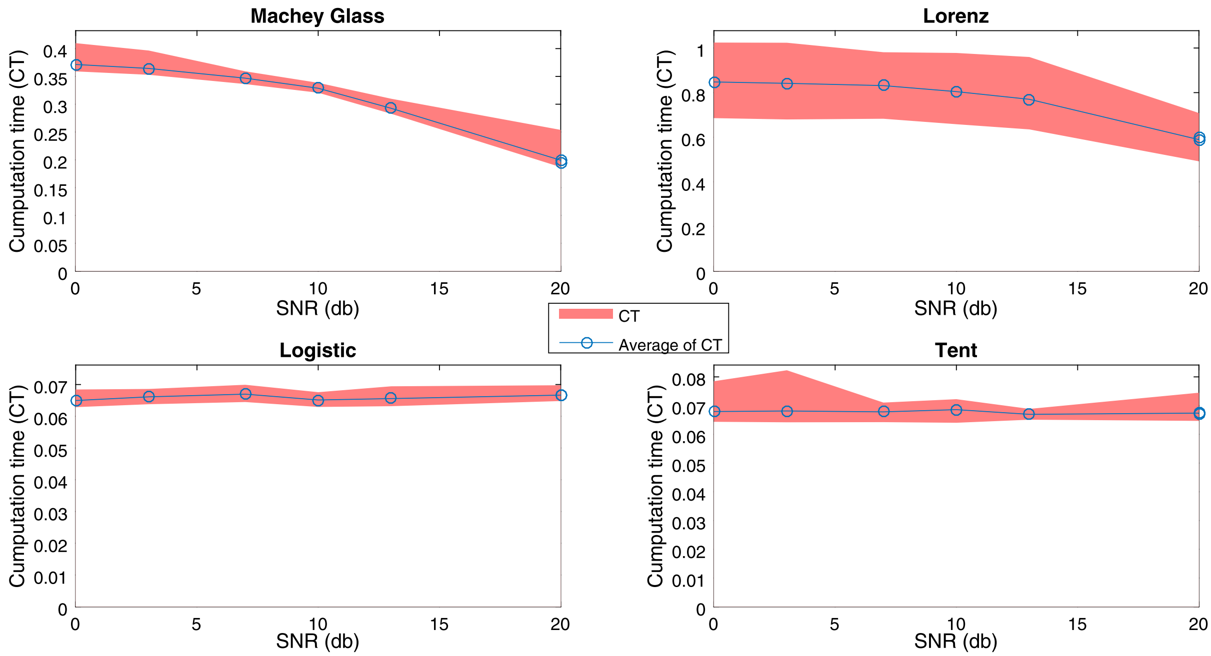

Fig. 5. The effect of adding noise with different SNR to computation time of extracting the similarity.

Table 4
The result of using the proposed similarity measure for supervised classification of nonlinear signals.

	The number of classes	Train-set size	Test-set size	Accuracy
Clean	60	300	300	0.913
Noisy, SNR = 10	60	300	300	0.883

The results in Table 3 show the ability of the proposed similarity measure in making a distinction between signals with different dynamics and as a tool for unsupervised clustering specially for discrete maps. As it can be seen, the presence of the noise does not make a significant effect on performance of clustering. However, because of increasing the number of local extrema computation time will increase. In some case such as “RC Nonlin Circ”, the performance in noisy signals is better, which is because of the random selection of the first cluster indicators and K-means algorithm.

3.2.2. Nonlinear signal classification

In the next evaluation, the similarity index is used in a supervised classification application. For all of the systems in Appendix A in four parameters, signals are generated ten times. Therefore, there are 600 signals and 60 classes. These signals are divided into two sets of 300 samples randomly: train-set and test-set. V_S^{signal} vector is extracted for all signals with $M = 3$, $N = 3$ and $S = 3$. For all V_S^{signal} vectors of the test-set signals the similarities to all V_S^{signal} vectors of the train-set are measured. Then, the label of each signal in the test set is considered as the label of the most similar signal in train-set similar to k-nearest neighbor routine (kNN) with $k = 1$. The results of classification of clean and noisy signals are presented in Table 4.

Similar to clustering evaluation, with a simple procedure, the results show the ability of the method for nonlinear signal classification. Furthermore, adding noise has a small effect on results.

