



Research article

Permutation Entropy and Bubble Entropy: Possible interactions and synergies between order and sorting relations

David Cuesta–Frau^{1,*}, and Borja Vargas²

¹ Technological Institute of Informatics(ITI), Universitat Politècnica de València, Campus Alcoi, Plaza Ferrándiz y Carbonell, 2, 03801, Alcoi, Spain

² Department of Internal Medicine, Móstoles Teaching Hospital, Móstoles, 28935, Madrid, Spain

* **Correspondence:** Email: dcuesta@disca.upv.es; Tel: 651028565.

Abstract: Despite its widely demonstrated usefulness, there is still room for improvement in the basic Permutation Entropy (PE) algorithm, as several subsequent studies have proposed in the recent years. For example, some improved PE variants try to address possible PE weaknesses, such as its only focus on ordinal information, and not on amplitude, or the possible detrimental impact of equal values in subsequences due to motif ambiguity. Other evolved PE methods try to reduce the influence of input parameters. A good representative of this last point is the Bubble Entropy (BE) method. BE is based on sorting relations instead of ordinal patterns, and its promising capabilities have not been extensively assessed yet. The objective of the present study was to comparatively assess the classification performance of this new method, and study and exploit the possible synergies between PE and BE. The claimed superior performance of BE over PE was first evaluated by conducting a series of time series classification tests over a varied and diverse experimental set. The results of this assessment apparently suggested that there is a complementary relationship between PE and BE, instead of a superior/inferior relationship. A second set of experiments using PE and BE simultaneously as the input features of a clustering algorithm, demonstrated that with a proper algorithm configuration, classification accuracy and robustness can benefit from both measures.

Keywords: Permutation Entropy; Bubble Entropy; signal classification; clustering

1. Introduction

Regardless of its relatively young age in comparison with other entropy statistics, Permutation Entropy (PE) has soon become one of the most utilised time series entropy–related measures. It was proposed in the well known paper by Bandt and Pompe [1] in 2002, and since then, it has given rise to a number of applications and further algorithm developments. This number is growing exponentially

[2], as a clear evidence of the utility of the PE approach.

Regarding PE applications, it has been used mainly in physiological records classification. There are many scientific papers that illustrate this point. For example, we have used PE to successfully classify body temperature records from febrile and healthy patients [3] and to classify glucose records from patients at diabetes risk [4]. More frequent applications are based on electroencephalogram [5, 6, 7, 8] and heart rate variability [9, 10, 11] analysis .

Other PE fields of application are also receiving a great deal of attention. In econometrics, there are quite many interesting examples already available. In [12], the authors used PE to try to unearth the dynamical properties of time series featuring Dow Jones Industrial Average data from 1901 to 2016, obtaining a very high degree of randomness related to the market efficiency. They also used PE on temporal windows to identify market events. Along this line, [13] applied PE to stock market time series of several countries to estimate their investment attractiveness. Their findings confirmed the validity of this approach, and they also found a clear correlation between market crisis and PE value decline. The paper [14] studied the evolution of stock market efficiency during the last financial crisis. It used some specific stock exchange indices, with its main focus on the differences between pre and post crisis data processed by PE. In the same context, [15] used the complexity-entropy causality plane and a permutation statistical complexity to analyse financial time series. Using permutation Shannon entropy and permutation Fisher information measure, [16] analysed the possible deterministic components in Libor rates time series induced by manipulation. The study in [17] applied a dynamic approach to detect structural changes in time series using different entropy measures related to PE, including Gaussian, Rényi, Tsallis, and Shannon entropies. They first developed a theoretical study over synthetic time series with abrupt changes in order to assess the transition detection power. Then they applied the methods to real time series, seismic data, and economic data (exchange rates between US dollar and gold, and Nasdaq time series).

In mechanical engineering, PE has also clearly demonstrated its usefulness, mainly in the framework of fault diagnosis. This is the case in [18]. This study uses vibration signals for bearing fault diagnosis based on multiscale PE and multinomial logit model. The classification accuracy achieved was close to 100%. The researchers in [19] also used multiscale PE for automatic recognition of weak faults in hydraulic systems from vibration signals as well. All the diagnosis methods tested achieved at least a recognition rate of 89%, and were able to discern among normal state, slight, moderate, and severe leakage.

The basic PE algorithm has also been improved since its initial version. Two possible algorithm weaknesses were almost immediately detected by researchers regarding PE. First, PE is based on relative frequency of ordinal patterns, but it does not take into account the possible influence of amplitude differences. For example, the subsequences $(0.25, -1.7, -0.33)$ and $(100, 99, 99.5)$ correspond to the same ordinal pattern $(1, 2, 0)$, but from an amplitude perspective they are very different. Some PE algorithm improvements have been proposed in the last years to have the amplitude patterns somehow covered in PE. For example, the Weighted Permutation Entropy method [20, 21, 22] applies a correction factor before computing the relative frequencies, based on the subsequence variance. The Fine Grained Permutation method [23] adds a new symbol to the ordinal motif of the subsequence that accounts for the amplitude differences. The amplitude information overlooked by the standard PE method has been demonstrated to be significant in many classification tests [24].

The other claimed PE flaw is the impossibility to assign a single motif to subsequences that contain

equal values or ties. For example, the subsequence (2, 1, 2) could be assigned the ordinal pattern or motif (1, 0, 2), but also the (1, 2, 0). In the seminal paper [1], Bandt and Pompe already acknowledged this possible issue, and proposed to add a small random noise to break ties in the unlikely event that equal values fall within the same subsequence. However, such ties are not that unlikely, and some improvements have also been suggested since then to minimise the possible histogram bias due to ordinal ambiguities. Thus, the modified PE method [11] generates more motifs corresponding to the possible ties. For example, the motif (0, 1, 2, 3) generates also ordinal patterns (0, 1, 22), (0, 11, 3), (0, 11, 2), and (0, 111). The method in [25] uses a Bayesian missing data imputation to learn from the unambiguous subsequences what are the most likely motifs in case of ties. The presence of such ties can lead to the incorrect interpretation of the chaotic nature of the time series, but they exert a minor impact on classification performance [26].

Another line of research related to PE is the use of complex networks to map time series [27]. Specifically, the use of transition networks with nodes featuring ordinal patterns as for PE, connected by edges based on temporal succession information from a time series, is a very promising research topic for the future in the field of non-linear dynamics analysis [28]. We also have results in this regard already, using Hidden Markov Models to synthetically generate time series based on the transition probabilities between consecutive ordinal patterns [29].

Other methods have been proposed to improve the performance and robustness of PE. Special attention deserves the Amplitude Aware Permutation Entropy method [30], a method that addresses simultaneously the two PE problems described above. This is also the case for the recently published Improved Permutation Entropy method [31], which adds an amplitude quantization stage to account for amplitude differences, ties, and noise. A different approach is the Bubble Entropy (BE) method [32], devised to reduce the dependence of PE on input parameters, such as the data length or the embedding dimension, by counting the number of sample swaps necessary to achieve the ordered subsequences instead of counting order patterns. It seems that BE exhibits more stability and discriminating power than the standard PE method [32].

Since the BE approach could be a game-changer, based on sorting relations instead of order relations, we hypothesized that PE and BE could exhibit some kind of synergy, since they are not looking exactly at the same aspects of the time series dynamics. For example, the patterns {0, 2, 1} and {1, 0, 2} are different for PE, but obtained with the same number of swaps (1) from the original ordinal pattern {0, 1, 2}, therefore, they are considered the same from the BE perspective.

