A new topological entropy-based approach for measuring similarities among piecewise linear functions

Matteo Rucco\textsuperscript{a,c,*}, Rocio Gonzalez-Diaz\textsuperscript{b,**}, Maria-Jose Jimenez\textsuperscript{b}, Nieves Atienza\textsuperscript{b}, Cristina Cristalli\textsuperscript{c}, Enrico Concettoni\textsuperscript{c}, Andrea Ferrante\textsuperscript{c}, Emanuela Merelli\textsuperscript{a}

\textsuperscript{a} University of Camerino, School of Science and Technology, Computer Science Division, Camerino, Italy
\textsuperscript{b} University of Seville, School of Computer Engineering, Applied Math Dept, Seville, Spain
\textsuperscript{c} Research for Innovation Group, Loccioni Group, Angeli di Rosora (AN), Italy

**MSC:** 55U10, 05E45, 62H30, 28D20

**Keywords:** Piecewise linear functions, Noisy signals, Persistent homology, Persistent entropy, Supervised classification

**ABSTRACT**

In this paper we present a novel methodology based on a topological entropy, the so-called persistent entropy, for addressing the comparison between discrete piecewise linear functions. The comparison is certified by the stability theorem for persistent entropy that is presented here. The theorem is used in the implementation of a new algorithm. The algorithm transforms a discrete piecewise linear function into a filtered simplicial complex that is analyzed via persistent homology and persistent entropy. Persistent entropy is used as a discriminant feature for solving the supervised classification problem of real long-length noisy signals of DC electrical motors. The quality of classification is stated in terms of the area under receiver operating characteristic curve (AUC=93.87%).

1. Introduction

Piecewise linear function (PL) is a powerful mathematical tool largely used for approximating signals. The task of measuring the similarity among piecewise linear functions (PLs) is still an open issue and a solution is strongly required in machine learning methods. The comparison between the area under the curves (AUCs) of discrete digital signals is a weak measure: for each value of AUC, a family of finite signals exists. Several approaches for measuring similarity among PLs have been reported in literature. To the best of our knowledge, the most relevant techniques are:

1. distance-based methods, e.g. Pompeiu-Hausdorff distance \cite{1};
2. similarity based on global descriptors: time-frequency analysis, wavelet analysis, mutual information, (Normalized/Zero-Mean) cross-correlation, sum of squared differences, etc \cite{2,3};
3. distances among “bags of local features” \cite{4}.

Generally, the identification of common patterns among signals suffers from the shifting problem \cite{5}. Formally, given two 2-dimensional signals \((S_1, S_2)\), that are two ordered collections of real points, signal \(S_2\) is shifted with respect to signal \(S_1\) if \((x_2 = x_1 + n a d l / o y_2 = y_1 + m,\) where \(m, n \in \mathbb{R}\)). For this reason, distance-based methods without pre-alignment cannot be directly applied. Pre-alignment techniques are in general highly time-consuming. A potential solution is represented by the dynamic time warping (DTW) \cite{6}. However, DTW is not computationally convenient for compiling a pair-wise distance matrix in the case of several long-length signals \cite{7}.

Both similarity based on global descriptors and distances among “bags of local features” need the extraction of local or global features from the PLs. The features can be used for extrapolating several useful information (e.g., periodicity, state transitions, chaotic behaviors, etc...) but there is not a unique criteria for deciding which features are completely informative and often it is mandatory to execute a feature selection (or reduction) procedure.

Mention should also be made on Fourier transformation and Wavelet decomposition as two main approaches in the literature for studying signals in general. Fourier transform decomposes a signal into a family of complex trigonometric functions (i.e., sinusoids), while Wavelet transform decomposes it into a family of wavelets. Sinusoids are typically symmetric, smooth and regular, while wavelets can be either symmetric or asymmetric, smooth or sharp, regular or irregular.
Hence, Fourier can be seen as a sub-case of Wavelet transform because sinusoids are one of the possible wavelets used by Wavelet analysis. Entropy measures can be defined from the Fourier transformation and from the Wavelet decomposition of signals. The former is known as *Power Spectral Entropy* (PSE), the latter is indicated as *Wavelet Shannon Entropy* (WSE) [8,9].