3.3. Computation time

The propose method consists of five steps. The main part of the computation time of the method belongs to the extraction of the V_S^{signal} from the time series, as the steps one to three mainly consist of logical and addition operands. The computation time of extracting the V_S^{signal} from a time series depends on M, N and S values. Changing the values of M, N and S mainly affects the computation time of the step four of the method. The computation time of extracting V_S^{signal}, after the three first steps, can be considered as Eq. (12).

$$Computation\ time_{V_S^{signal}} \propto \sum_{s=1}^{S} (M * N)^s \tag{12}$$

As a experimental test, the computation time of extracting V_S^{signal} (the step four of the method) from a sample Lorenz signal (Fig. 6) with different M, N and S parameters is measured in a computer with a core of i7 2.2 GHz CPU and 8 Gigabyte of RAM. For each set of parameters computation time is measured ten times and average values are presented in Fig. 6.

Fig. 6 shows the huge impact of the S value on the computation time.

3.4. Epileptic seizure prediction

Epilepsy is among the most common neurological disturbances, showing temporary, reversible, and abnormal electrical activity in the brain. It is characterized by occasional, excessive and synchronous discharging of neurons, which can be detected by clinical appearances [21]. Since epilepsy is a condition related to the electrical activity of the brain, it can be studied by analyzing electroencephalogram (EEG) signals. Based on some studies of the community of neurophysiology researchers, EEG signal is multivariate time series that stems from a highly nonlinear and multidimensional system when the brain activity is normal [22]. To evaluate the proposed similarity measure as a tool for real-world application, it is used to classify EEG signals, which is an important task in epileptic seizure detection studies.