2. Materials and methods

2.1. Permutation Entropy

The present study is based on the original PE algorithm described in [1], for a single time scale. This method computes a normalised histogram of ordinal patterns found in the subsequences drawn from a time series, when sorted in ascending order, from which the Shannon Entropy is calculated. The length of these subsequences is defined by an input parameter, the embedded dimension m .

Formally, the input time series under analysis is defined as a vector of N components $\mathbf{x} = \{x_0, x_1, \dots, x_{N-1}\}$. A generic subsequence extracted commencing at sample x_j of \mathbf{x} is defined as a vector of m components $\mathbf{x}_j^m = \{x_j, x_{j+1}, \dots, x_{j+m-1}\}$. In its original state, the samples in \mathbf{x}_j^m can be assigned a default growing set of indices given by $\boldsymbol{\pi}^m = \{0, 1, \dots, m-1\}$. The subsequence \mathbf{x}_j^m undergoes then an

ascending sorting process, and the sample order changes in it, are mirrored in the vector of indices π^m . The resulting new version of this vector, $\pi_j^m = \{\pi_0, \pi_1, \dots, \pi_{m-1}\}$, with $x_{j+\pi_0} \leq x_{j+\pi_1} \leq x_{j+\pi_2} \dots \leq x_{j+\pi_{m-1}}$, is compared, in principle, with all the possible $m!$ ordinal patterns of length m . When a coincidence is found, a specific associated counter to that pattern, $c_i \in \mathbf{c}$, is increased. This process is repeated with all the possible $N - (m - 1)$ subsequences ($0 \leq j < N - m + 1$) until the complete histogram is obtained. Each bin of the histogram is finally normalised by $N - (m - 1)$ in order to obtain an estimation of the probability of each ordinal pattern: $\mathbf{p} = \{p_0, p_1, \dots, p_{m!-1}\} \left| p_i = \frac{c_i}{N-(m-1)} \right.$. This vector of probabilities is used to calculate PE as (assuming $\log 0 = 0$):

$$\text{PE}(\mathbf{x}, m, N) = - \sum_{k=0}^{m!-1} p_k \log p_k \quad (1)$$

For example, let $\mathbf{x} = \{-0.45, 1.9, 0.87, -0.91, 2.3, 1.1, 0.75, 1.3, -1.6, 0.47, -0.15, 0.65, 0.55, -1.1, 0.3\}$ be a time series of length 15 whose PE has to be calculated using $m = 3$. The procedure of subsequence extraction and sorting is illustrated in Table 1. Many other numerical examples of PE computation can be found in the literature. For example, in works such as [33, 34].

Table 1. Motifs found for PE computation for the example input series $\mathbf{x} = \{-0.45, 1.9, 0.87, -0.91, 2.3, 1.1, 0.75, 1.3, -1.6, 0.47, -0.15, 0.65, 0.55, -1.1, 0.3\}$.

Subsequence	Swap 1	Swap 2	Swap 3	Motif
$\mathbf{x}_0^3 = \{-0.45, 1.9, 0.87\}$	$\{-0.45, 0.87, 1.9\}$			$\{0, 2, 1\}$
$\mathbf{x}_1^3 = \{1.9, 0.87, -0.91\}$	$\{0.87, 1.9, -0.91\}$	$\{0.87, -0.91, 1.9\}$	$\{-0.91, 0.87, 1.9\}$	$\{2, 1, 0\}$
$\mathbf{x}_2^3 = \{0.87, -0.91, 2.3\}$	$\{-0.91, 0.87, 2.3\}$			$\{1, 0, 2\}$
$\mathbf{x}_3^3 = \{-0.91, 2.3, 1.1\}$	$\{-0.91, 1.1, 2.3\}$			$\{0, 2, 1\}$
$\mathbf{x}_4^3 = \{2.3, 1.1, 0.75\}$	$\{1.1, 2.3, 0.75\}$	$\{1.1, 0.75, 2.3\}$	$\{0.75, 1.1, 2.3\}$	$\{2, 1, 0\}$
$\mathbf{x}_5^3 = \{1.1, 0.75, 1.3\}$	$\{0.75, 1.1, 1.3\}$			$\{1, 0, 2\}$
$\mathbf{x}_6^3 = \{0.75, 1.3, -1.6\}$	$\{0.75, -1.6, 1.3\}$	$\{-1.6, 0.75, 1.3\}$		$\{2, 0, 1\}$
$\mathbf{x}_7^3 = \{1.3, -1.6, 0.47\}$	$\{-1.6, 1.3, 0.47\}$	$\{-1.6, 0.47, 1.3\}$		$\{1, 2, 0\}$
$\mathbf{x}_8^3 = \{-1.6, 0.47, -0.15\}$	$\{-1.6, -0.15, 0.47\}$			$\{0, 2, 1\}$
$\mathbf{x}_9^3 = \{0.47, -0.15, 0.65\}$	$\{-0.15, 0.47, 0.65\}$			$\{1, 0, 2\}$
$\mathbf{x}_{10}^3 = \{-0.15, 0.65, 0.55\}$	$\{-0.15, 0.55, 0.65\}$			$\{0, 2, 1\}$
$\mathbf{x}_{11}^3 = \{0.65, 0.55, -1.1\}$	$\{0.55, 0.65, -1.1\}$	$\{0.55, -1.1, 0.65\}$	$\{-1.1, 0.55, 0.65\}$	$\{2, 1, 0\}$
$\mathbf{x}_{12}^3 = \{0.55, -1.1, 0.3\}$	$\{-1.1, 0.55, 0.3\}$	$\{-1.1, 0.3, 0.55\}$		$\{1, 2, 0\}$

2.2. Bubble Entropy

BE is a very recently proposed entropy measure [32] that has not received the attention it deserves yet, but it will surely become an indispensable tool in the field of non-linear dynamics analysis due to the possible improvements over PE it introduces.

The main objective of BE was to minimise the dependence of entropy measures on input parameters. In general, many of the most utilised measures require at least two for their computation, an

embedded dimension m , and a similarity threshold r . Depending on the values of these parameters, the performance can vary significantly in terms of discriminating power, robustness to artifacts, or any other desirable feature. Along this line, the Rank–Entropy method (RankEn) described in [35] tried to reduce the dependence on r by calculating the amount of shuffling that the distances between two subsequences under comparison had to undergo to sort such distances in ascending order. In this case, the parameter r determines the maximum rank of the set of distances that contribute to the entropy measure, and therefore, the r value is less critical.

In [32], researchers went one step further in order to remove completely the r parameter from the entropy computation. Taking advantage of the PE method, already independent of r , and the shuffling used for RankEn, the authors proposed first a new method, called Conditional Rényi Permutation Entropy, which combined Conditional Entropy (CE) [36], and Rényi Permutation Entropy (RPE) [37]. CE can be computed as $CE(m) = PE(m + 1) - PE(m)$. RPE achieved the best results when using a quadratic approach, as for Sample Entropy, and outperformed other entropy methods in terms of discriminating power [38]. Then, BE was finally defined by using two consecutive m values as CE, with a quadratic Rényi definition of entropy, and processing ordinal patterns, as PE. This way, BE does not need parameter r , is less dependent on m , or even independent for large m values, it gives more emphasis to peaks without neglecting lower values, the number of unique possible states is reduced (ties and amplitude influence are less critical since more matches to compute relative frequencies can be found), and converges faster than PE, with a higher discriminating power [32].

