An Experimental Evaluation of Time Series Classification Using Various Distance Measures

Tomasz Górecki and Paweł Piasecki

Abstract In recent years a vast number of distance measures for time series classification has been proposed. Obviously, the definition of a distance measure is crucial to further data mining tasks, thus there is a need to decide which measure should we choose for a particular dataset. The objective of this study is to provide a comprehensive comparison of 26 distance measures enriched with extensive statistical analysis. We compare different kinds of distance measures: shape-based, edit-based, feature-based and structure-based. Experimental results carried out on 34 benchmark datasets from UCR Time Series Classification Archive are provided. We use an one nearest neighbour (1NN) classifier to compare the efficiency of the examined measures. Computation times were taken into consideration as well.

Tomasz Górecki · Paweł Piasecki
Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland
✉ tomasz.gorecki@amu.edu.pl
✉ pawel.piasecki@amu.edu.pl

Archives of Data Science, Series A
(Online First) DOI: 10.5445/KSP/1000087327/07
KIT Scientific Publishing ISSN 2363-9881
Vol. 5, No. 1, 2018

1 Introduction

Nowadays, we are collecting more and more data of different types. Typically, we differentiate time series problems from other data analysis tasks, because the attributes are ordered and we may look for a discriminatory feature that depends on the ordering (Bagnall et al, 2017). In the last 20 years interest in the area of time series has soared and many tasks have been deeply investigated, such as:

- classification (Bagnall et al, 2017),
- clustering (Keogh and Lin, 2005),
- indexing (Keogh, 2006),
- prediction (Weigend and Gershenfeld, 1994),
- anomaly detection (Weiss, 2004),
- motif discovery (Lin et al, 2004) and more.

In our opinion, there is a problem that appears throughout almost all of these topics: How to compare two given time series in the most appropriate way?

The problem of pairwise similarity of time series is based on the underlying distance measure (which are not necessarily metrics or even dissimilarity measures). To the best of our knowledge, about 40 distance measures have already been proposed in the literature. Some of them are based on certain features of data, while others use predictions, underlying models or some transformations. Such a variety may be confusing and makes it hard to find the most appropriate measure, especially for application-oriented scientists. Available research include only 2 papers providing a partial comparison of selected distance measures.

Wang et al (2013) provide an extensive comparison of 9 different similarity measures and their 4 variants, which was applied to 38 time series datasets from UCR archive (Dau et al, 2018). The authors of the paper conclude, that they did not find any measure, that is "universally better" for all datasets – some of them are better than the rest, while being worse on other datasets. However, Dynamic Time Warping (DTW; Berndt and Clifford (1994)) – slightly ahead of some edit based measures: Longest Common Subsequence distance (LCSS), Edit Distance for Real Sequences (EDR) and Edit Distance with Real Penalty (ERP) – seems to be superior to others. And it is in line with the widespread opinion – that DTW is not always the best but in general hard to beat (Xi et al, 2006; Spiegel et al, 2014). On the other hand, the study points out that Euclidean distance remains a quick and efficient way of measuring distances between time series. Especially, when the training set increases, the accuracy of elastic measures converges to that of Euclidean distance.

Serrà et al (2012) compare 7 similarity measures on 45 datasets from UCR archive. The authors of the paper suggest that, in the set of investigated distances, there is a group of measures with no statistically significant differences: DTW, EDR and MJC. Another finding is that the Time Warp Edit Distance (TWED) measure seems to consistently outperform all the considered distances. Euclidean distance is said to perform statistically worse than TWED, DTW, EDR and Minimum Jump Cost distance (MJC), and even its performance on large datasets was “not impressive”. What is more, an interesting remark is made about various post-processing steps that may increase classification accuracy:

– the complexity-invariant correction (Batista et al, 2011),
– the hubness correction for time series classification (Radovanović et al, 2010) and
– unsupervised clustering algorithms to prune nearest neighbor candidates (Serrà et al, 2012).

Despite giving interesting results, both studies take into account only some distance measures, while in the meantime, there are about 40 of them available. As it is computationally expensive, in this paper we compare 26 of them, but we plan to continue our experiment in the nearest future. Our contribution is to give an extensive comparison, supported by deep statistical analysis. We would like to create a benchmark study, that could be used not only by researchers from different application fields, but as well by authors of new distance measures, to assess their effectiveness. We are going to give only basic descriptions of the used similarity measures, provided along with some reference, as our intention is not to develop distance measures themselves, but rather to compare their efficacy. In out paper we focused on classification, however we should be aware that there are a lot of other tasks which can be done with time series, e.g.

– clustering,
– segmentation,
– registration,

– prediction,
– anomaly detection or
– change point detection,

where distance measures may be used in one way or the other.

2 Distances' Classification and Description

To the best of our knowledge there exist about 40 distance measures, thus there is a strong need to classify them. Montero and Vilar (2014) proposed to group measures into four categories: model free measures, model-based measures, complexity-based measures and prediction-based measures. Wang et al (2013) in their research named four groups of distance measures:

a) lock-step measures,
b) elastic measures,
c) threshold-based measures and
d) pattern-based measures.