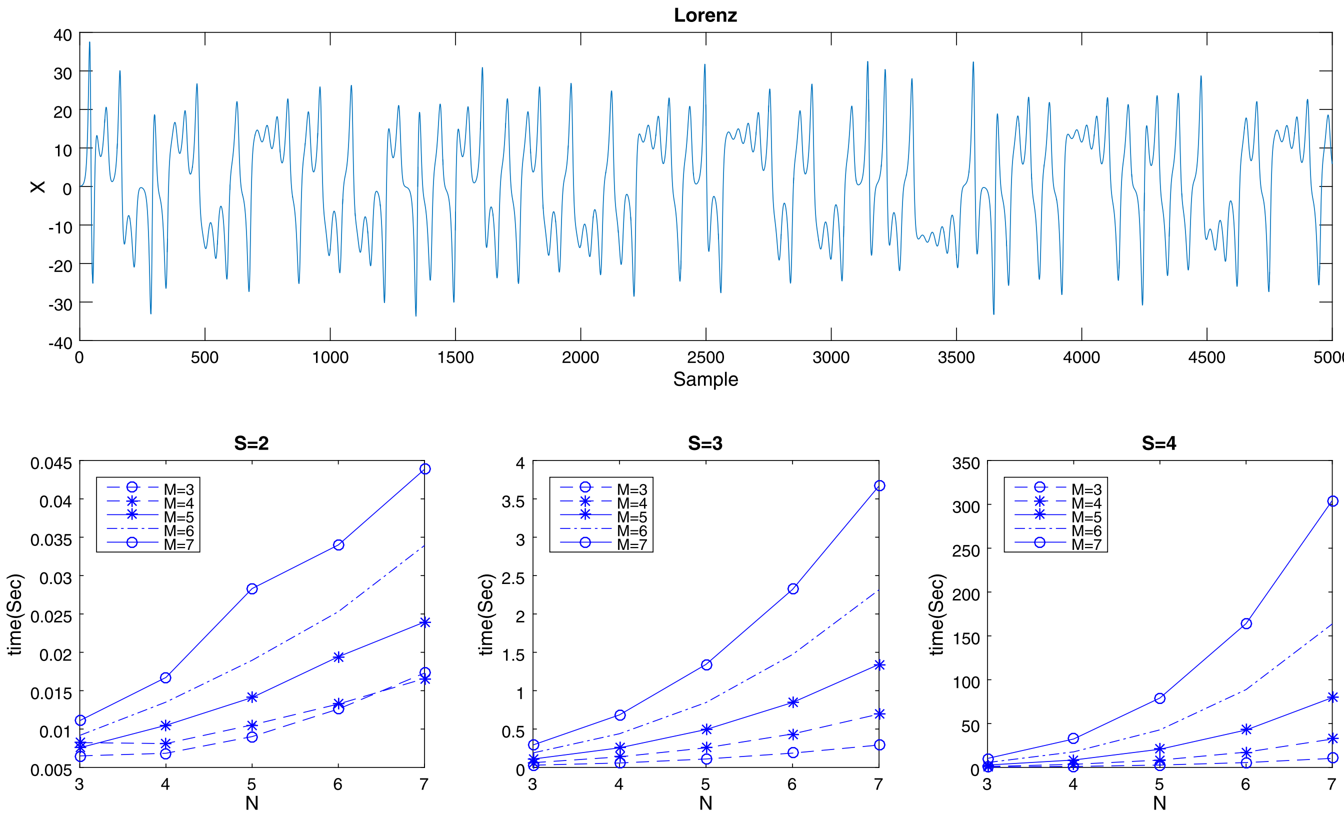

Fig. 6. Average of ten times computation time of extracting V_S^{signal} from a sample Lorenz signal in various M, N and S.

Table 5
Description of five data sets.

Subjects	Set A	Set B	Set C	Set D	Set E
	Five healthy volunteers		Five epileptic patients		
Patient state	Eyes open	Eyes closed	Pre-ictal	Pre-ictal	Ictal
Electrode types	Surface	Surface	Intracranial	Intracranial	Intracranial
Electrode placement	International 10/20 systems	International 10/20 systems	Within epileptogenic zone	Opposite to epileptogenic zone	Within epileptogenic zone
No. of samples	100	100	100	100	100
Sampling points	4096	4096	4096	4096	4096

3.4.1. Dataset

The EEG data used in this study, which is publicly available, were taken from University of Bonn, Germany [23]. The complete database contains five sets denoted as A–E, with 100 samples of 23.6 s duration. The description of five data sets is shown in Table 5. The signals were recorded with the same 128-channel amplifier system and digitized at 173.61 samples per second. Sets C and D denoted as inter-ictal data are recorded during the patients in pre-ictal. Set E, which is called ictal data, contains signals recorded during the epileptic seizure. In this study, five cases including all of the data sets are presented.

3.4.2. Procedure and result

Five different classification combinations are built by the dataset in this subsection as shown in Table 6. This classification problem is very common and used in various papers in the field of epilepsy diagnosis and seizure detection.

For all signals in each case, V_S^{signal} is extracted with $M = 10$, $N = 2$ and $S = 4$ parameters. The K-fold cross validation with $K = 10$ is used to evaluate classification accuracy. After extracting V_S^{signal} for training and testing sets in each fold, similarities of every test signal to all train signals have to be computed and the label of the most similar signal for the train set is assigned to each test signal.

Table 6
Description of five cases.

Cases	Sets	Description
Case 1	Set A, B versus Set C, D versus Set E	Healthy, inter-ictal and ictal
Case 2	Set A versus Set D versus Set E	Healthy, inter-ictal and ictal
Case 3	Set A, B, C, D versus Set E	Non-seizure and seizure
Case 4	Set C versus Set E	Inter-ictal and ictal
Case 5	Set D versus Set E	Inter-ictal and ictal

This procedure may be the simplest way to use the proposed similarity measure as classifier and Table 7 reports the accuracy of this method in comparison with some recent studies that used the same cases.

The result of the method is not the best in comparison with all of the other works. Nevertheless, as Table 7 shows; the proposed method can achieve acceptable accuracy with simple methodology without using any powerful classifier such as ANN and SVM.

The proposed method can be employed in other more complex circumstances. For example, V_S^{Signal} can be considered as

Table 7
Performance comparison with existing methods that used the same data.

