New Approaches in Ordinal Pattern Representations for Multivariate Time Series

Marisa Mohr, [1,2,*] **Florian Wilhelm,**[1] **Mattis Hartwig,**[2] **Ralf Möller,**[2] **Karsten Keller**[3]
[1] inovex GmbH, 76131 Karlsruhe, Germany
[2] Institute of Information Systems, University of Lübeck, 23562 Lübeck, Germany
[3] Institute of Mathematics, University of Lübeck, 23562 Lübeck, Germany
[*] Correspondence: mohr@ifis.uni-luebeck.de

Abstract

Many practical applications involve classification tasks on time series data, e.g., the diagnosis of cardiac insufficiency by evaluating the recordings of an electrocardiogram. Since most machine learning algorithms for classification are not capable of dealing with time series directly, mappings of time series to scalar values, also called representations, are applied before using these algorithms. Finding efficient mappings, which capture the characteristics of a time series is subject of the field of representation learning and especially valuable in cases of few data samples. Time series representations based on information theoretic entropies are a proven and well-established approach. Since this approach assumes a total ordering it is only directly applicable to univariate time series and thus rendering it difficult for many real-world applications dealing with multiple measurements at the same time. Some extensions were established which also cope with multivariate time series data, but none of the existing approaches take into account potential correlations between the movement of the variables. In this paper we propose two new approaches, considering the correlation between multiple variables, which outperform state-of-the-art algorithms on real-world data sets.

1 Introduction

Time series data is part of many real-world applications, e.g., in weather prediction, stock markets, energy production, medical recordings, sales and websites activity, political or sociological factors. Classification of time series is a challenging and important subject, e.g., to diagnose cardiac insufficiency by evaluating the recordings of an electrocardiogram (ECG). Classical machine learning models for classification such as k-nearest neighbor, support vector machines or random forest, can't process time series directly. It is necessary to extract scalar-valued representations (or features) from time series before using these algorithms.

The choice of suitable representations strongly influences the quality of a model and its predictions. Ideally, representations are chosen without domain knowledge of specialists and yet contain as much information as possible. An approach for automatic extraction of representations is given by the application of deep neural networks, which learn appropriate representations of time series within their hidden layers (Franceschi, Dieuleveut, and Jaggi 2019). Deep neural networks, however, require many data samples for model training, which are rarely available in real-world tasks.

In general, the goal is to find efficient mappings from a time series to a scalar representation that capture as many characteristics about the time series as possible. Information theoretic entropies are promising through an encoding that preserves information content (Amigó 2010). The concept of Permutation Entropy (PE) has already been successfully used in time series analysis and classification on many univariate real-world data sets (Antonelli, Meschino, and Ballarin 2019; J. Weck et al. 2014; Xue et al. 2019). PE uses total ordering of data points in a univariate time series to encode the ups and downs of neighboring points.

The concept of PE in its original form is not able to handle multivariate time series, because of the inability to determine a total order between vector-valued time points. Nevertheless, in real-world applications we often have to deal with high-dimensional multivariate time series. For example, medical measurements stored as ECG data are usually not determined from a single electrode, but from multiple electrodes. There are extensions of PE which also cope with multivariate time series data, but none of the existing approaches take into account potential correlations between the simultaneous movement of the variables over the time.

This paper introduces two new approaches considering simultaneous movement and correlations of variables. Section 2 introduces the ordinal pattern formalism and a definition of PE. Section 3 presents related work. Section 4 details our new approaches. To prove their contribution, we show in Section 5, that our approaches outperform in classification tasks on many data sets out of UEA Multivariate Time Series Classification (MTSC) archive (Bagnall et al. 2018). In the last Section, we discuss limitations, as well as further research.

2 Preliminaries

For using entropies as representations, time series observations are encoded as sequences of symbolic abstractions. As far as current research is concerned, there are two general

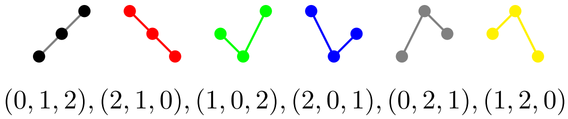

Figure 1: All possible ordinal patterns of order $d = 3$.