The core of the BE algorithm is that of PE, but instead of computing the relative frequency of the ordinal patterns, it computes the relative frequency of the necessary swaps to get an ordered subsequence. First, each subsequence \mathbf{x}_j^m is also sorted in ascending order using a bubble sort algorithm. The counter vector \mathbf{c} stores instead all the swaps necessary in each case, with a maximum given by $[0, \frac{m(m-1)}{2}]$. Each bin is normalised by $N - m + 1$ too. From all the resulting relative frequencies p_i , accounting for how likely a number of swaps is [32], the Rényi entropy of order 2 is computed as:

$$H_2^m(\mathbf{x}) = -\log \sum_{k=0}^{\frac{m(m-1)}{2}} p_k^2 \quad (2)$$

The embedding dimension is then increased by 1, $m \rightarrow m + 1$, and the procedure is repeated again to obtain a new entropy value from Eq. 2, H_2^{m+1} . Finally, BE is obtained in a similar way as for Approximate Entropy, ApEn [39]:

$$BE(\mathbf{x}, m, N) = \frac{(H_2^{m+1} - H_2^m)}{\log \frac{m+1}{m-1}} \quad (3)$$

The procedure of subsequence extraction, sorting, and swap computation, for the same input time series as for PE, is illustrated in Table 2. The maximum number of swaps in this case is 3.

From the results in Table 2, the counter vector components for the first iteration of the method are obtained as: $c_0 = 0$, $c_1 = 7$, $c_2 = 3$, and $c_3 = 3$. These values, once normalised, will become the probability vector p from which entropy H^m can be computed. The process is then repeated for $m \rightarrow m + 1 = 4$ to obtain H^{m+1} , with results: $c_0 = 0$, $c_1 = 1$, $c_2 = 2$, $c_3 = 2$, $c_4 = 6$, $c_5 = 1$, and $c_6 = 0$. Finally, applying Eq. 3, the value for BE obtained is 0.3115, with $H^4 = 1.1411$, and $H^3 = 0.9252$. An implementation of this method is detailed in Algorithm 1.

Algorithm 1 Bubble Entropy (BE) Algorithm

Input: \mathbf{x} , $m > 2$, $N > m + 1$

Initialisation: $\text{BE} \leftarrow 0$, $\mathbf{c} \leftarrow \{0\}$, $\mathbf{p} \leftarrow \{0\}$

for $M \leftarrow m, \dots, m + 1$ **do**

for $j \leftarrow 0, \dots, N - M$ **do**

$\mathbf{x}_j^M \leftarrow \{x_j, x_{j+1}, \dots, x_{j+M-1}\}$

$\text{bSorted} \leftarrow \text{false}$

$\text{nSwaps} \leftarrow 0$

while ($\text{bSorted}=\text{false}$) **do**

 ▶ Bubble sort

$\text{bSorted} \leftarrow \text{true}$

for $i \leftarrow j, \dots, j + M - 2$ **do**

if ($x_i > x_{i+1}$) **then**

$\text{swap}(x_i, x_{i+1})$

$\text{nSwaps}=\text{nSwaps}+1$

$\text{bSorted} \leftarrow \text{false}$

end if

end for

end while

$c_{\text{nSwaps}} \leftarrow c_{\text{nSwaps}} + 1$

 ▶ Update counter

end for

$H_2^M \leftarrow 0$

for $k \leftarrow 0, \dots, \frac{m(m-1)}{2}$ **do**

$p_k \leftarrow \frac{c_k}{N-M+1}$

$H_2^M \leftarrow H_2^M - (p_k^2)$

end for

$H_2^M \leftarrow -\log H_2^M$

end for

$\text{BE} = H_2^{m+1} - H_2^m$

$\text{BE} = \text{BE} / \log \frac{m+1}{m-1}$

Output: $\text{BE}(\mathbf{x}, m, N)$

Table 2. Necessary swaps to sort in ascending order all the subsequences that can be extracted from the example data to compute BE.

Subsequence	Swap 1	Swap 2	Swap 3	Swaps
$\mathbf{x}_0^3 = \{-0.45, 1.9, 0.87\}$	$\{-0.45, 0.87, 1.9\}$			1
$\mathbf{x}_1^3 = \{1.9, 0.87, -0.91\}$	$\{0.87, 1.9, -0.91\}$	$\{0.87, -0.91, 1.9\}$	$\{-0.91, 0.87, 1.9\}$	3
$\mathbf{x}_2^3 = \{0.87, -0.91, 2.3\}$	$\{-0.91, 0.87, 2.3\}$			1
$\mathbf{x}_3^3 = \{-0.91, 2.3, 1.1\}$	$\{-0.91, 1.1, 2.3\}$			1
$\mathbf{x}_4^3 = \{2.3, 1.1, 0.75\}$	$\{1.1, 2.3, 0.75\}$	$\{1.1, 0.75, 2.3\}$	$\{0.75, 1.1, 2.3\}$	3
$\mathbf{x}_5^3 = \{1.1, 0.75, 1.3\}$	$\{0.75, 1.1, 1.3\}$			1
$\mathbf{x}_6^3 = \{0.75, 1.3, -1.6\}$	$\{0.75, -1.6, 1.3\}$	$\{-1.6, 0.75, 1.3\}$		2
$\mathbf{x}_7^3 = \{1.3, -1.6, 0.47\}$	$\{-1.6, 1.3, 0.47\}$	$\{-1.6, 0.47, 1.3\}$		2
$\mathbf{x}_8^3 = \{-1.6, 0.47, -0.15\}$	$\{-1.6, -0.15, 0.47\}$			1
$\mathbf{x}_9^3 = \{0.47, -0.15, 0.65\}$	$\{-0.15, 0.47, 0.65\}$			1
$\mathbf{x}_{10}^3 = \{-0.15, 0.65, 0.55\}$	$\{-0.15, 0.55, 0.65\}$			1
$\mathbf{x}_{11}^3 = \{0.65, 0.55, -1.1\}$	$\{0.55, 0.65, -1.1\}$	$\{0.55, -1.1, 0.65\}$	$\{-1.1, 0.55, 0.65\}$	3
$\mathbf{x}_{12}^3 = \{0.55, -1.1, 0.3\}$	$\{-1.1, 0.55, 0.3\}$	$\{-1.1, 0.3, 0.55\}$		2

2.3. Performance analysis

The performance analysis of the methods under analysis will be assessed both in quantitative and qualitative terms. The classification accuracy will be quantified as the ratio of correctly assigned time series to classes. For example, a 0.80 classification accuracy will refer to the fact that 80% of the time series were labelled as their true class after the classification analysis has been completed. The remaining 20% has been incorrectly included in another class distinct to the true class to which they belong. The classification accuracy was quantified using the classical sensitivity, specificity and accuracy performance indicators. They were obtained using a threshold from the closest point in the ROC curve [40] to the (0,1) point [41]. We first proceeded by computing the number of True Positives (TP), False Negatives (FN), True Negatives (TN), and False Positives (FP). Then, the classical performance metrics Sensitivity = $TP / (TP+FN)$, Specificity = $TN / (TN+FP)$, and Accuracy = $(TN + TP) / (TN + TP + FN + FP)$ [42], were obtained.

However, the numerical result of the classification accuracy does not suffice to provide a complete picture of the discriminating power of the method. For example, a relatively high accuracy of 0.75 can be achieved for the classification of two size-balanced classes, but with a sensitivity close to 0.5, pure guess, and a specificity close to 1, a useless performance since there must be a trade-off between both parameters to avoid statistical uncertainty.

Therefore, in order to assess the statistical significance of the results quantified in terms of classification accuracy, we used a bootstrap version [43] of the equality of means (there is no difference between the classes under analysis according to the values of PE or BE obtained for classification) hypothesis testing. By using this bootstrap method, it is not necessary to assume any specific distribution of the data, and it can be applied even to small sizes of data.