In our opinion the most universal categorization, covering almost all distances, has been proposed by Esling and Agon (2012):

– shape-based measures,
– edit-based measures,
– feature-based measures and
– structure-based measures.

We are going to follow the last classification. In this following subsections we give a brief description of the 26 distance measures compared in this paper. For the interested reader, we shortly list distances that are already known, while for the moment, we did not include them into our research due to computational / technical reasons:

– Derivative Dynamic Time Warping (DDTW; Keogh and Pazzani (2001)),
– Parametric Derivative Dynamic Time Warping (DD_{DTW}; Górecki and Łuczak (2013)),

- dissimilarity measure based on a combination of DTW and LCSS distances (Górecki, 2018),
- LB-Keogh for DTW (Keogh and Ratanamahatana (2005)),
- Edit Distance on Real Sequence (EDR; Chen et al (2005)),
- Edit Distance with a Real Penalty (ERP; Chen and Ng (2004)),
- Time Warp Edit Distance (TWED; Marteau (2009)),
- prediction-based (Vilar et al, 2010),
- Dissim distance (Frentzos et al, 2007),
- Maharaj distance (Maharaj, 2000),
- Cepstral-based distance (Kalpakis et al, 2001),
- Shape-based Pattern Detection distance (SpADe; Chen et al (2007)),
- Global Alignment Kernels (GAK, Cuturi (2011)).

2.1 Shape-Based Distance Measures

This group of distance measures compare the overall shape of series, focusing mostly on the raw values.

L_p Distances

L_p distances are directly derived from L_p norms. They are widely used mainly thanks to their simplicity and ease of computation (Yi and Faloutsos, 2000). However, their drawbacks are: poor performance (Antunes and Oliveira, 2001), measuring only time series of equal length and being highly influenced by outliers, noise, scaling or warping.

Table 1: L_p distances.

Distance	p	Formula
Manhattan (MAN)	$p = 1$	$\sum_{i=1}^{T} \lvert x_i - y_i \rvert$
Euclidean (ED)	$p = 2$	$\sqrt{\sum_{i=1}^{T} (x_i - y_i)^2}$
Minkowski (MIN)	$1 \leq p < \infty$	$\sqrt[p]{\sum_{i=1}^{T} \lvert x_i - y_i \rvert^p}$
Infinite norm (INF)	$p = \infty$	$\max_{i=1,\ldots,T} \lvert x_i - y_i \rvert$

Note: An overview of the used abbreviations can be found in the appendix on page 19.

Given two time series $\mathbf{X}_T = (x_1, x_2, ..., x_T)$, $\mathbf{Y}_T = (y_1, y_2, ..., y_T)$ we compute them with formulas from Table 1. For the Minkowski distance we made an arbitrary choice of $p = 3$.

Short Time Series Distance (STS)

The Short Time Series distance was proposed by Möller-Levet et al (2003). The main idea of STS is to adapt measurement to irregularly sampled series. It is given by

$$d_{\mathrm{STS}}(\mathbf{X}_T, \mathbf{Y}_T) = \sqrt{\sum_{i=1}^{T-1} \left(\frac{y_{i+1} - y_i}{t_{i+1} - t_i} - \frac{x_{i+1} - x_i}{t'_{i+1} - t'_i} \right)}, \tag{1}$$

where t and t' are the temporal indexes of series $\mathbf{X}_T$ and $\mathbf{Y}_T$ respectively.

Dynamic Time Warping Distance (DTW)

DTW (Berndt and Clifford, 1994) is one of the most popular distance measures due to its ability to deal with warping of the time axis. The objective of DTW is to find the optimal alignment between two series by looking for the shortest warping path in a distance matrix.

As one of the main drawback of DTW is long computation time (Aßfalg et al, 2008; Papadopoulos, 2008), several lower bounding and temporal constraints techniques have been proposed. In Section 4 we denote DTW with Sakoe-Chiba Band as "DTWc" and we use the window size as in Dau et al (2018). For more details about DTW we refer to (Bagnall et al, 2017; Keogh and Ratanamahatana, 2005; Mori et al, 2016).

Complexity-Invariant Dissimilarity Measure

Many dissimilarity measures tend to put time series with high complexity level further apart than simple ones (Batista et al, 2011). In order to fix this distortion, a correction factor has been proposed by Batista et al (2011). A general complexity-invariant dissimilarity measure (CID) is defined as follows

$$d_{\mathrm{CID}}(\mathbf{X}_T, \mathbf{Y}_T) = CF(\mathbf{X}_T, \mathbf{Y}_T) \cdot d(\mathbf{X}_T, \mathbf{Y}_T), \tag{2}$$

where $d(\mathbf{X}_T, \mathbf{Y}_T)$ is a distance which may be adjusted (in our experiment we use Euclidean distance), $\mathrm{CF}(\mathbf{X}_T, \mathbf{Y}_T)$ is a complexity correction factor defined as

$$\mathrm{CF}(\mathbf{X}_T, \mathbf{Y}_T) = \frac{\max\{\mathrm{CE}(\mathbf{X}_T), \mathrm{CE}(\mathbf{Y}_T)\}}{\min\{\mathrm{CE}(\mathbf{X}_T), \mathrm{CE}(\mathbf{Y}_T)\}}, \tag{3}$$