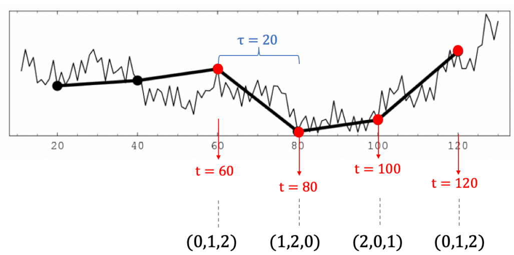

Figure 2: Ordinal pattern determination of order $d = 3$ and time delay $\tau = 20$ in a univariate time series.

approaches of symbolization. Classical symbolization approaches use threshold values and data range partitioning for symbol assignment such as the well-know SAX representation introduced by Chiu, Keogh, and Lonardi (2003). The ordinal pattern symbolization approach, describing the up and down, is based on Bandt and Pompe (2002). The formalism and the advantages of this symbolization scheme are introduced in the following.

2.1 Ordinal Pattern Symbolization

Ordinal patterns describe the total order between two or more neighbors, coded by permutations.

Definition 1 (Univariate Ordinal Pattern)**.** A vector $(x_1, ..., x_d) \in \mathbb{R}^d$ has *ordinal pattern* $(r_1, ..., r_d) \in \mathbb{N}^d$ of order $d \in \mathbb{N}$ if $x_{r_1} \geq ... \geq x_{r_d}$ and $r_{l-1} > r_l$ in the case $x_{r_{l-1}} = x_{r_l}$.

Note that the equality of two values within a pattern is not allowed. In this case, for example, the newer value is regarded as a smaller value. Figure 1 shows all possible ordinal patterns of order $d = 3$ of a vector (x_1, x_2, x_3). To symbolize a time series $(x_1, x_2, ..., x_T) \in \mathbb{R}^T$ each time point $t \in \{d, ..., T\}$ is assigned its ordinal pattern of order d. The order d is chosen much smaller than the length T of the time series to look at small windows in a time series and their distributions of up and down movements. To access the overarching trend delayed behavior is of interest. The time delay $\tau \in \mathbb{N}$ is the delay between successive points in the symbol sequences. Different delays show different details of the structure of a time series. Figure 2 visualizes the ordinal pattern determination of order $d = 3$ and time delay $\tau = 20$ of four different time points in a univariate time series.

The ordinal approach has notable advantages in application. First of all, the method is conceptionally simple. Second, it is not necessary to have previous knowledge about the data range or type of time series. Third, the ordinal approach supports robust and fast implementations (Keller et al. 2017; Piek, Stolz, and Keller 2019). Fourth, it allows an easier estimation of a good symbolization scheme compared to the classical symbolization approaches (Keller, Maksymenko, and Stolz 2015; Stolz and Keller 2017).

2.2 Ordinal Pattern Distributions

Not the ordinal patterns themselves, but their distribution in different parts of a *univariate time series* $(x_t)_{t=1}^T$ are of interest. Thus, each pattern is identified with exactly one of the *ordinal pattern symbols* $j = 1, 2, ..., d!$. Using the distribution of symbols, the entropy of ordinal pattern symbols is calculated as in the following Definition.

Definition 2 (Permutation Entropy (PE))**.** The *permutation entropy* of order $d \in \mathbb{N}$ and delay $\tau \in \mathbb{N}$ of a univariate time series $(x_t)_{t=1}^T$, $T \in \mathbb{N}$ is defined by

$$\mathrm{PE}(d, \tau) = -\sum_{j}^{d!} p_j^{\tau,d} \ln p_j^{\tau,d}, \quad (1)$$

where

$$p_j^{\tau,d} = \frac{\#\{t | (x_{t-(d-1)\tau}, ..., x_{t-\tau}, x_t) \text{ has pattern } j\}}{T - (d-1)\tau} \quad (2)$$

is the relative frequency of ordinal pattern j in the time series.

Depending on the area of research, entropy is a measure for quantifying inhomogeneity, impurity, complexity and uncertainty or unpredictability. In this paper, we use PE as a representation (also known as feature) in a learning model that measures the complexity of a time series, which is measured in the (in)regularity of ordinal patterns occurring. For time series with maximum random ordinal pattern symbols (resulting in a uniform pattern distribution due to uniqueness), PE is $ln(d!)$. For a time series with regular pattern, e.g., in case of monotony, PE is equal to zero (Amigó 2010).