Authors	Year	Method	Accuracy (%)
The classification problem of Case 1 AB/CD/E			
Acharya et al. [24]	2012	Wavelet packet decomposition + Gaussian mixture model	99.0
Alam et al. [25]	2013	Empirical mode decomposition + artificial neural network	80.0
Niknazar et al. [26]	2013	RQA in EEG signal and its wavelet-based sub-bands + ECOC	98.67
Riaz et al. [27]	2016	Empirical mode decomposition based temporal and spectral features + SVM	82.5
Das et al. [28]	2016	Dual-tree complex wavelet transform + SVM	96.28
The proposed similarity measure (averag ± standard deviation)			92.2 ± 2.2
The proposed similarity measure + ECOC-SVM (averag ± standard deviation)			98.6 ± 1.3
The classification problem of Case 2 A/D/E			
Acharya et al. [29]	2013	Continuous wavelet transform based high order spectrum and textures + SVM	96.0
Kaya et al. [30]	2014	1-D local binary patterns + BayesNet	95.67
Riaz et al. [27]	2016	Empirical mode decomposition based temporal and spectral features + SVM	84.0
Martis et al. [31]	2015	Wavelet packet decomposition based non-linear features + SVM	98.0
The proposed similarity measure (averag ± standard deviation)			90.0 ± 1.6
The proposed similarity measure + ECOC-SVM (averag ± standard deviation)			99.0 ± 1.6
The classification problem of Case 3 ABCD/E			
Alam et al. [25]	2013	Empirical mode decomposition + artificial neural network	100
Zhu et al. [32]	2014	Weighted horizontal visibility algorithm + K-nearest neighbor	95.4
Kumar et al. [33]	2014	Fuzzy approximate entropy + SVM	97.38
Riaz et al. [27]	2016	Empirical mode decomposition based temporal and spectral features + SVM	95.6
The proposed similarity measure (averag ± standard deviation)			96.8 ± 1.4
The proposed similarity measure + ECOC-SVM (averag ± standard deviation)			99.2 ± 1.0
The classification problem of Case 4 C/E			
Nicolaou et al. [34]	2012	Permutation entropy + SVM	100
Das et al. [28]	2014	Dual-tree complex wavelet transform + SVM	100
Kumar et al. [33]	2014	Fuzzy approximate entropy + SVM	99.60
The proposed similarity measure (averag ± standard deviation)			96 ± 2.1
The proposed similarity measure + ECOC-SVM			100
The classification problem of Case 5 D/E			
Alam et al. [25]	2013	Weighted horizontal visibility algorithm + K-nearest neighbor	100
Zhu et al. [32]	2014	Weighted horizontal visibility algorithm + K-nearest neighbor	95.4
Kumar et al. [33]	2013	Fuzzy approximate entropy + SVM	97.38
Riaz et al. [27]	2016	Empirical mode decomposition based temporal and spectral features + SVM	95.6
The proposed similarity measure (averag ± standard deviation)			95.5 ± 1.6
The proposed similarity measure + ECOC-SVM			100

a feature vector of the signals. Therefore, any other classifier such as ANN and SVM can be applied on these feature vectors. The SVM classifier was also used to classify signals by using V_S^{Signal} vectors as features. SVM is a two-class classifier and needs a method like error correcting output codes (ECOC) [35] to be used as a multi-class classifier. The ECOC technique can be broken down into two distinct stages, encoding and decoding. Given a set of classes, the coding stage designs a code word (a sequence of bits representing each class, where each bit identifies the membership of the class for a given binary classifier) for each class based on different binary problems. The decoding stage makes a classification decision for a given test sample based on the output code [36]. Table 7 also reports the accuracy of using V_S^{Signal} vectors with ECOC-SVM classifier. The accuracy of this method is 100% in two cases and in the other cases is one of the best.

4. Conclusion

In this study, we propose a similarity index based on extracting information from amplitude and time of local extrema. This similarity measure can be used in nonlinear signal processing in classification and clustering problems. Using fuzzy values in the method, the sensitivity to noise was decreased in comparison with the other symbolic methods such as SAX and SBLE. We evaluated the method in practical applications by synthetic nonlinear signals as a core of clustering and classification system. Moreover, by using EEG data it was shown that the proposed similarity measure could be used in real-world application and provided an acceptable result, especially when it is mixed to an efficient classifier.

Acknowledgements

This study was supported by Cognitive Sciences and Technologies Council of Iran according to the contract No. 2688.

Appendix A

The equation and parameters of the nonlinear systems that are used in this study are presented in Table A.8. Each system has some parameters that are set to the fixed values and one variable parameter. The initial point in every run is set randomly.