There are many works in the literature dealing with numerous applications of different entropy measures. For instance, in [10] the concept of fuzzy entropy is used to effectively segment an image by multiple thresholding; in [11], using theory of entropy-based information fusion, a dynamic trust model for P2P network societies is proposed, in which multiple complementary decision factors are incorporated to reflect complexity and uncertainty of trust relationship; in [12], in the context if iris recognition, they propose a fusion method from four matching algorithms, using relative entropy measure to assign low weighting coefficients to features with less information and higher weights to features with more information.

In this paper we present a new method that carries out the comparison among the shape of PLs. The methodology described in this article is based on topology. Roughly speaking, topology is a branch of pure mathematics that deals with the analysis of connectivities on spaces. Briefly, our approach represents a PL as a topological space, namely a filtered simplicial complex, that is qualitatively described by *persistent homology* (under the form of a persistence barcode) and quantitatively measured by *persistent entropy*. Thanks to the *stability theorem for persistent entropy* that is presented in this paper, persistent entropy can be used as a unique global feature for comparing signals. We observe that this quantity is invariant with respect to translation of axis. It means that given two signals with the same shape, even if they are shifted along one or both directions, they have the same persistent entropy. Hence, persistent entropy is a prominent feature for signals characterization.

A precursor of the persistent entropy definition was given in [13] to measure how different bars of a persistence barcode are in length. Persistent entropy was defined in [14], where it was used for the analysis of the behavior of the human immune system. The authors dealt with a collection of networks representing the temporal evolution of the human immune system in a simulated setting. The nodes of the networks represented the antibodies and an edge existed if and only if two antibodies interacted. In order to capture higher order relationships among antibodies, the network was endowed with a structure of a filtered simplicial complex that was studied by persistent homology.

From the persistence barcodes, they computed the persistent entropy and, by looking and the chart of the persistent entropy vs. time, they recognized peaks or plateaus. The peaks represent when the immune system reacts against external stress (antigens). The observations obtained from the chart were used for modeling the behavior of the immune system by using automata theory. It was the first approach for extracting a *language* from the data produced by a time-dependent system. The automata that encodes the temporal evolution of the persistent entropy for the analysis of the immune system is known as Persistent Entropy Automaton (PEA) and it was published in [15]. In [16], the authors projected the signals produced by small DC motors (the same type of data analyzed in this paper) into a new \( \mathbb{R}^2 \) space by using time-delay embedding techniques. Vietoris-Rips complexes (a specific type of simplicial complexes) were constructed out of that produced data. The classification of the motors was accomplished then by using persistent homology of the new space.

The paper is organized as follows, in Section 2 we recall the mathematical background for understanding the methodology described in Section 3. Section 3 also contains both the proof of the *stability theorem for persistent entropy* and an algorithm for the comparison between PLs. The application of the methodology to a real case study is reported in SubSection Section 3.2. Moreover, we report on the comparison of the methodology with the Fourier and Wavelet based entropies. Section 4 is devoted to briefly highlight the main results of our paper and point out relevant observations.

### 2. Background

#### 2.1. Topology

A topological space is a powerful mathematical concept for describing the connectivity of a space. Informally, a *topological space* is a set of points, each of them equipped with the notion of *neighboring*. More formally, it is defined as a set endowed with a *topology* as follows:

**Topology.** A topology on a set of points \( X \) is a family of subsets \( T \subseteq 2^X \) satisfying that

- a) If \( S_1, S_2 \in T \), then \( S_1 \cap S_2 \in T \);
- b) If \( \{S_i| i \in J\} \subseteq T \), then \( \bigcup_{i \in J} S_i \in T \);
- c) \( \emptyset, X \in T \).

**Topological space.** The pair \( (X, T) \) of a set \( X \) and a topology \( T \) is a topological space \( \mathcal{E} \).

One way to represent a topological space is by decomposing it into *simple* pieces such that their common intersections are lower-dimensional pieces of the same kind. In this paper, we use (abstract) simplicial complexes as the data structure to represent topological spaces.