A *multivariate time series* $((x_t^i)_{i=1}^m)_{t=1}^T$ has more than one time-dependent variable. Each variable x^i for $i \in 1, ..., m$ depends not only on its past values but also has some dependency on other variables. Considering two time points $(x_t^i)_{i=1}^m$ and $(x_{t+1}^i)_{i=1}^m$ with m variables, it is not possible to establish a total order between them. A total order is only possible if $x_t^i > x_{t+1}^i$ or $x_t^i < x_{t+1}^i$ for all $i \in 1, ..., m$. Therefore, there is no trivial generalization of the PE algorithm to the multivariate case.

3 Related Work

In order to determine ordinal patterns on multivariate time series, two cases have been discussed in literature so far:

1. The multivariate time series is *projected into ordinal space directly* by either
   a. determining univariate ordinal pattern *between values in time* of a single variable i (row marked in blue in Figure 3), and then averaging over all m variables, or
   b. determining univariate ordinal pattern *between values in time* of a single variable i (box marked in green in Figure 3), storing all m pattern at a fixed time point t in one matrix, or

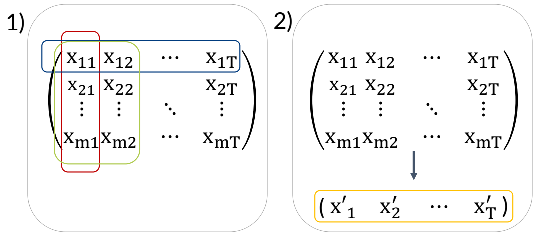

Figure 3: Two possibilities of univariate ordinal pattern determination in a multivariate time series. Case 1 shows ignoring variables (blue) and time (red) dependencies respectively. Our first proposed approach considers both (green). Case 2 shows a previous dimensionality reduction (yellow).

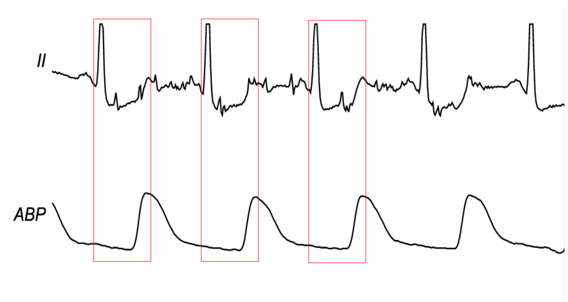

Figure 4: Medical time series (ECG signal II and ABP) from a patient of the MIMIC III waveform database with identification id 3900006_0029 published by Johnson et al. (2016).

c. determining univariate ordinal pattern *between values of all variables* at a single time point t (column marked in red in Figure 3), and then averaging over all T time points.

2. The multivariate time series is *reduced to a single-dimensional projection* first, and then transformed into ordinal space.

Consequently, Definition 2 can be used for PE calculation in all cases. The first approach 1(a) is implemented by Keller and Lauffer (1999) as Pooled Permutation Entropy and hereafter referred to as PE_{pool}. Morabito et al. (2012) expand the concept of PE_{pool} by taking into account variations over multiple scales, named Multivariate Multi-Scale Permutation Entropy (MMSPE). Some theoretical basis for 1(b) is set by Keller; Antoniouk, Keller, and Maksymenko (2012; 2013). A variant of approach 1(c) is implemented by He, Sun, and Wang (2016) as Multivariate Permutation Entropy (MVPE). The second case is implemented by Rayan, Mohammad, and Ali (2019). They propose to reduce the number of variables m variable first by applying different distance measures between all points of the time series and a reference point before calculating PE on the reduced univariate time series. Depending on the distance measure and reference point we denote the calculated PE-variants as PE_{eucl} (Euclidian distance with reference point $(x_0^i)_{i=1}^m$)), PE_{manh} (Manhattan distance with reference point $(x_0^i)_{i=1}^m$)), and PE_{norm} (Euclidian distance with reference point $\mathbf{0}$). Note that the length T of the time series in variable-dimensions must be the same in order to perform a dimensionality reduction. The well-known Matrix Profile by Yeh et al. (2016) pursues a different goal and is therefore not considered in this paper.