Given two samples $\mathbf{Y} = \{Y_1, \dots, Y_{n_Y}\}$ and $\mathbf{Z} = \{Z_1, \dots, Z_{n_Z}\}$, the steps to carry out this test are:

1. Calculate the T statistic as in the standard t–Student test [44].
2. Resample the input data randomly with replacement to create bootstrapped versions of the data, \mathbf{Y}^* and \mathbf{Z}^* .
3. Calculate a bootstrap statistic T^* as:

$$T^* = \frac{(\hat{\mu}_{\mathbf{Y}^*} - \hat{\mu}_{\mathbf{Z}^*}) - (\hat{\mu}_{\mathbf{Y}} - \hat{\mu}_{\mathbf{Z}})}{\sqrt{\frac{\hat{\sigma}_{\mathbf{Y}^*}^2}{n_{\mathbf{Y}^*}-1} + \frac{\hat{\sigma}_{\mathbf{Z}^*}^2}{n_{\mathbf{Z}^*}-1}}},$$

where μ is the mean, and σ^2 the variance of the data.

4. Repeat steps 2 and 3 several times ($k = 200$ in this case) to obtain the corresponding statistics $T_1^*, T_2^*, \dots, T_k^*$.
5. Sort the previous statistics in increasing order, $T_{(1)}^*, T_{(2)}^*, \dots, T_{(k)}^*$.
6. Reject the equal means hypothesis, $H : \mu_{\mathbf{Y}} = \mu_{\mathbf{Z}}$, if $T < T_{(q_1)}^*$ or $T > T_{(q_2)}^*$, where $q_1 = \lfloor k\alpha/2 \rfloor$ and $q_2 = k - q_1 + 1$.

The null hypothesis of equal means of the classes in terms of PE or BE was assessed using this test. The acceptance threshold was set at $\alpha = 0.05$.

The possible synergy between PE and BE can be exploited in many ways. They can become independent variables of the same classification function, as in logistic regression applications [45, 3], among many other similar methods. The input dataset can be split into training and test sets to obtain more complex polynomial classification functions, as is the case when using neural networks [46]. They can also conform the input feature vector for an unsupervised classification approach featured by a clustering algorithm [47].

We have chosen the clustering approach for its simplicity and good results in previous similar studies [48]. The specific method employed is the Max–Min algorithm [49]. The input number of clusters was set to 2, and therefore 2 centroids are selected, one from each experimental dataset. In order to further simplify the clustering method, since the goal was not to design a classification method but rather explore the possible synergy between PE and BE, the centroids are chosen from those time series with the maximum and minimum PE value, the furthest points in the PE feature space. Finally, each time series is assigned to the class of the closest centroid in a single iteration, quantified using a Euclidean distance. If the goal was to maximise the classification performance, a more evolved clustering algorithm could be chosen, such as a genetic method that converges to a global minimum cost [50], or a density–based method, that enables the use of highly complex partition regions [51]. Other distance metrics, or more iterations with centroid updating could also be used, such as in the k –Means method or its variants [52, 53].

3. Experiments

3.1. Experimental dataset

The experimental dataset was composed of 7 biomedical records previously used in other studies. In addition, 4 out of the 7 datasets are publicly available at Physionet [54]. The information regarding the specific datasets employed is summarised in Table 3, for our proprietary data, and in Table 4, for the publicly available data at Physionet. Each row includes the dataset assigned name, a short description,

and references to further studies or information about the specific dataset in each case. An example plot of one record is also included.

Table 3. Summary of the own experimental datasets employed in the present study.

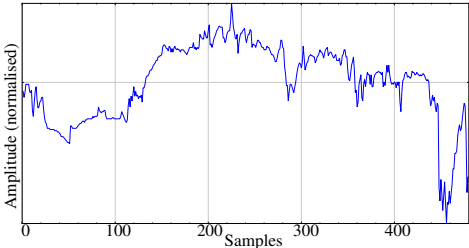
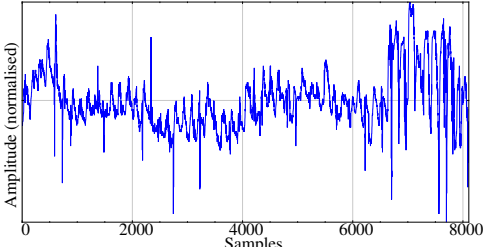
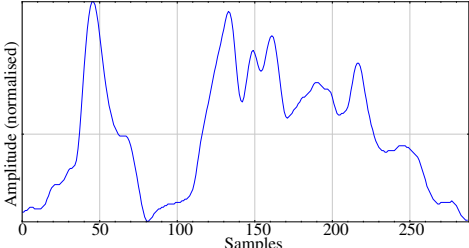
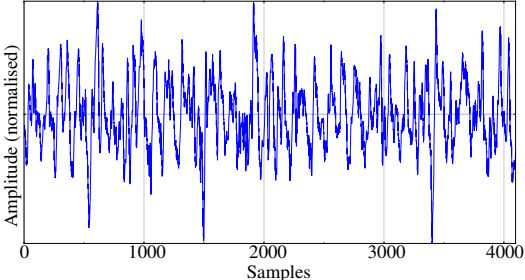
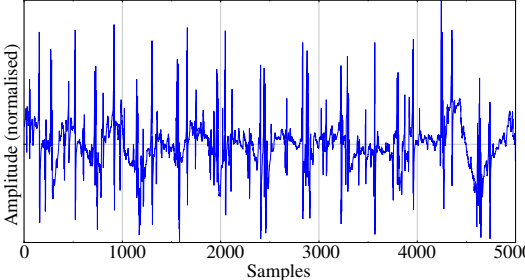
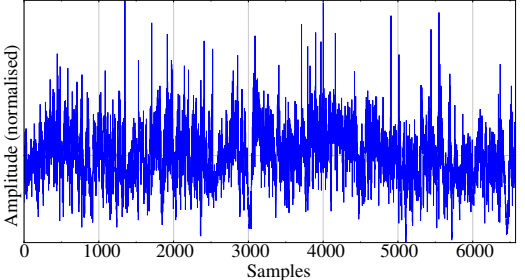
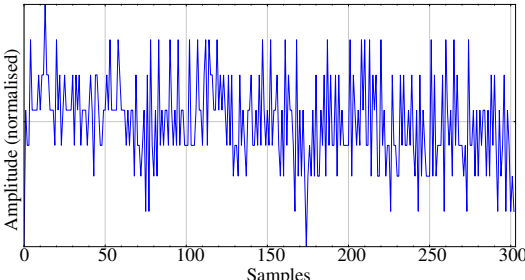
Name	Description	Plot
TEMPERATURE	8h temperature records of 16 healthy patients and 14 patients that developed a fever 24h before the recording. Sampling rate of 1 sample per minute [55, 3].	 <p data-bbox="954 927 1409 987">Example signals of the TEMPERATURE experimental database.</p>
SURVIVAL	Several days of temperature records of critically ill patients. A total of 36 records, 18 for survivors, and 18 for non-survivors. Sampling rate of 10 minutes [56].	 <p data-bbox="986 1375 1377 1435">Example signals of the SURVIVAL experimental database.</p>
GLUCOSE	24h records of 206 patients at risk of developing a diabetes. After 3 years, 18 of these patients became diabetic. Sampling rate of 5 minutes [4, 57].	 <p data-bbox="986 1823 1377 1883">Example signals of the GLUCOSE experimental database.</p>

Table 4. Summary of the publicly available experimental datasets employed in the present study.