**Abstract simplicial complex.** An abstract simplicial complex \( K \) is given by:

- a set \( V \) of vertices (also called *vertices*);
- for each \( k \geq 1 \) a set of \( k \) vertices \( \sigma = \{v_0, v_1, \ldots, v_k\} \), where \( v_i \in V \);
- each \( k \) simplex has \( k+1 \) faces obtained by removing one of the vertices;
- if \( \sigma \) belongs to \( K \), then all the faces of \( \sigma \) must belong to \( K \).

**Simplicial complex.** A simplicial complex \( K \) is a geometric realization of an abstract simplicial complex \( K \). A simplicial complex is obtained by a nested family of simplices: a \( 0 \)-simplex can be thought as a point, a \( 1 \)-simplex as an edge, a \( 2 \)-simplex as a filled triangle and a \( 3 \)-simplex as a filled tetrahedron.

From now on, both the abstract simplicial complex and its geometric representation will be called simplicial complex and denoted by \( K \), to simplify the exposition. See [17,18] for an introduction to algebraic topology.

#### 2.2. Persistent homology

Homology is an algebraic machinery used for describing a topological space \( \mathcal{E} \). The \( k \) \textquotesingle{}\textquotesingle{} Betti number represents the rank of the \( k \) \textquotesingle{}\textquotesingle{} dimensional homology group. Informally, for a fixed \( k \), the \( k \) \textquotesingle{}\textquotesingle{} Betti number \( \beta_k \) counts the number of \( k \) \textquotesingle{}\textquotesingle{} dimensional holes characterizing \( \mathcal{E} \); \( \beta_0 \) is the number of connected components, \( \beta_1 \) counts the number of holes in 2D or tunnels in 3D, \( \beta_2 \) can be thought as the number of voids in geometric solids.

Persistent homology is a method for computing \( k \) \textquotesingle{}\textquotesingle{} dimensional holes at different spatial resolutions. Persistent holes are more likely to represent true features of the underlying space, rather than artifacts of sampling (noise), or particular choice of parameters. For a more formal description we refer the reader to [19]. In order to compute persistent homology, we need a distance function on the underlying space. This can be obtained constructing a *filtration* on the simplicial complex, that is a nested sequence of increasing subcomplexes. More formally, a filtered simplicial complex is a collection of subcomplexes \( \{K(t) : t \in \mathbb{R}\} \) of \( K \) such that \( K(t) \subset K(s) \) for \( t < s \) and there exists \( t_{\max} \in \mathbb{R} \) such that

\[ \text{if } d \leq t_{\max} \text{ then } K(t) \text{ is connected}. \]
The filtration time (or filter value) of a simplex \( \sigma K \in K \) is the smallest \( t \) such that \( \sigma K t \in (K(t)) \).

Persistent homology describes how the homology of \( K \) changes along a filtration \( K(t); t \in \mathbb{R} \). A \( k \)-dimensional Betti interval, with endpoints \( [t_{\text{start}}, t_{\text{end}}] \), corresponds to a \( k \)-dimensional hole that appears at filtration time \( t_{\text{start}} \) and remains until time \( t_{\text{end}} \). We refer to the holes that are still present at \( t = t_{\text{max}} \) (the value in \( \mathbb{R} \) such that \( K_{\text{max}} = K \)) as persistent topological features, otherwise they are considered topological noise [20]. The set of intervals representing birth and death times of homology classes is called the persistence barcode associated to the corresponding filtration.

Instead of bars, we sometimes draw points in the plane such that a point \( (x, y) \in \mathbb{R}^2 \) (with \( x < y \)) corresponds to a bar \([x, y]\) in the persistence barcode. This set of points is called persistence diagram.

In the case of an interval with no death time, \([t_{\text{start}}, +\infty)\), several approaches can be considered, such as extending real numbers including \( +\infty \), removing or truncating infinite intervals or using extended persistence [21–23]. In this paper, we truncate infinite intervals and replace \([t_{\text{start}}, +\infty)\) by \([t_{\text{start}}, m]\) in the persistence barcode, where \( m = t_{\text{max}} + 1 \).

2.3. Persistent entropy

In order to measure how much is ordered the construction of a filtered simplicial complex, a new entropy measure, the so-called persistent entropy, was defined in [14]. A precursor of this definition was given in [13] to measure how different bars of a persistence barcode are in length. Here we recall the definition.