4 Ordinal Pattern Representations Considering Correlations of Variables

The *simultaneous movement pattern of several variables over time* is important information, that should be encoded in a mathematical representation. For example, each of ECG and Atrial Blood Pressure (ABP) signals contain information of cardiac status, which can be used for diagnosis of diseases. The depolarization of ventricles and contraction of the large ventricular muscles of the human heart (so-called QRS complex) can be observed during an ECG signal by the highest rash, the central and most visually obvious part of the tracing as it is shown in the signal at top in Figure 4. Shortly after the electrical activity the blood pressure rises as it is shown in the second signal in Figure 4. Both movements of variables depend on each other and should be considered together (red boxes in Figure 4), rather than separately.

In the following, we propose two new approaches taking into account the correlation of two or more variables. In Section 4.1, we introduce our first approach adapting the first case in Section 3 by natural extension of dimensions to include both, variable and time. In Section 4.2, we present a second approach following the second case in Section 3 by an improvement of dimensional reduction via Principal Component Analysis.

4.1 Permutation Entropy Based on Multivariate Ordinal Pattern

The intuitive idea is to store the univariate ordinal patterns of all variables at a time point t together into one multivariate pattern.

Definition 3 (Multivariate Ordinal Pattern)**.** A matrix $(x_1, ..., x_d) \in \mathbb{R}^{m \times d}$ has *multivariate ordinal pattern*

$$\begin{pmatrix} r_{11} & \cdots & r_{1d} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{md} \end{pmatrix} \in \mathbb{N}^{m \times d} \tag{3}$$

of order $d \in \mathbb{N}$ if $x_{r_{i1}} \geq ... \geq x_{r_{id}}$ for all $i = 1, ..., m$ and $r_{il-1} > r_{il}$ in the case $x_{r_{il-1}} = x_{r_{il}}$.

Figure 5 shows all possible multivariate ordinal patterns of order $d = 2$ and number of variables $m = 2$.

With the natural extension of Definition 1, which leads to Definition 3, it is possible to apply the PE algorithm in Definition 2 in its original form to multivariate time series. We denote PE_{multi} as PE that have been calculated on multivariate ordinal patterns. Obviously, this approach has the disadvantage that the number of possible patterns $d!$ increases exponentially with the number of variables m, in summary $(d!)^m$. In this setting, we quickly end up with an under-sampling and a uniform distribution of patterns, resulting in

$$\begin{pmatrix}1 & 0\\1 & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}, \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\0 & 1\end{pmatrix}$$

Figure 5: All possible multivariate ordinal patterns of order $d = 2$ and variable-dimension $m = 2$.

maximum complexity. Nevertheless, with small order d and sufficiently large length T of the time series, more information can be obtained than with its averaged variant $\mathrm{PE}_{\mathrm{pool}}$. The following evaluation confirms this.

4.2 Permutation Entropy Based on Principal Component Analysis

The disadvantages of the approach in the previous Section limit its applicability on real-world data sets. In order to minimize the combinatorial possibilities of ordinal patterns, we consider the univariate ordinal pattern case. The idea is to transform a multivariate time series $((x_t^i)_{i=1}^m)_{t=1}^T$ into a univariate time series $((x_t')_{t=1}^T$ and then calculate the PE from Definition 2 on the reduced time series as usual. To order time points based on a single-dimensional values, Rayan, Mohammad, and Ali (2019) propose to use the Euclidian or Manhattan distance between all points of the time series and a reference point for reduction. The calculation of distances can be interpreted as the strength of a signal or time series, on the basis of which PE is then calculated. This approach does not take into account correlations between variables.