Name	Description	Plot
BONN	4097 samples long electroencephalograph records. Two classes, with and without epileptic seizures, 100 records each. Sampling rate of 173.61 Hz [58, 59].	 <p data-bbox="887 674 1414 741">Example signals of the BONN experimental database.</p>
EMG	Three classes, healthy (10 records), myopathy (22 records), and neuropathy (29 records) electromyogram records of 5000 samples. Sampling rate of 4KHz [54, 59].	 <p data-bbox="887 1070 1414 1137">Example signals of the EMG experimental database.</p>
FANT	Two classes of RR-interval records from 20 young and 20 elderly healthy subjects monitored during 120 minutes. The length of the records varies, but they are longer than 5000 samples in all cases [54, 24, 60].	 <p data-bbox="887 1467 1414 1534">Example signals of the FANT experimental database.</p>
PAF	Two classes of RR-interval records, 25 paroxysmal atrial fibrillation episode free, and 25 with. The duration of the records is 5 minutes, with lengths mostly around 400 samples [54, 59].	 <p data-bbox="887 1859 1414 1926">Example signals of the PAF experimental database.</p>

3.2. Results

The experiments were devised to assess the performance of BE and PE using a different and more varied dataset than in the seminal paper of BE [32], and also to explore the possible synergies between both measures. These experiments also included a variation of the parameter m , from 3 to 8, since input parameter influence is another topic of intense debate and research in the scientific literature. All the experiments and algorithms were implemented in C++ programming language, using MinGW 4.9.2 32 bit compiler (www.mingw.org).

In this regard, Table 5 shows the classification performance (if statistically significant) of the results achieved by PE and BE using the TEMPERATURE dataset. In this case, PE exhibits a high classification accuracy, around 0.8 in some m cases. BE performance is not significant for any m value.

Table 5. Classification results achieved using PE and BE, applied to the TEMPERATURE database.

m	PE			BE		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
3	–	–	–	–	–	–
4	–	–	–	–	–	–
5	–	–	–	–	–	–
6	0.81	0.57	0.70	–	–	–
7	0.93	0.64	0.80	–	–	–
8	0.81	0.85	0.83	–	–	–

For the SURVIVAL dataset, the classification results achieved are shown in Table 6. The results for BE are not statistically significant for $m = 8$, being PE more stable with m . The maximum performance achieved by both methods was 0.77.

Table 6. Classification results achieved using PE and BE, applied to the SURVIVAL database.

m	PE			BE		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
3	0.72	0.72	0.72	0.72	0.77	0.75
4	0.72	0.77	0.75	0.72	0.77	0.75
5	0.72	0.77	0.75	0.77	0.77	0.77
6	0.72	0.77	0.75	0.72	0.77	0.75
7	0.72	0.83	0.77	0.61	0.61	0.61
8	0.72	0.83	0.77	–	–	–

For the last dataset in Table 3, GLUCOSE, the results are listed in Table 7. Both statistics yielded significant classification results in all m cases tested, but the performance of BE was slightly superior to that of PE, 0.80 vs. 0.79. This small difference is more significant if sensitivity is taken into account,

since with PE it was 0.61, but with BE it was 0.66, a clear difference for a small class size of 18 records.

Table 7. Classification results achieved using PE and BE, applied to the GLUCOSE database.

m	PE			BE		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
3	0.77	0.56	0.58	0.66	0.71	0.70
4	0.72	0.67	0.67	0.72	0.69	0.69
5	0.61	0.80	0.79	0.72	0.77	0.77
6	0.77	0.63	0.65	0.77	0.66	0.67
7	0.72	0.73	0.73	0.66	0.74	0.73
8	0.72	0.72	0.72	0.66	0.81	0.80

Table 8 shows the classification performance achieved by PE and BE using the BONN dataset. In this case, PE exhibits a very high classification accuracy, close to 0.9 in all m cases. BE performance is lower, 0.785 at most, and even with no significant results for $m = 5, 6, 7$.

Table 8. Classification results achieved using PE and BE, applied to the BONN database.

m	PE			BE		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
3	0.93	0.90	0.91	0.74	0.83	0.78
4	0.93	0.89	0.91	0.70	0.67	0.68
5	0.92	0.89	0.90	–	–	–
6	0.91	0.89	0.90	–	–	–
7	0.93	0.85	0.89	–	–	–
8	0.90	0.85	0.87	0.75	0.52	0.63

In Table 9, the classification results for the three classes in the EMG database are shown (healthy, myopathy, neuropathy). PE was not able to achieve significant results for the second case (Healthy vs. Myopathy), whereas BE found differences in all cases for $m = 3$ and 5. Therefore, BE can be considered to outperform PE in this experiment.

Table 10 shows the results for the FANT database. In this case, BE was clearly better than PE, with significant results for $m = 7, 8$, and a maximum classification accuracy of 0.875. PE was not capable of achieving the significance threshold, probably because in many experiments either Sensitivity or Specificity were close to 0.5, despite a good accuracy in the other measure.

Finally, the results of the last dataset are shown in Table 11. The classification accuracy of PE was quite stable around 0.8, but BE clearly failed in this case.

A summary of the best statistically significant accuracy results achieved using the two measures assessed (PE, BE) is shown in Table 12. Of the 7 datasets, PE yielded a better accuracy in 3 cases, and BE in other 3. In 2 datasets each, PE and BE were not capable of finding significant differences. There was no dataset where none of the measures studied failed to find significant differences, but in 4 of

Table 9. Classification results achieved using PE and BE, applied to the EMG database. There are three rows for each case, first row for results comparing healthy and myopathy records, second row for healthy and neuropathy, and the third row for myopathy and neuropathy comparison.

m	PE			BE		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
3	1	1	1	1	1	1
	–	–	–	0.9	0.58	0.66
	0.55	0.77	0.67	0.96	1	0.98
4	1	1	1	1	1	1
	–	–	–	–	–	–
	1	1	1	1	1	1
5	1	1	1	0.86	0.90	0.88
	–	–	–	0.75	0.80	0.78
	1	1	1	0.51	0.77	0.66
6	1	1	1	0.70	0.77	0.75
	–	–	–	–	–	–
	1	1	1	0.75	0.77	0.76
7	1	1	1	1	1	1
	–	–	–	–	–	–
	1	1	1	1	1	1
8	1	1	1	1	1	1
	–	–	–	–	–	–
	1	1	1	1	1	1

the experiments, one of the measures achieved a good performance, whereas the other was unable to achieve statistically significant results. It is not possible to say what measure was better in this case. However, what became more apparent was the fact that there exists a clear complementarity between PE and BE, each one seems to look at different properties of the time series dynamics, and this should be exploited.

The second group of experiments were focused on exploring the possible synergy between PE and BE, used jointly for classification. Without any specific customisation, the clustering algorithm described in Section 2.3 was applied to some datasets in a completely unsupervised way. The quantitative results achieved are shown in Tables 13, 14 and 15 to provide an example of beneficial synergy or no synergy.

The results for the GLUCOSE dataset in Table 13 probably best illustrate the optimal synergy between BE and PE. The classification performance achieved combining these two measures is far higher than that achieved individually, which confirms the hypothesis that there exists at least some

Table 10. Classification results achieved using PE and BE, applied to the FANT database.

m	PE			BE		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
3	–	–	–	–	–	–
4	–	–	–	–	–	–
5	–	–	–	–	–	–
6	–	–	–	–	–	–
7	–	–	–	0.75	1	0.875
8	–	–	–	0.80	0.85	0.82

Table 11. Classification results achieved using PE and BE, applied to the PAF database.

m	PE			BE		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
3	0.76	0.88	0.82	–	–	–
4	0.80	0.84	0.82	–	–	–
5	0.80	0.80	0.80	–	–	–
6	0.92	0.72	0.82	–	–	–
7	0.96	0.68	0.82	–	–	–
8	0.92	0.68	0.80	–	–	–

PE–BE complementarity. Moreover, it is independent of the m value, an almost ideal case.