**Persistent entropy.** Given a filtered simplicial complex \( [K(t); t \in \mathbb{R}] \), and the corresponding persistence diagram \( D = \{(x_i, y_i); i \in I\} \) (being \( x_i < y_i \) for all \( i \in I \)), the persistent entropy \( H \) of the filtered simplicial complex is calculated as follows:
\[ H = \sum_{i \in I} p_i \log(p_i) \]

where \( p_i = \frac{\ell_i}{L} \), \( \ell_i = y_i - x_i \), and \( L = \sum_{i \in I} \ell_i \).

Note that, when topological noise is present, there can be more than one point, denoted by \((x_i, y_i)\), with multiplicity greater than 1 (see [19], page 152).

Observe that the maximum persistent entropy corresponds to the situation in which all the intervals in the barcode are of equal length. In that case, \( H = \log n \) if \( n \) is the number of elements of \( I \). Conversely, the value of the persistent entropy decreases as more intervals of different length are present.

3. Topological comparison of plots

The aim of our paper is to address the problem of comparison between the shape of plots. In order to satisfy this task, instead of using metric spaces (e.g., DTW), we propose to study the topology of the shape of the plots. The methodology is completely based on algebraic topology: first, transform a plot into a filtered 1-dimensional simplicial complex that is analyzed by persistent homology. Then, from the homological groups, compute the persistent entropy and finally, persistent entropy is used as a one-dimensional feature that characterizes the signal.

This section is organized as follows: in Section 3.1 we introduce a formal description of the methodology and we derive an algorithm for its computation; in Section 3.2 we apply the methodology to a real case.
stability of persistence diagram \[24\], page 105

Given two PL functions \(f, g : K \to \mathbb{R}\) respectively computed: Suppose that we have two discrete signals \(f, g\) and, in particular, at least 40 times the difference between \(f\) and \(g\), and, in particular, at least 40 times the machine precision \(c\). In practice, this is a common case.

Consider, for example, the two signals \(f_1\) and \(f_2\) in Fig. 5. In this signals, the machine precision is \(c = 1.4386 \times 10^{-14}\). The number of points in the respective persistent diagrams are: \(|D(f_1)| = 421\) and \(|D(f_2)| = 7213\), being the total lengths: \(L_1 = 488.5810\) and \(L_2 = 9852.5\). Average lengths are: \(\bar{c} = 1.6053\) and \(\bar{c} = 1.3285\). Observe that \(f_1\) is the 0.814044 times \(c\) and \(f_2\) is the 0.900414 times \(c\). Finally, \(H(f_1) = 5.9977\) and \(H(f_2) = 8.7997\).

Notice that the application of the theorem above to two discrete signals is possible whenever they are defined on the same support set, union of the first \(i\) lower stars.

- Finally, persistent entropy \(H(f)\) is computed.

**Persistent Entropy Stability Theorem.** Given two PL functions on simplicial complexes embedded in \(\mathbb{R}^d\), \(f, g : K \to \mathbb{R}\), for every \(\epsilon > 0\), there exists \(\delta > 0\) such that

\[\|f - g\|_\infty \leq \delta \Rightarrow |H(f) - H(g)| \leq \epsilon.\]

**Proof.** The bottleneck distance between the persistence diagrams \(D(f)\) and \(D(g)\) associated to the lower-star filtrations of \(f\) and \(g\) is:

\[d_B(D(f), D(g)) = \inf_{\gamma} \sup_{a, b} \|a - \gamma(a)\|_\infty,\]

where, for points \(a = (x, y)\) and \(\gamma(a) = (\gamma_x, \gamma_y)\), \(\|a - \gamma(a)\|_\infty = \max\{|x - \gamma_x|, |y - \gamma_y|\}\) and the bijection \(\gamma : D(f) \to D(g)\) can associate a point off the diagonal with another on or off the diagonal. See \[19\], page 229.

Now, let \(n = \max\{|D(f)|, |D(g)|\}\) (here, \(|D(f)|\) and \(|D(g)|\) are, respectively, the number of points in \(D(f)\) and \(D(g)\)). Let \(\tau : D(f) \to D(g)\) be the bijection such that \(d_B(D(f), D(g)) = \sup_{a \in D(f)} \|a - \tau(a)\|_\infty\). Then \(D(f) = \{a_1, \ldots, a_n\}\) and \(D(g) = \{a'_1, \ldots, a'_n\}\), being \(a'_i = \tau(a_i)\) for all \(i\).