The aim, however, should be to sustain as much information of variables correlations as possible to understand their simultaneous movement. Principal Component Analysis (PCA) is a well-known method converting a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables by an orthogonal transformation. The total variance is an information that describes one kind of characteristics of the time series (Hastie, Tibshirani, and Friedman 2009). For m-dimensional data, there are basically m basis vectors that are orthogonal. The variance of data points along each basis vector is the total variance of the data. If the first $r < m$ basis vectors cover a sufficiently large percentage of the total variance, then the major components represented by the new r basis vectors will be sufficient for the information content of the data. Keeping only the first r principal components gives the truncated transformation $\mathbf{X'}_r = \mathbf{X}\mathbf{W}_r$, where $W \in \mathbb{N}^{r \times r}$ is a matrix of weights whose columns are the eigenvectors of $X^T X$ sorted in descending order of the r highest corresponding eigenvalues.

For applying PCA, we assume that the time series $((x_t^i)_{i=1}^m)_{t=1}^T$ is stationary for all $i = 1, .., m$. An appropriate algorithm to reduce the data by PCA can be found in all popular statistical textbooks, e.g. in Hastie, Tibshirani, and Friedman (2009). We denote $\mathrm{PE}_{\mathrm{PCA}}$ as PE that have been calculated on the single-dimensional projection by PCA. We have $r = 1$ for reducing a multivariate into a univariate time series. The reduction from high number of variables m to only one dimension can lead to a high loss of information of variance in the data, provided that the first main component explains very little variance of the data. In order to show that the addition of further main components does not lead to better results, we denote $\mathrm{PE}_{\mathrm{PCA2}}$ as the calculation of $\mathrm{PE}_{\mathrm{multi}}$ on the first two principal components. A possible explanation for this could be the orthogonality of the main components.

5 Experimental Results

We have conducted several experiments to investigate the relevance of the different ordinal pattern representations for multivariate time series. We consider a representation as being *good*, if it is flexible and realiable applicable on different real-world data sets. To show the flexibility, we compare it on 25 different real-world use cases. To show the reliability, we compare the accuracy of seven different PE-variants. Higher accuracy means that the representation is able to identify the underlying explanatory factors better than other representations and discriminatory properties can be identified as useful inputs for supervised predictors. (Bengio, Courville, and Vincent 2012). We challenge our ordinal pattern representations in classification tasks on the UEA MTSC archive, a collection of different time series data sets from many real-world cases, released in October 2018 by Bagnall et al. (2018). The archive consists of 30 data sets with a wide range of series lengths, dimensions and cases from human activity recognition, motion classification, ECG classification, electroencephalography (EEG) and magnetoencephalography (MEG) classification to audio spectra classification and many others. In the following setting 5 of 30 data sets of the archive could not be used without compromising comparability due to different length T of the time series in its m variables.

5.1 Deep Dive: AtrialFibrillation Classification

Before we start with a general evaluation of different data sets, we give a detailed insight in one specific data set out of the 30 for a better understanding. The AtrialFibrillation data set contains two-channel ECG recordings for predicting spontaneous termination of atrial fibrillation (AF). The class labels are: t, s and n. Class t contains data, where the AF terminates at the latest within one second after the recording ending. Class s is described as an AF that self terminates at least one minute after the recording ending. In Class n, the AF does not terminate for at least one hour after the recording of the data. An example of the recordings for each class is shown in Figure 6. More details are in Moody (2004). Below we examine the separability of the three classes based on the different PE-variants, $\mathrm{PE}_{\mathrm{pool}}$, $\mathrm{PE}_{\mathrm{multi}}$ and $\mathrm{PE}_{\mathrm{PCA}}$. The variance ratio for the first main component is greater than 74.17% for every $l = 1, .., n_{\mathrm{train}}$, where n_{train} is the number of train samples. So a dimensionality reduction via PCA can be applied without losing too much information. Figure 7 shows boxplots for the calculated values of three different PE-variants for each class. While $\mathrm{PE}_{\mathrm{pool}}$ does not allow any separability of the classes, $\mathrm{PE}_{\mathrm{multi}}$ allows classes n and t to be separated relatively well. An even better separation of classes s and t is achieved by $\mathrm{PE}_{\mathrm{PCA}}$.

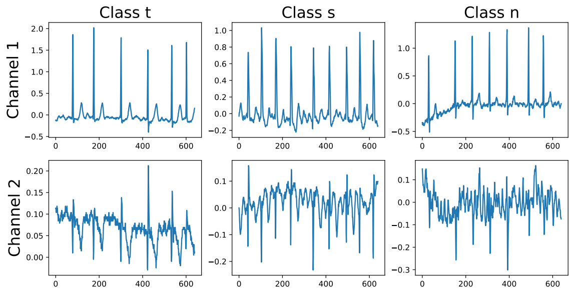

Figure 6: Example of 2-channel recordings for all three classes.