However, one size does not fit all, and the method should be customised for each dataset. For example, what worked in Table 13, for the GLUCOSE dataset, did not work in other cases. Table 14 shows the classification accuracy achieved using the TEMPERATURE dataset, where the individual performance using PE is higher than using both measures. Since BE did not achieve any significant result, it seems it has a blurring effect on the differences between classes. The discerning method should be customised to automatically detect the measure that better features the differences, and therefore increase not only the accuracy, but also the robustness of the classification by reducing the number of datasets where no significance is achieved.

The results in Table 15 also correspond to a case where the possible synergy is not seamlessly exploited. The clustering method achieves a lower performance than the methods individually, but with a small difference for $m < 6$. As in the previous case, with clustering method optimisation and customisation, or using other more sophisticated pattern recognition methods, it can be hypothesised that there is room for performance improvement.

4. Discussion

The experiments in Section 3 were first aimed at confirming the performance of the BE approach from a broader point of view than in [32], using a set of records with markedly different properties

Table 12. Summary of the performance achieved by PE and BE. Only the best significant case is reported in terms of overall accuracy.

Dataset	PE	BE
TEMPERATURE	0.80	–
SURVIVAL	0.77	0.77
GLUCOSE	0.79	0.80
BONN	0.91	0.78
EMG	–	1,0.66,0.98
FANT	–	0.87
PAF	0.82	–

Table 13. Clustering results using the GLUCOSE database and features PE and BE.

m	Clustering	PE	BE
3	0.91	0.58	0.70
4	0.91	0.67	0.69
5	0.91	0.79	0.77
6	0.91	0.65	0.67
7	0.91	0.73	0.73
8	0.91	0.72	0.80

and behaviour: slow varying signals such as TEMPERATURE, SURVIVAL, and GLUCOSE datasets, faster varying signals such as the BONN dataset, spiky records in the EMG dataset, and a combination of probably all the previous, in the RR datasets.

According to the results, the performance of BE was, in general, similar but complementary to that of PE, in terms of classification accuracy. Namely, both statistics achieved similar performance, but on different datasets. Taking into account all the m values tested, PE seems to be more robust and stable with slow varying signals, and BE for spiky records (as theoretically stated in Section 2.2), with the rest of records in between. The dependence of both PE and BE on parameter m remains an unsolved problem, since there was a great accuracy variability depending on the specific m value chosen in many cases.

PE yielded a higher accuracy than BE for datasets TEMPERATURE (Table 5), BONN (Table 8), and PAF (Table 11), whereas BE outperformed PE when applied to GLUCOSE (Table 7), EMG (Table 9), and FANT (Table 10) time series. The performance was the same for the SURVIVAL dataset (Table 6). Obviously, a generalisation can not be made however representative the experimental dataset is, but it can arguably be stated that the combination of PE and BE is more likely to provide significant results than each measure isolatedly, mainly based on the fact that non-significance on one measure is often accompanied by high performance on the other (Table 12).

Along this line, we explored a potential synergy using a very simple approach based on a clustering algorithm, and PE and BE as the two extracted features. If the object distribution in the PE–BE

Table 14. Clustering results using the TEMPERATURE database and features PE and BE.

m	Clustering	PE	BE
3	0.53	–	–
4	0.60	–	–
5	0.53	–	–
6	0.53	0.70	–
7	0.56	0.80	–
8	0.70	0.83	–

Table 15. Clustering results using the SURVIVAL database and features PE and BE.

m	Clustering	PE	BE
3	0.69	0.72	0.75
4	0.75	0.75	0.75
5	0.72	0.75	0.77
6	0.58	0.75	0.75
7	0.58	0.77	0.61
8	0.50	0.77	–

plane matches the centroid computation and the clustering algorithm geometric properties, the possible synergy can be exploited successfully, as demonstrated in Table 13. However, this is not always the case, and the combination of PE and BE can also exert a blurring effect and have a detrimental impact on classification performance (Table 14 and Table 15). Therefore, whatever the discriminating function chosen is (linear, polynomial, logistic, etc.), or the classification algorithm (clustering, neural networks, random forest, etc.), it should be customised in each case to make the most of the information provided by PE and BE. Obviously, there will be cases where no synergy will be found, whatever the method is, but in some cases as in Table 13, combining more than a single method can make a significant difference.

Fig. 1 depicts an example of the interaction between BE and PE for the dataset GLUCOSE, with $m = 8$. With a suitable computed class segmentation function, it is possible to increase the accuracy achieved by each method individually. As stated above, this is not always the case, and the possible synergy has to be assessed on a case-by-case basis. It will also depend on the complexity of the segmentation function employed, and on the match between the shape of the clusters and the clustering algorithm applied: spherical, linear, or other non-linear shapes. In a more general way, when one of the methods fails, either PE or BE, it is more likely the other yields a significantly better performance, since complementarity seems to be even more frequent than synergy (Tables 5, 8, 10 11).

5. Conclusions

Every year quite many new tools to quantify the dynamical features of time series are described in the scientific literature. These new tools are claimed to be more efficient in algorithmic terms, more

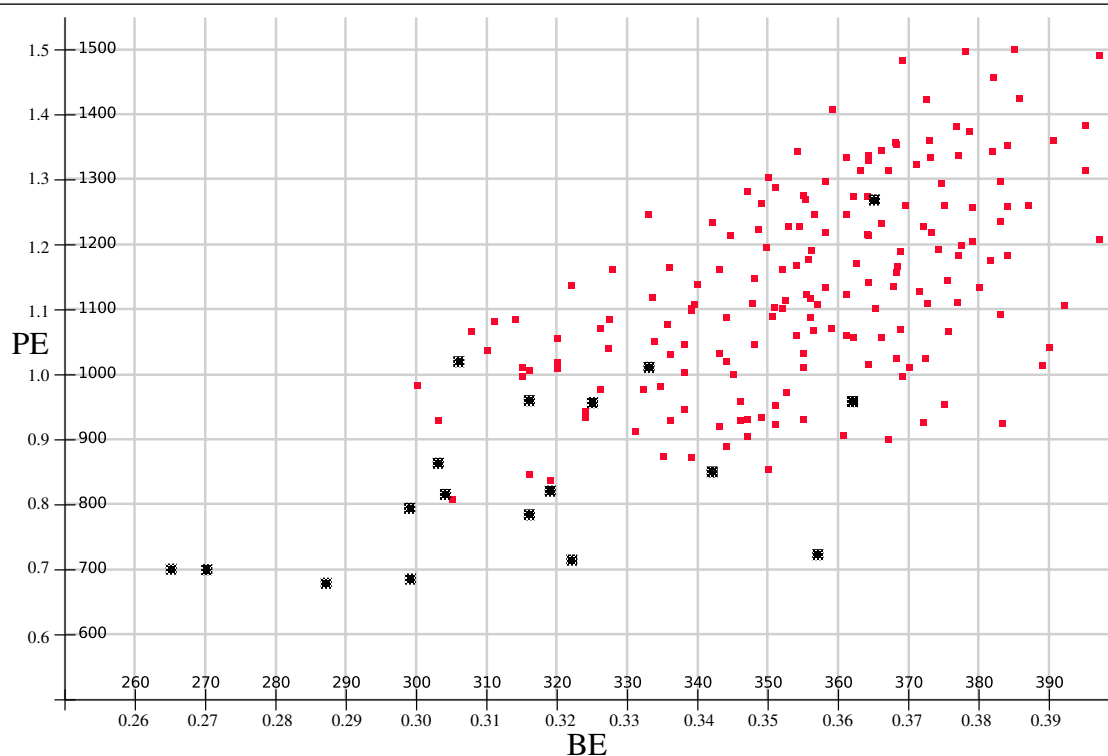


Figure 1. Example of BE–PE plane for GLUCOSE dataset, with $m = 8$. Big black squares represent patients that became diabetic. Small red squares correspond to patients that did not develop a diabetes during the study.

sensitive, more robust, or less dependent on input parameters, among many other possible benefits. In this regard, BE is a recently proposed measure that exploits the effort to sort subsequences instead of the ordinal patterns obtained [32].