Since \(h(x) = -\log x\) is a continuous function in \([0, 1]\) (redefining \(h(0)\) as 0), for \(\epsilon' > 0\), there exists \(\delta \in (0, 1]\) such that if \(|x - x'| \leq \delta\) then \(|h(x) - h(x')| \leq \epsilon'.\)

Stability of persistence diagram \[24\], page 105 establishes that \(d_B(D(f), D(g)) \leq \|f - g\|_\infty\), where

\[\|f - g\|_\infty = \sup_{q \in \mathbb{R}^d} |f(q) - g(q)|, \quad q \in \mathbb{R}^d.\]

So if \(\|f - g\|_\infty \leq \delta\), then \(||f - \tau|| \leq \epsilon'\) for all \(i\) and \(L = L' \leq 2\delta n\).

Recall that if \(a_i = (x_i, y_i)\), then \(c_i = y_i - x_i\) and \(L = \sum c_i L_i\).

Without loss of generality, assume that \(L \geq L'\). Let \(\delta = \frac{L - L'}{4\epsilon'} > 0\). Then, \(|p_i - p'_i| \leq \delta\) for all \(i\), since:

- \(p_i - p'_i = \frac{c_i}{L} - \frac{c'_i}{L'} = \frac{Lc'_i - L'c_i}{LL'} \leq \frac{c_i - c'_i}{L} + \frac{L'c'_i - Lc_i}{L} \leq \frac{\epsilon'}{L} + \frac{\epsilon'}{L'} \leq \delta\)
- \(p'_i - p_i \leq \frac{2(L'c_i - Lc'_i)}{LL'} \leq \frac{\epsilon'}{L} + \frac{\epsilon'}{L'} \leq \delta\),

\[\leq \delta\]

Therefore,

\[|H(f) - H(g)| = \sum_{i=1}^n p_i \log p_i - \sum_{i=1}^n p'_i \log p'_i \leq \sum_{i=1}^n p_i \log p_i - p'_i \log p'_i \leq \epsilon\]

which concludes the proof.$^2$

In practical terms, the parameter \(\delta\) in the above theorem can be effectively computed: Suppose that we have two discrete signals \(\tau : \mathbb{R}^d \to \mathbb{Z}\) and \(\tau : \mathbb{R}^d \to \mathbb{Z}\), being \(c\) a small positive number (machine precision). Suppose that \(D(f) \leq D(g)\) and \(D(g)\) has \(n\) points. Since \(h(x) = -\log x\) in \([0, 1]\), then \(c' \in (0, \frac{1}{e})\) and \(c \in (0, \frac{1}{e})\). If, for example, \(n = 1000\), then \(c' \in (0, 0.368)\). Write \(L' = n'c'\), being \(c'\) the average length of all the intervals in \(D(g)\). Then \(\delta = \frac{4d}{4d + \frac{d}{c'}} = \frac{\epsilon}{c'}\) to have \(0 < \delta < 4d\). We need \(c'\) to be large with respect to \(\delta\). For example, for \(\delta \leq 0.1\), we need \(c' > 40d\), which is equivalent to say that the average between consecutive local max and local min in \(g\) should be at least 40 times the difference between \(f\) and \(g\), and, in particular, at least 40 times the machine precision \(c\). In practice, this is a common case.

Consider, for example, the two signals \(f_1\) and \(f_2\) in Fig. 5. In this signals, the machine precision is \(c = 1.4386 \times 10^{-14}\). The number of points in the respective persistent diagrams are: \(|D(f_1)| = 421\) and \(|D(f_2)| = 7213\), being the total lengths: \(L_1 = 488.5810\) and \(L_2 = 9852.5\). Average lengths are: \(\bar{c} = 1.6053\) and \(\bar{c} = 1.3285\). Observe that \(f_1\) is the 0.814044 times \(c\) and \(f_2\) is the 0.900414 times \(c\). Finally, \(H(f_1) = 5.9977\) and \(H(f_2) = 8.7997\).