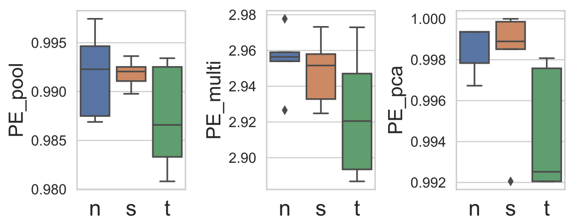

Figure 7: Boxplots for the values of three different PE-variants of order $d = 2$ and delay $\tau = 1$ for classes n (blue), s (orange) and t (green).

5.2 A Comparison of PE-variants on the UEA Data Sets

We perform a classification on the UEA MTSC data sets to compare the ability of separation through seven ordinal pattern representations $\mathrm{PE}_{\mathrm{pool}}$, $\mathrm{PE}_{\mathrm{multi}}$, $\mathrm{PE}_{\mathrm{eucl}}$, $\mathrm{PE}_{\mathrm{manh}}$, $\mathrm{PE}_{\mathrm{norm}}$, $\mathrm{PE}_{\mathrm{PCA}}$ and $\mathrm{PE}_{\mathrm{PCA2}}$ on arbitrary data sets. The initial benchmarking on the UEA MTSC data sets by Bagnall et al. (2018) is with the standard 1-nearest neighbor (NN) classifier with three different distance functions. To make a certain comparability we also use the 1-NN classifier. As model input, all seven PE-variants were used individually for the evaluation. Finding an optimal order d and time delay τ is a challenging problem in research (Riedl, Müller, and Wessel 2013; Myers and Khasawneh 2019). For simplicity, we have done an extensive hyperparameter optimization.

Table 1 lists the accuracy of the performed 1-NN classification on all data sets with train size n_{train}, number of variables m, length T of the time series and number of classes C. The highest accuracy scores per data set are bold. Besides, the names of data sets whose benchmark accuracy we outperform are italics.

$\mathrm{PE}_{\mathrm{multi}}$ outperforms on two data sets with low number m of variables and high length T of the time series, which confirms that the representation is successful in this special case, especially more successful than its averaged variant $\mathrm{PE}_{\mathrm{pool}}$. $\mathrm{PE}_{\mathrm{PCA}}$ outperforms the other state-of-the-art PE variants in most cases. The results show that a dimensionality reduction by decorrelation of variables is more successful than by different distance measures. The accuracy scores of $\mathrm{PE}_{\mathrm{PCA2}}$ confirm that adding more major components does not produce better results than $\mathrm{PE}_{\mathrm{PCA}}$.

6 Conclusion and Future Work

Throughout this paper, we have discussed ordinal pattern representations for multivariate time series both from the viewpoint of their theoretical foundation and their application. We have shown that our approaches, especially $\mathrm{PE}_{\mathrm{PCA}}$, outperform state-of-the-art algorithms on many real-world data sets. $\mathrm{PE}_{\mathrm{multi}}$ is a valuable representation in case of small number m of variables and high length T of the time series. Due to an improvement in prediction results and easy handling, the integration into toolboxes for representation learning of multivariate time series is indispensable.

Furthermore, we will study, if an optimized method for dealing with PCA in case of high-dimension, low-sample-size (HDLSS) data suggested by Yata and Aoshima (2012) can improve our results of $\mathrm{PE}_{\mathrm{PCA}}$. In addition, it may be worth taking a closer look at the individual PE of all m main components to understand how movements on the decorrelated variables of the multivariate time series change.

The advantages of ordinal patterns for analysis and prediction of time series is well known in research. The approach presented here has the disadvantage that representation and prediction model are chosen independently. We currently work on adapting the ordinal pattern approach to automatic representation learning. Instead of first calculating PE and applying an ML model in a second step, this will allow to both learn the representation and use them to perform a specific problem *in one task*.