Table A.8
Name, equation and parameters of 15 nonlinear systems.

Name	Equation	Parameter values
Autonomous 4D Circ	$\frac{dX}{dt} = \alpha(X + \gamma Y) - Z - n_x$ $\frac{dY}{dt} = \beta(\delta(X + \gamma Y)) - W - \gamma n_y$ $\frac{dZ}{dt} = X$ $\frac{dW}{dt} = \alpha Y$ $n_x = \frac{\lvert X-1\rvert + X - 1}{2\epsilon_1}$ $n_y = \frac{\lvert Y-1\rvert + Y - 1}{2\epsilon_2}$	$\alpha = 2$ $\beta = \{0.15, 0.16, 0.17, 0.18\}$ $\gamma = 3$ $\delta = 0.27$ $\epsilon_1 = 0.8$ $\epsilon_2 = 0.2$
Chua Circ	$\frac{dX}{dt} = -\alpha X + \alpha Y - \alpha n$ $\frac{dY}{dt} = X - Y + Z$ $\frac{dZ}{dt} = -\beta Y$ $n = aX + 0.5(a-b)(\lvert X+1\rvert - \lvert X-1\rvert)$	$\alpha = \{8, 8.5, 9, 9.5\}$ $\beta = 15$ $a = -1.3$ $b = -0.7$
Chua Circ CN	$\frac{dX}{dt} = -\alpha X + \alpha Y - \alpha n$ $\frac{dY}{dt} = X - Y + Z$ $\frac{dZ}{dt} = -\beta Y$ $n = aX + bX^3$	$\alpha = \{8.6, 9.05, 9.5, 9.95\}$ $\beta = 15$ $a = -1.3$ $b = 0.07$
Colpitts Osc	$\frac{dX}{dt} = Z - \beta n$ $\frac{dY}{dt} = \alpha(-r(Y + \gamma) - Z - n)$ $\frac{dZ}{dt} = \delta(-X + Y + \epsilon) - \rho Z$ $n = \begin{cases} 0, & \text{if } Y \le 1 \\ 1, & \text{if } Y > 1 \end{cases}$	$\alpha =$ $\{0.05, 0.126, 0.202, 0.278\}$ $\beta = 15$ $a = -1.3$ $b = -0.7$
RC Colpitts	$\frac{dX}{dt} = kk_1(X - Z) - X$ $\frac{dY}{dt} = 2k_2 n$ $\frac{dZ}{dt} = 2k_1(X - Z) - 2k_2 n$ $n = \begin{cases} Z - Y, & \text{if } Z - Y \le 1 \\ 1, & \text{if } Y > 1 \end{cases}$	$k = \{1.75, 1.82, 1.89, 1.96\}$ $k_1 = 11$ $k_2 = 0.9$
RC Hysteresis	$\frac{dX}{dt} = -X - Y$ $\frac{dY}{dt} = (2\delta + 2)X + (2\delta + 2)Y$ $\quad + p * state$ $state = \begin{cases} 1, & \text{if } Y \le -1 \& state = -1 \\ -1, & \text{if } Y \ge 1 \& state = 1 \end{cases}$	$\delta = \{0.005, 0.013$ $0.021, 0.029\}$ $p = 1$
RC Nonlin Circ	$\frac{dX}{dt} = -2X + Y + n$ $\frac{dY}{dt} = X - 2Y + Z$ $\frac{dY}{dt} = Y - Z$ $n = m_b Z + 0.5(m_a - m_b)(\lvert Z+1\rvert - \lvert Z-1\rvert)$	$m_b = \{150, 198, 246, 294\}$ $m_a = -33.03$
Simple Chaotic Circ	$\frac{dX}{dt} = Y - n$ $\frac{dY}{dt} = -\beta(X + Y) + Fsin(\omega t)$ $n = bX + 0.5(a-b)(\lvert X+1\rvert - \lvert X-1\rvert)$	$F = \{0.25, 0.4, 0.55, 0.7\}$ $\beta = 1$ $\omega = 0.6$ $a = -1.27$ $b = -0.68$

Table A.8 (*continued*)