Notice that the application of the theorem above to two discrete signals is possible whenever they are defined on the same support set,
what means equal length signals and with equal sampling rates.

For practical scope, the methodology explained above can be translated into the following algorithm designed for analyzing a 2-dimensional plot.

Suppose that the first coordinate of a point in $\mathbb{R}^2$ represents time.

Given a set of sample points of a signal $S \subset \mathbb{R}^2$:

- order the points in $S$ with respect to their first coordinate (i.e., order the points in $S$ by time);
- transform $S$ into a filtered simplicial complex:
  1. each point of $S$ is a 0-simplex with filter value equals to its second coordinate.
  2. Each pair formed by two consecutive points in $S$: $(x_i, y_i)$ and $(x_{i+1}, y_{i+1}) \in S$ where $x_i < x_{i+1}$, forms a 1-simplex $\sigma$ with filter

\(Fig. 6.\) Entropies for the 23 good signal (blue points) and for the 23 faulty signals (red crosses). From top to bottom: Persistent entropy, Spectral entropy, Wavelet entropy. It is well evident that Persistent entropy and Spectral entropies clearly identify the two classes. However, for the dataset under examination Spectral entropy counts more false positive than Persistent entropy (red points below the dotted line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
value \( f(\sigma) = \max \{y, y_+\} \). Note that the resulting filtration is obtained by presenting at the beginning the simplices formed with the lowest second coordinate (i.e., the filter-value set \( F \) is obtained by spanning the \( y \)-axis in upward direction).

The resulting filtration is a lower-star filtration.

- compute persistent entropy. Fig. 1 shows a toy example of the application of the methodology. From left to right: a) The input signal formed by three time points, respectively with coordinates: (1, 0), (2, 2), (3, 1). b) The filtered simplicial complex formed by three 0 − simplices: \( \{v_0, v_1, v_2\} \) with filter values \( f(v_0) = 0, f(v_1) = 1, f(v_2) = 2 \) and two 1 − simplices: \( \{e_1, e_2\} \), with filter values \( f(e_1) = f(e_2) = 2 \), so the filter-value set is \( F = \{0, 1, 2\} \). c) As for the persistence barcode: at \( F=0 \), there is only one topological feature corresponding to \( v_0 \); at \( F=1 \), \( v_0 \) is still in the space but also a new component is introduced that corresponds to \( v_2 \); for \( F=2 \), the new 0 − simplex \( v_1 \) is added, together with the two 1 − simplices \( e_1 = \{v_0, v_2\} \) and \( e_2 = \{v_2, v_1\} \) that connect the two connected component formed at previous filter values. From this filter value \( F = 2 \) and successive, the space is described by only one persistent connected component, that is, \( \beta_0 = 1 \). Visually there is only one infinite line in the barcode.

The persistent entropy of the space is computed as follows. The maximum filter value is 2, so the symbol “\( \infty \)” representing the persistent topological feature is substituted with the value \( m=3 \). Then, the barcode is formed by two lines with lengths \( \ell_1 = 1 \) and \( \ell_2 = 3 \), respectively, so the total length \( L = 1 + 3 = 4 \). For each line, the probability is given by \( p_1 = 1/4 \) and \( p_2 = 3/4 \), and finally the persistent entropy is \( H=0.5623 \).

In Fig. 2, three examples of different signals with the same support set are shown so that the reader can appreciate that “similar” signals have “similar” entropies while more different signals have more different entropies.

It is also possible to compare signals acquired with the same temporal periods and with different sampling frequency or signals with different temporal periods and with the same sampling frequency. In the former case (see for example Fig. 3) a possible solution, among others, is given by interpolating the more frequently sampled signal (blue) and to evaluate the interpolation on the same points of the reference signal (red). The evaluation returns a piece-wise linear function (green). Now in order to compare the two signals, it is possible to apply our methodology. In the latter case, a possible solution is to segment the longest signal by a sliding window with the same length of the shortest signal and for each window, apply our methodology. This approach should find segments that are similar to the reference signal. This is possible because the entropy is invariant to the position of the signal with respect to both the temporal and the \( y \)-axis (see Fig. 4).