References

Amigó, J. 2010. *Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That*. Springer Series in Synergetics. Berlin Heidelberg: Springer-Verlag.

Antonelli, A.; Meschino, G.; and Ballarin, V. 2019. Mammographic Density Estimation Through Permutation Entropy. In Lhotska, L.; Sukupova, L.; Lacković, I.; and Ibbott, G. S., eds., *World Congress on Medical Physics and Biomedical Engineering 2018*, IFMBE Proceedings, 135–141. Singapore: Springer.

Antoniouk, A.; Keller, K.; and Maksymenko, S. 2013. Kolmogorov-Sinai entropy via separation properties of order-generated sigma-algebras. *Discrete and Continuous Dynamical Systems* 34(5):1793–1809.

Bagnall, A.; Dau, H. A.; Lines, J.; Flynn, M.; Large, J.; Bostrom, A.; Southam, P.; and Keogh, E. 2018. The UEA multivariate time series classification archive, 2018. *arXiv:1811.00075 [cs, stat]*.

Bandt, C., and Pompe, B. 2002. Permutation Entropy: A Natural Complexity Measure for Time Series. *Physical Review Letters* 88(17).

Bengio, Y.; Courville, A.; and Vincent, P. 2012. Representation Learning: A Review and New Perspectives. *arXiv:1206.5538 [cs]*.

Chiu, B.; Keogh, E.; and Lonardi, S. 2003. Probabilistic Discovery of Time Series Motifs. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, 493–498. New York, NY, USA: ACM.

Franceschi, J.-Y.; Dieuleveut, A.; and Jaggi, M. 2019. Unsupervised Scalable Representation Learning for Multivariate Time Series. *arXiv:1901.10738 [cs, stat]*.

Hastie, T.; Tibshirani, R.; and Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. New York: Springer-Verlag, 2 edition.

					1-NN based on						
Datset	n_{train}	m	T	C	PE_{pool}	PE_{multi}	PE_{eucl}	PE_{manh}	PE_{norm}	PE_{PCA}	PE_{PCA2}
ArticularyWordR.	275	9	144	25	**0.137**	0.073	0.116	0.11	0.11	0.127	0.064
AtrialFibr.	15	2	640	3	0.667	0.6	0.6	0.6	0.666	**0.8**	0.4
BasicMotions	40	6	100	4	0.625	0.525	0.525	0.45	0.475	**0.675**	0.45
DuckDuckGeese	50	1345	270	5	0.3	0.2	0.28	0.24	0.26	**0.36**	0.28
EigenWorms	128	6	17984	5	0.557	**0.611**	0.578	0.584	0.578	0.595	0.443
Epilepsy	137	3	206	4	0.5	0.478	0.507	0.48	0.485	**0.514**	0.384
ERing	30	4	65	6	**0.392**	0.351	0.385	0.374	0.351	0.381	0.274
EthanolCon.	261	3	1751	4	0.323	0.312	0.304	0.323	0.330	**0.331**	0.300
FaceDetection	5890	144	62	2	0.504	0.500	0.508	0.506	0.504	**0.512**	0.510
FingerMovem.	360	28	50	2	0.57	0.54	0.6	**0.62**	0.55	0.58	0.56
HandMovem.	160	10	400	4	0.419	0.419	0.392	0.392	**0.432**	0.364	0.405
Handwriting	150	3	152	26	0.072	0.06	**0.075**	0.069	0.074	0.071	0.051
Heartbeat	204	61	405	2	0.692	0.722	0.682	0.712	0.698	**0.731**	0.727
Libras	180	2	45	15	0.322	0.266	0.311	0.295	**0.35**	0.277	0.117
LSST	2459	6	36	14	0.238	0.212	0.201	0.236	0.198	0.260	**0.309**
MotorImagery	278	64	3000	2	0.52	0.56	0.54	0.49	0.55	**0.57**	0.56
NATOPS	180	24	51	6	0.239	0.2	0.25	**0.272**	**0.272**	0.266	0.2
PEMS-SF	267	963	144	7	0.671	0.173	0.563	0.647	0.524	**0.676**	0.266
PenDigits	7494	2	8	10	0.203	0.190	0.201	0.198	0.176	**0.229**	0.107
PhonemeSpectra	3315	11	217	39	0.060	0.052	0.049	0.045	0.060	**0.061**	0.034
RacketSports	151	6	30	4	0.348	0.296	0.355	0.349	0.368	**0.381**	0.362
SelfR.SCP1	268	6	896	2	0.580	0.556	0.659	**0.662**	0.607	0.611	0.539
SelfR.SCP2	200	7	1152	2	0.588	0.577	0.561	0.572	0.584	**0.6**	0.577
StandWalkJ.	12	4	2500	3	0.666	**0.733**	0.666	0.533	**0.733**	0.666	0.533
UWaveGest.	120	3	315	8	0.243	0.234	0.225	0.216	0.206	**0.244**	0.184