Name	Equation	Parameter values
Lorenz	$\frac{dX}{dt} = \sigma(Y - X)$ $\frac{dY}{dt} = \rho X - Y - XZ$ $\frac{dZ}{dt} = XY - \beta Z$	$\sigma = \{8, 10.5, 13, 15.5\}$ $\beta = 8/3$ $\rho = 28$
Genhaos	$X(n) = c_1 X(n-1) + c_2 X(n-2) + c_3 X(n-3)$	$c_1 = \{1.5, 2, 2.5, 3\}$ $c_2 = 1$ $c_3 = 1$
Henon map	$X(n) = 1 - \alpha X(n-1)^2 + \beta X(n-2)$	$\alpha = \{1.1, 1.2, 1.3, 1.4\}$ $\beta = 0.3$
Logistic map	$X(n) = rX(n-1)(1 - X(n-1))$	$r = \{3.5, 3.66, 3.82, 3.98\}$
PWAM map	$X(n) = \begin{cases} B\lvert X(n-1)\rvert, & \text{if } \lvert X(n-1)\rvert \le D \\ -B(\lvert X(n-1)\rvert - 2D), & \text{if } \lvert X(n-1)\rvert > D \end{cases}$	$B = \{1.5, 1.9, 2.3, 2.7\}$ $D = 1$
Tent map	$X(n) = \mu(1 - \lvert 1 - 2X(n-1)\rvert)$	$\mu = \{0.7, 0.78, 0.86, 0.94\}$
Bernoulli map	$X(n) = (\mu X(n-1)) \setminus 1$	$\mu = \{1.5, 1.63, 1.76, 1.89\}$

References

[1] C. Faloutsos, M. Ranganathan, Y. Manolopoulos, Fast subsequence matching in time-series databases, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, 1994, pp. 419–429.

[2] K. Chan, A. Fu, Efficient time series matching by wavelets, in: Proceedings of the IEEE International Conference on Data Engineering, 1999, pp. 126–133.

[3] K. Ravi Kanth, D. Agrawal, A. Singh, Dimensionality reduction for similarity searching in dynamic databases, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, 1998, pp. 166–176.

[4] J. Lin, E. Keogh, S. Lonardi, B. Chiu, A symbolic representation of time series, with implications for streaming algorithms, in: Proceedings of the ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 2003, pp. 2–11.

[5] M. Le Van Quyen, J. Martinerie, V. Navarro, P. Boon, M. D'Havé, C. Adam, B. Renault, F. Varela, M. Baulac, Anticipation of epileptic seizures from standard EEG recordings, Lancet 357 (9251) (2001) 183–188.

[6] G. Ouyang, X. Li, X. Guan, Use of fuzzy similarity index for epileptic seizure prediction, in: Fifth World Congress on Intelligent Control and Automation, vol. 6, WCICA 2004, IEEE, 2004, pp. 5351–5355.

[7] H. Niknazar, A.M. Nasrabadi, Epileptic seizure prediction using a new similarity index for chaotic signals, Int. J. Bifurc. Chaos 26 (11) (2016) 1650186.

[8] J. Lin, E. Keogh, L. Wei, S. Lonardi, Experiencing sax: a novel symbolic representation of time series, Data Min. Knowl. Discov. 15 (2) (2007) 107–144.

[9] H. Tayebi, S. Krishnaswamy, A.B. Waluyo, A. Sinha, M.M. Gaber, RA-SAX: resource-aware symbolic aggregate approximation for mobile ECG analysis, in: 2011 IEEE 12th International Conference on Mobile Data Management, vol. 1, IEEE, 2011, pp. 289–290.

[10] A. Canelas, R. Neves, N. Horta, A new SAX-GA methodology applied to investment strategies optimization, in: Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation, ACM, 2012, pp. 1055–1062.

[11] T. Rakthanmanon, E. Keogh, Fast shapelets: a scalable algorithm for discovering time series shapelets, in: Proceedings of the 13th SIAM International Conference on Data Mining, SIAM, 2013, pp. 668–676.

[12] G. Balakrishnan, A. Shoeb, Z. Syed, Creating symbolic representations of electroencephalographic signals: an investigation of alternate methodologies on intracranial data, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010, pp. 4683–4686.

[13] B. Lkhagva, Y. Suzuki, K. Kawagoe, New time series data representation ESAX for financial applications, in: 22nd International Conference on Data Engineering Workshops, ICDEW'06, 2006.