### 3.2. The case study: classification of DC motors

We applied our methodology to 46 small DC motors (see two examples in Fig. 5). For each motor we analyzed the acceleration that has been measured with a B & K single axis 4514-001 IEPE accelerometer for acquiring the radial component of vibration. Signals were sampled at a rate of 50 kHz with a total number of 180,000 time points. For the sake of preciseness, in this work we used a subset of the set reported in [25,16] that is formed by signals with the same length (we only consider signals that has been recorded with the same sampling frequency and for the same temporal period). Each signal is formed by 180,000 points equally time-spaced. The signals were classified by an expert operator in two classes: good motors, and faulty motors based on their vibration and noise level. The number of simplices (vertices

### Table 1

<table>
<thead>
<tr>
<th>Entropy</th>
<th>Hypothesis</th>
<th>p-Value</th>
<th>Confidence Interval</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistent Entropy</td>
<td>1</td>
<td>3.6017e-08</td>
<td>-0.1764; -0.0944</td>
<td></td>
</tr>
<tr>
<td>Spectral Entropy</td>
<td>1</td>
<td>6.3579e-05</td>
<td>-0.5343; -0.1997</td>
<td></td>
</tr>
<tr>
<td>Wavelet Entropy</td>
<td>0</td>
<td>0.1271</td>
<td>-0.4359; 3.3804</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Classifier</th>
<th>AUC</th>
<th>S.E.</th>
<th>95% C.I.</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistent Entropy</td>
<td>0.93870</td>
<td>0.03749</td>
<td>0.86522; 1.00000</td>
<td>Excellent test</td>
</tr>
<tr>
<td>Spectral Entropy</td>
<td>0.82023</td>
<td>0.06306</td>
<td>0.69664; 0.94383</td>
<td>Good test</td>
</tr>
</tbody>
</table>

![ROC curve of the classifier based on Persistent entropy (left) and Spectral entropy (right). The area under curves for Persistent entropy is \( AUC = 93.87\% \), while for Spectral entropy \( AUC = 82.02\% \).](image-url)
and edges) for each signal is 359,999 (the signals have the same number of vertices and edges but the length of the edges are different). The machine precision (minimum value between two signal values) is: $1.4386 \times 10^{-14}$. The algorithm has been coded in MATLAB and for the topological analysis we used the Java package Javaplex [20].

On the same dataset we also computed the Power Spectral entropy $^3$ and the Wavelet entropy $^2$. We briefly recall the mathematical definitions of Spectral Entropy and Wavelet Shannon Entropy, for a complete explanation we refer the reader to [8,9].

**Definition 1 (Power Spectral Entropy (PSE)).** Given a signal $s$ and its spectrum of frequencies $X(\omega_i)$. The Power Spectral Density (PSD) of $s$ is given by:

$$P(\omega_i) = \frac{1}{N}X(\omega_i)^2$$

where $N$ is the number of frequencies. The normalization of PSD is a Probability Density Function with integral equal to 1:

$$P_i = \frac{P(\omega_i)}{\sum P(\omega_i)}$$

The Power Spectral Entropy (PSE) is given by:

$$PSE = -\sum_{i=1}^{n} P_i \ln P_i$$

**Definition 2 (Wavelet Shannon Entropy (WSE)).** Given a signal $s$ and the coefficients of $s$ in an orthonormal basis $\{s_i\}$. Wavelet Shannon Entropy of $s$ is given by:

$$E(s) = -\sum_i s_i^2 \log(s_i^2)$$

The average computational time for calculating persistent entropy, spectral entropy and wavelet entropy for each motor is of 180 s on the following notebook: MacBook Air, 1.7 GHz i5, 8 GB RAM, Hard Disk SSD.