Table 1: Accuracy scores of multivariate PE variants on 25 UEA MTSC data sets.

He, S.; Sun, K.; and Wang, H. 2016. Multivariate permutation entropy and its application for complexity analysis of chaotic systems. *Physica A: Statistical Mechanics and its Applications* 461:812–823.

J. Weck, P.; Schaffner, D.; Brown, M.; and Wicks, R. 2014. Permutation Entropy and Statistical Complexity Analysis of Turbulence in Laboratory Plasmas and the Solar Wind. *arXiv*.

Johnson, A. E. W.; Pollard, T. J.; Shen, L.; Lehman, L.-w. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data* 3:160035.

Keller, K., and Lauffer, H. 1999. Symbolic analysis of high-dimensional time series. In *Int. J. Bifurcation Chaos*, 2657–2668.

Keller, K.; Mangold, T.; Stolz, I.; and Werner, J. 2017. Permutation Entropy: New Ideas and Challenges. *Entropy* 19(3):134.

Keller, K.; Maksymenko, S.; and Stolz, I. 2015. Entropy determination based on the ordinal structure of a dynamical system. *Discrete and Continuous Dynamical Systems - Series B* 20(10):3507–3524.

Keller, K. 2012. Permutations and the Kolmogorov-Sinai entropy. *Discrete & Continuous Dynamical Systems - A* 32(3):891.

Moody, G. 2004. Spontaneous termination of atrial fibrillation: A challenge from physionet and computers in cardiology 2004. In *Computers in Cardiology, 2004*, 101–104.

Morabito, F. C.; Labate, D.; La Foresta, F.; Bramanti, A.; Morabito, G.; and Palamara, I. 2012. Multivariate Multi-Scale Permutation Entropy for Complexity Analysis of Alzheimer's Disease EEG. *Entropy* 14(7):1186–1202.

Myers, A., and Khasawneh, F. 2019. On the Automatic Parameter Selection for Permutation Entropy. *arXiv:1905.06443 [nlin, physics:physics]*.

Piek, A. B.; Stolz, I.; and Keller, K. 2019. Algorithmics, Possibilities and Limits of Ordinal Pattern Based Entropies. *Entropy* 21(6):547.

Rayan, Y.; Mohammad, Y.; and Ali, S. A. 2019. Multidimensional Permutation Entropy for Constrained Motif Discovery. In Nguyen, N. T.; Gaol, F. L.; Hong, T.-P.; and Trawiński, B., eds., *Intelligent Information and Database Systems*, Lecture Notes in Computer Science, 231–243. Cham: Springer International Publishing.

Riedl, M.; Müller, A.; and Wessel, N. 2013. Practical considerations of permutation entropy: A tutorial review. *The European Physical Journal Special Topics* 222.

Stolz, I., and Keller, K. 2017. A General Symbolic Approach to Kolmogorov-Sinai Entropy. *Entropy* 19(12):675.

Xue, X.; Li, C.; Cao, S.; Sun, J.; and Liu, L. 2019. Fault Diagnosis of Rolling Element Bearings with a Two-Step Scheme Based on Permutation Entropy and Random Forests. *Entropy* 21(1):96.

Yata, K., and Aoshima, M. 2012. Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *Journal of Multivariate Analysis* 105(1):193–215.

Yeh, C.-C. M.; Zhu, Y.; Ulanova, L.; Begum, N.; Ding, Y.; Dau, H. A.; Silva, D. F.; Mueen, A.; and Keogh, E. 2016. Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 1317–1322.