BE was presented as an improvement over PE, less influenced by the time series length, the specific value of the input parameter m , and with a better discriminating power [32]. The present study assessed this discriminant power using several different datasets, and we would conclude that the BE discriminating capability was not clearly better, but complementary, BE succeeded where PE failed, and the other way round. This led us to try to combine both measures in a single method to take advantage of BE and PE strengths and simultaneously minimise their weaknesses. The scheme used was based on a clustering algorithm, equivalent to a linear discriminant function.

According to the results obtained, the combination of methods can be considered a new strategy worth examining in the field of time series classification. It should not be claimed to be a *cure-all* method, but the classification performance confirmed the hypothesis in some cases, and it seems to be able to exploit the synergy between PE and BE provided there is some customisation to the problem at hand.

Moreover, classification performance should not be the only item to be maximised in quantitative terms. Since there are many datasets where only one measure is able to find significant differences, other customisations should be devised to increase the robustness of the classification. In other words, new algorithms able to automatically exploit the differences provided by a measure, and minimise the influence of the confounding measure, should be proposed. This way, it would be possible to

have a method able to work with a wider input range of time series or properties, with, for example, statistically significant results in more datasets used in the present study.

There are many clustering algorithms described in the scientific literature, most of them more robust than the example used in this paper. In addition to implementation issues to save memory requirements or computational cost, the algorithms could be improved in terms of accuracy, unknown number of classes, optimality or convergence.

This approach will need further studies using other databases and other non-linear features. The direct combination of ordinal patterns and sorting relations in a single statistic should also be investigated, or the introduction of the BE information into PE computation as a some kind of histogram weighting. The influence of time series length [59], equal values in the subsequences [26], and time delay τ , could also be characterised. Further integration with other PE improvements could be worth exploring [30, 23, 20, 4, 11].

Acknowledgements

No funding was received to support this research work.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. C. Bandt and B. Pompe, Permutation entropy: A natural complexity measure for time series, *Phys. Rev. Lett.*, **88** (2002), 174102.
2. M. Zanin, L. Zunino, O. A. Rosso and D. Papo, Permutation entropy and its main biomedical and econophysics applications: A review, *Entropy*, **14** (2012), 1553–1577.
3. D. Cuesta-Frau, P. Miró-Martínez, S. Oltra-Crespo, J. Jordán-Núñez, B. Vargas, P. González, et al., Model selection for body temperature signal classification using both amplitude and ordinality-based entropy measures, *Entropy*, **20**, (2018).
4. D. Cuesta-Frau, P. Miró-Martínez, S. Oltra-Crespo, J. Jordán-Núñez, B. Vargas and L. Vigil, Classification of glucose records from patients at diabetes risk using a combined permutation entropy algorithm, *Comput. Meth. Program. Biomed.*, **165** (2018), 197–204.
5. D. Mateos, J. Diaz and P. Lamberti, Permutation entropy applied to the characterization of the clinical evolution of epileptic patients under pharmacological treatment, *Entropy*, **16** (2014), 5668–5676.
6. N. Nicolaou and J. Georgiou, The use of permutation entropy to characterize sleep electroencephalograms., *Clin. EEG Neurosci.*, **421** (2011), 24–28.
7. A. Martínez-Rodrigo, B. García-Martínez, L. Zunino, R. Alcaraz and A. Fernández-Caballero, Multi-lag analysis of symbolic entropies on eeg recordings for distress recognition, *Front. Neuroinform.*, **13** (2019), 40.

8. D. Li, X. Li, Z. Liang, L. J. Voss and J. W. Sleight, Multiscale permutation entropy analysis of EEG recordings during sevoflurane anesthesia, *J. Neural Eng.*, **7** (2010), 046010.
9. C. C. Naranjo, L. M. Sanchez-Rodriguez, M. B. Martínez, M. E. Báez and A. M. García, Permutation entropy analysis of heart rate variability for the assessment of cardiovascular autonomic neuropathy in type 1 diabetes mellitus, *Comput. Biol. Med.*, **86** (2017), 90–97.
10. A. G. Ravelo-Garcia, J. L. Navarro-Mesa, U. Casanova-Blancas, S. González, P. Quintana, I. Guerra-Moreno, et al., Application of the permutation entropy over the heart rate variability for the improvement of electrocardiogram-based sleep breathing pause detection, *Entropy*, **17** (2015), 914–927.
11. C. Bian, C. Qin, Q. D. Y. Ma and Q. Shen, Modified Permutation-entropy analysis of heartbeat dynamics, *Phys. Rev. E*, **85** (2012), 021906.
12. M. Henry and G. Judge, Permutation entropy and information recovery in nonlinear dynamic economic time series, *Econometrics*, **7** (2019).
13. H. Danylchuk, N. Chebanova, N. Reznik and Y. Vitkovskyi, Modeling of investment attractiveness of countries using entropy analysis of regional stock markets, *Global J. Environ. Sci. Manag.*, **5** (2019), 227–235.
14. F. Siokis, Credit market jitters in the course of the financial crisis: A permutation entropy approach in measuring informational efficiency in financial assets, *Phys. A Statist. Mechan. Appl.*, **499** (2018).
15. A. F. Bariviera, L. Zunino, M. B. Guercio, L. Martinez and O. Rosso, Efficiency and credit ratings: A permutation-information-theory analysis, *J. Statist. Mechan. Theory Exper.*, **2013** (2013), P08007.
16. A. F. Bariviera, M. B. Guercio, L. Martinez and O. Rosso, A permutation information theory tour through different interest rate maturities: the libor case, *Philos. Transact. Royal Soc. A Math. Phys. Eng. Sci.*, **373** (2015).
17. J. Cánovas, G. García-Clemente and M. Muñoz-Guillermo, Comparing permutation entropy functions to detect structural changes in time series, *Phys. A Statist. Mechan. Appl.*, **507** (2018), 153–174.
18. J. Zhang, Y. Zhao, M. Liu and L. Kong, Bearings fault diagnosis based on adaptive local iterative filtering–multiscale permutation entropy and multinomial logistic model with group–lasso, *Advan. Mechan. Eng.*, **11** (2019), 1687814019836311.
19. J. Huang, X. Wang, D. Wang, Z. Wang and X. Hua, Analysis of weak fault in hydraulic system based on multi-scale permutation entropy of fault-sensitive intrinsic mode function and deep belief network, *Entropy*, **21** (2019).
20. B. Fadlallah, B. Chen, A. Keil and J. Principe, Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information, *Phys. Rev. E*, **87** (2013), 022911.
21. J. Garland, T. R. Jones, M. Neuder, V. Morris, J. W. C. White and E. Bradley, Anomaly detection in paleoclimate records using permutation entropy, *Entropy*, **20** (2018).