In Fig. 6 we plot the 46 values of Persistent entropy, Spectral entropy and Wavelet entropy. It is well evident that both Persistent entropy and Spectral entropy are able to discriminate the two classes of motors (“good” and “faulty”). We certified the statistical difference among entropies for “good motors” and “faulty motors” by using the Wilcoxon test [26], see Table 1. Persistent entropy and Spectral entropy passed the test. We interpreted these two entropies as two one-dimensional features and we used them separately for signal classification. The quality of these entropy-based classifiers is evaluated by ROC curves (see Fig. 7) [27,28]. The Area Under Curves (AUCs) as well as other information are reported $^2$ in Table 2. Persistent entropy outperforms Spectral entropy: the AUC of Persistent-entropy-based classifier is 93.87% with a standard error (S.E.) equal to 0.03749 with a Confidence Interval (C.I.), that is a sort of index of precision, in the range $[0.8962; 1.0000]$, while Spectral-entropy-based classifier is characterized by: $AUC = 82.02\%$, $S. E. = 0.06306$ and $C. I. \in [0.69664; 0.94383]$. The last column labeled Comment reports a qualitative description of the performance: “excellent test” means that 0.90 $\leq$ AUC $< 1$, while “good test” means 0.80 $\leq$ AUC $< 0.90$. Moreover, the analysis of the ratio between the number of correct assessments and the number of all assessments reveals the threshold “$\theta$” that maximizes the separation between the classes. For Persistent entropy $\theta = 0.7211$, while for Spectral entropy $\theta = 1.6502$. They can be used for classifying future motors.

Furthermore, in order to compare the classification obtained by the method introduced in this paper (based on persistent entropy) with the one discussed in [16], we compute the Jaccard coefficient. The Jaccard coefficient for both classifiers is $J = 96\%$. Our new classifier does not outperform the previous one but we have to remark that the method discussed in this paper is more suitable, since, first, the previous method needs to find the right parameter set for doing the time delay embedding of the signals. Second, from the embedded set of points, Vietoris-Rips complexes are built, which also depend on a parameter, the so-called proximity parameter $\epsilon$. Different $\epsilon$ returns a different complex and statistical methods would be required for finding the right $\epsilon$ for the data-set under analysis. However, the method that we have introduced in Section 3.1 is completely data-driven, it does not need any user-defined parameter and further analysis is not necessary for the interpretation of the result.

4. Discussion

In this paper we reported on the development of a new technique, based on persistent homology and information theory, for comparing discrete signals. The transformation of a signal into a filtered simplicial complex of dimension 1 allows to study its topology in terms of persistent homology. The persistence barcodes characterize the signals and they are used for calculating an entropy measure, the so-called persistent entropy. Persistent entropy is used as feature for comparing signals. One of the main contributions of this paper is the stability theorem for persistent entropy. This theorem gives the formal support for the comparison of the persistent entropy of two signals. The methodology presented in this paper is able to decide if two signals have the same shape even if one is shifted with respect to the other. More rigorously, given two signals ($s_1$, $s_2$) with the same shape but $s_2$ is shifted with respect to $s_1$ ($s_2 = s_1 + \text{nandory}_2 = y_1 + m$, where $m, n \in \mathbb{R}$), the persistent entropy for $s_1$ and $s_2$ is the same. The y-shifting increases the filter values ($F_2 = F_1 + m$) but the number of lines within a barcode and their lengths is completely preserved. We obtained numerical evidences that for the dataset under analysis, our methodology outperforms the Fourier-based entropy. The comparison is stated in terms of Area under ROC curves. The area under curves obtained for persistent-entropy-based classifier is $AUC = 93.87\%$, while for spectral entropy $AUC = 82.02\%$. In future investigations we will verify if our methodology is more computationally convenient with respect to the distances usually used for measuring similarities between two persistence barcodes (e.g., Wasserstein and Bottleneck). Persistent entropy seems to be an expressive feature that can be used for linking data analysis with formal methods. In fact, persistent entropy was used for characterizing the Persistent Entropy Automaton (PEA) [15]. Based on the stability theorem for persistent entropy, we intend to extend the bioshape language, that is a formal language for dealing with low dimensional geometrical structures [29], to a new language based on simplices, that are higher dimensional objects, and operational timed semantics [30].

Acknowledgements

We acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme (FP7) for Research of the European Commission, under the FP7 FET-Proactive Call 8 - DyMCs, Grant Agreement TOPDRIM, number FP7-ICT-318121 and MINECO, FEDER/UE under project number MTM2013-67072-P. We also thank the reviewers for their valuable comments and suggestions.

---

$^2$ http://www.fttw.org
$^4$ http://www.mathworks.com/matlabcentral/fileexchange/19950
References