[14] Y. Sun, J. Li, J. Liu, B. Sun, C. Chow, An improvement of symbolic aggregate approximation distance measure for time series, Neurocomputing 138 (2014) 189–198.

[15] H. Niknazar, K. Maghooli, A.M. Nasrabadi, Epileptic seizure prediction using statistical behavior of local extrema and fuzzy logic system, Int. J. Comput. Appl. 113 (2) (2015), https://doi.org/10.5120/19799-1578.

[16] R.C. Hilborn, Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers, Oxford University Press on Demand, 2000.

[17] E.N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. 20 (2) (1963) 130–141.

[18] M.C. Mackey, L. Glass, et al., Oscillation and chaos in physiological control systems, Science 197 (4300) (1977) 287–289.

[19] P.C.-J. Eckmann, Iterated Maps on the Interval as Dynamical Systems, Progress in Physics, vol. 1, 1980.

[20] W.H. Kruskal, W.A. Wallis, Use of ranks in one-criterion variance analysis, J. Am. Stat. Assoc. 47 (260) (1952) 583–621.

[21] T.R. Browne, G.L. Holmes, Handbook of Epilepsy, Jones & Bartlett Learning, 2008.

[22] N. Thomasson, T.J. Hoeppner, C.L. Webber, J.P. Zbilut, Recurrence quantification in epileptic EEGs, Phys. Lett. A 279 (1) (2001) 94–101.

[23] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state, Phys. Rev. E 64 (6) (2001) 061907.

[24] U.R. Acharya, S.V. Sree, A.P.C. Alvin, J.S. Suri, Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework, Expert Syst. Appl. 39 (10) (2012) 9072–9078.

[25] S.S. Alam, M.I.H. Bhuiyan, Detection of seizure and epilepsy using higher order statistics in the EMD domain, IEEE J. Biomed. Health Inform. 17 (2) (2013) 312–318.

[26] M. Niknazar, S. Mousavi, B.V. Vahdat, M. Sayyah, A new framework based on recurrence quantification analysis for epileptic seizure detection, IEEE J. Biomed. Health Inform. 17 (3) (2013) 572–578.

[27] F. Riaz, A. Hassan, S. Rehman, I.K. Niazi, K. Dremstrup, EMD-based temporal and spectral features for the classification of EEG signals using supervised learning, IEEE Trans. Neural Syst. Rehabil. Eng. 24 (1) (2016) 28–35.

[28] A.B. Das, M.I.H. Bhuiyan, S.S. Alam, Classification of EEG signals using normal inverse Gaussian parameters in the dual-tree complex wavelet transform domain for seizure detection, Signal Image Video Process. 10 (2) (2016) 259–266.

[29] U.R. Acharya, R. Yanti, J.W. Zheng, M.M.R. Krishnan, J.H. Tan, R.J. Martis, C.M. Lim, Automated diagnosis of epilepsy using CWT, hos and texture parameters, Int. J. Neural Syst. 23 (03) (2013) 1350009.

[30] Y. Kaya, M. Uyar, R. Tekin, S. Yıldırım, 1d-local binary pattern based feature extraction for classification of epileptic EEG signals, Appl. Math. Comput. 243 (2014) 209–219.

[31] R.J. Martis, J.H. Tan, C.K. Chua, T.C. Loon, S.W.J. Yeo, L. Tong, Epileptic EEG classification using nonlinear parameters on different frequency bands, J. Mech. Med. Biol. 15 (03) (2015) 1550040.

[32] G. Zhu, Y. Li, P.P. Wen, Epileptic seizure detection in EEGs signals using a fast weighted horizontal visibility algorithm, Comput. Methods Programs Biomed. 115 (2) (2014) 64–75.

[33] Y. Kumar, M. Dewal, R. Anand, Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine, Neurocomputing 133 (2014) 271–279.

[34] N. Nicolaou, J. Georgiou, Detection of epileptic electroencephalogram based on permutation entropy and support vector machines, Expert Syst. Appl. 39 (1) (2012) 202–209.

[35] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[36] S. Escalera, O. Pujol, J. Mauri, P. Radeva, Intravascular ultrasound tissue characterization with sub-class error-correcting output codes, J. Signal Process. Syst. 55 (1–3) (2009) 35–47.