22. B. Deng, L. Cai, S. Li, R. Wang, H. Yu, Y. Chen, et al., Multivariate multi-scale weighted permutation entropy analysis of eeg complexity for alzheimer's disease, *Cogn. Neurodyn.*, **11** (2017), 217–231.
23. L. Xiao-Feng and W. Yue, Fine-grained permutation entropy as a measure of natural complexity for time series, *Chinese Phys. B*, **18** (2009), 2690.
24. D. Cuesta-Frau, Permutation entropy: Influence of amplitude information on time series classification performance, *Math. Biosci. Eng.*, **5** (2019), 1–16.
25. F. Traversaro, M. Risk, O. Rosso and F. Redelico, An empirical evaluation of alternative methods of estimation for Permutation Entropy in time series with tied values, *arXiv e-prints*, arXiv:1707.01517 (2017).
26. D. Cuesta-Frau, M. Varela-Entrecanales, A. Molina-Picó and B. Vargas, Patterns with equal values in permutation entropy: Do they really matter for biosignal classification?, *Complexity*, **2018** (2018), 1–15.
27. Y. Zou, R. V. Donner, N. Marwan, J. F. Donges and J. Kurths, Complex network approaches to nonlinear time series analysis, *Phys. Rep.*, **787** (2019), 1–97, Complex network approaches to nonlinear time series analysis.
28. M. McCullough, M. Small, T. Stemler and H. Iu, Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems, *Chaos Interdisciplin. J. Nonlinear Sci.*, **25** (2015).
29. D. Cuesta-Frau, A. Molina-Picó, B. Vargas and P. González, Permutation entropy: Enhancing discriminating power by using relative frequencies vector of ordinal patterns instead of their shannon entropy, *Entropy*, **21** (2019).
30. H. Azami and J. Escudero, Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation, *Comput. Meth. Program. Biomed.*, **128** (2016), 40–51.
31. Z. Chen, L. Yaan, H. Liang and J. Yu, Improved permutation entropy for measuring complexity of time series under noisy condition, *Complexity*, **2019** (2019), 1–12.
32. G. Manis, M. Aktaruzzaman and R. Sassi, Bubble entropy: An entropy almost free of parameters, *IEEE Transact. Biomed. Eng.*, **64** (2017), 2711–2718.
33. M. Riedl, A. Müller and N. Wessel, Practical considerations of permutation entropy, *European Phys. J. Special Topics*, **222** (2013), 249–262.
34. L. Zunino, F. Olivares, F. Scholkmann and O. A. Rosso, Permutation entropy based time series analysis: Equalities in the input signal can lead to false conclusions, *Phys. Lett. A*, **381** (2017), 1883–1892.
35. L. Citi, G. Guffanti and L. Mainardi, Rank-based multi-scale entropy analysis of heart rate variability, in *Computing in Cardiology 2014*, 2014, 597–600.
36. A. M. Unakafov and K. Keller, Conditional entropy of ordinal patterns, *Phys. D Nonlinear Phenom.*, **269** (2014), 94–102.
37. Z. Liang, Y. Wang, X. Sun, D. Li, L. Voss, J. Sleight, et al., Eeg entropy measures in anesthesia, *Front. Comput. Neurosci.*, **9** (2015), 16.

38. D. E. Lake, J. S. Richman, M. P. Griffin and J. R. Moorman, Sample entropy analysis of neonatal heart rate variability, *Am. J. Physiology-Regulatory Integrat. Comparat. Physiol.*, **283** (2002), R789–R797, PMID: 12185014.
39. S. M. Pincus, Approximate entropy as a measure of system complexity., *Proceed. Nat. Acad. Sci.*, **88** (1991), 2297–2301.
40. T. Fawcett, An introduction to ROC analysis, *Patt. Recogn. Lett.*, **27** (2006), 861–874, ROC Analysis in Pattern Recognition.
41. I. Unal, Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach, *Comput. Math. Methods Med.*, **2017** (2017), 14.
42. A. Tharwat, Classification assessment methods, *Appl. Comput. Inform.*, (2018).
43. A. M. Zoubir and D. R. Iskander, *Bootstrap Techniques for Signal Processing*, Cambridge University Press, 2004.
44. D. Kalpić, N. Hlupić and M. Lovrić, *Student's t-Tests*, 1559–1563, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
45. C.-Y. J. Peng, K. L. Lee and G. M. Ingersoll, An introduction to logistic regression analysis and reporting, *J. Educat. Res.*, **96** (2002), 3–14.
46. C. M. Bishop, *Neural Networks Patt. Recogn.*, Oxford University Press, Inc., New York, NY, USA, 1995.
47. A. K. Jain, M. N. Murty and P. J. Flynn, Data clustering: A review, *ACM Comput. Surv.*, **31** (1999), 264–323.
48. J. Rodríguez-Sotelo, D. Peluffo-Ordoñez, D. Cuesta-Frau and G. Castellanos-Domínguez, Unsupervised feature relevance analysis applied to improve ecg heartbeat clustering, *Comput. Meth. Program. Biomed.*, **108** (2012), 250–261.
49. D. Cuesta-Frau, J. C. Pérez-Cortés and G. Andreu-García, Clustering of electrocardiograph signals in computer-aided Holter analysis, *Comput. Meth. Program. Biomed.*, **72** (2003), 179–196.
50. C. Murthy and N. Chowdhury, In search of optimal clusters using genetic algorithms, *Patt. Recogn. Lett.*, **17** (1996), 825–832.
51. J. Sander, M. Ester, H.-P. Kriegel and X. Xu, Density-based clustering in spatial databases: The algorithm gbscan and its applications, *Data Min. Knowl. Discov.*, **2** (1998), 169–194.
52. J. Wu, *Advances in K-means Clustering: A Data Mining Thinking*, Springer Publishing Company, Incorporated, 2012.
53. S. Panda, S. Sahu, P. Jena and S. Chattopadhyay, Comparing fuzzy-c means and k-means clustering techniques: A comprehensive study, in *Advances in Computer Science, Engineering & Applications* (eds. D. C. Wyld, J. Zizka and D. Nagamalai), Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, 451–460.
54. A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, et al., PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, *Circulation*, **101** (2000), 215–220.

55. J. Jordán-Núñez, P. Miró-Martínez, B. Vargas, M. Varela-Entrecanales and D. Cuesta-Frau, Statistical models for fever forecasting based on advanced body temperature monitoring, *J. Crit. Care*, **37** (2017), 136–140.
56. D. Cuesta-Frau, M. Varela, P. Miró-Martínez, P. Galdós, D. Abásolo, R. Hornero, et al., Predicting survival in critical patients by use of body temperature regularity measurement based on approximate entropy, *Med. Biol. Eng. Comput.*, **45** (2007), 671–678.
57. C. Rodríguez de Castro, L. Vigil, B. Vargas, E. García Delgado, R. García-Carretero, J. Ruiz-Galiana, et al., Glucose time series complexity as a predictor of type 2 diabetes, *Diabetes Metab. Res. Rev.*, **30** (2016).
58. R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David and 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** (2001), 061907.
59. D. Cuesta-Frau, J. P. Murillo-Escobar, D. A. Orrego and E. Delgado-Trejos, Embedded dimension and time series length. practical influence on permutation entropy and its applications, *Entropy*, **21** (2019).
60. N. Iyengar, C. K. Peng, R. Morin, A. L. Goldberger and L. A. Lipsitz, Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics, *Am. J. Physiology-Regulatory Integrat. Comparat. Physiol.*, **271** (1996), R1078–R1084, PMID: 8898003.



AIMS Press

©2020 the author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)