where $\mathrm{CE}(\mathbf{X}_T)$ stands for a complexity estimator of $\mathbf{X}_T$. From (2), we can observe, that when the complexity of both time series is equal, we get $d_{\mathrm{CID}}(\mathbf{X}_T, \mathbf{Y}_T) = d(\mathbf{X}_T, \mathbf{Y}_T)$ and from (3) that an increase of complexity difference results in increase of distance between time series. As a complexity estimator Batista et al (2011) proposed

$$\mathrm{CE}(\mathbf{X}_T) = \sqrt{\sum_{t=1}^{T-1} (X_t - X_{t+1})^2}. \tag{4}$$

2.2 Edit-Based Distance Measures

Edit-based distances use the minimal number of operation necessary to transform one series into another.

Longest Common Subsequence Distance (LCSS)

The LCSS distance was proposed by Vlachos et al (2002) and measures the similarity between time series in terms of the longest common subsequence, with addition that gaps and unmatched regions are permitted. LCSS is robust to noise and we expect that it should be more accurate than DTW in the presence of outliers and noise. We set δ parameter to 100 % and as ε parameter we used a value equal to the smallest standard deviation between the two sequences that were examined at any time (Górecki, 2018).

2.3 Feature-Based Distances

These distances looks at some aspect of the time series by extracting certain feature. Then, based on it, a similarity measure is calculated.

Distance Based on Pearson's Correlation

Based on Pearson's correlation coefficient, we can define two distance measures (Golay et al, 2005):

$$d_{\mathrm{PC1}}(\mathbf{X}_T, \mathbf{Y}_T) = \left(\frac{1 - PC}{1 + PC}\right)^{\beta}, \tag{5}$$

$$d_{\mathrm{PC2}}(\mathbf{X}_T, \mathbf{Y}_T) = 2\,(1 - PC), \tag{6}$$

where PC denotes Pearson's correlation coefficient and β is a positive parameter. When the parameter is specified, we use the d_{PC1}, in the other case d_{PC2}.

Distance Based on the Cross-Correlation

Based on cross-correlation, Warren Liao (2005) defined the following distance measure

$$d_{\text{CC}}(\mathbf{X}_T, \mathbf{Y}_T) = \sqrt{\frac{(1 - CC_0(X,Y))}{\sum_{k=1}^{\max} CC_k(X,Y)}}, \tag{7}$$

where $\text{CC}_k(X,Y)$ is the cross correlation between two series at lag k. By default, we set the maximum lag to $T-1$.

Autocorrelation-Base and Partial Autocorrelation-Based Distances

Let $\hat{\boldsymbol{\rho}}_{X_T} = (\hat{\rho}_{1,X_T}, \ldots, \hat{\rho}_{L,X_T})^T, \hat{\boldsymbol{\rho}}_{Y_T} = (\hat{\rho}_{1,Y_T}, \ldots, \hat{\rho}_{L,Y_T})^T$ be the estimated autocorrelation vectors of $\mathbf{X}_T, \mathbf{Y}_T$ (respectively), for some L such that $\hat{\rho}_{i,X_T}, \hat{\rho}_{i,Y_T} \approx 0$ for $i > L$. Peña and Galeano (2000) proposed the following distance:

$$d_{\text{ACF}}(\mathbf{X}_T, \mathbf{Y}_T) = \sqrt{(\hat{\boldsymbol{\rho}}_{X_T} - \hat{\rho}_{Y_T})^T \boldsymbol{\Omega} (\hat{\boldsymbol{\rho}}_{X_T} - \hat{\rho}_{Y_T})}, \tag{8}$$

where $\boldsymbol{\Omega}$ is a matrix of weights, which define the importance of correlation at different lags. We set $\boldsymbol{\Omega}$ as an identity matrix. Similarity measure based on partial autocorrelation function may be defined analogously, taking PACFs instead of ACFs.

An Adaptive Dissimilarity Index Combining Temporal Correlation and Raw Value Behaviors

The first order temporal correlation coefficient is defined by

$$\text{CORT}(\mathbf{X}_T, \mathbf{Y}_T) = \frac{\sum_{i=1}^{T-1}(X_{t+1} - X_t)(Y_{t+1} - Y_t)}{\sqrt{\sum_{i=1}^{T-1}(X_{t+1} - X_t)^2}\sqrt{\sum_{i=1}^{T-1}((Y_{t+1} - Y_t)^2}}. \tag{9}$$

The CORT coefficient reflect the dynamic behavior of the series (Montero and Vilar, 2014) and is similar to Pearson's coefficient: it belongs to the interval $[-1, 1]$, the value of 1 indicates similar dynamic behavior (their growths in time are similar in direction and rate), the value of -1 implies opposite behavior, while the value of 0 shows no relation. A related dissimilarity measure was proposed by Chouakria and Nagabhushan (2007) and it is defined as

$$d_{\mathrm{CORT}}(\mathbf{X}_T, \mathbf{Y}_T) = \phi_k\left[\mathrm{CORT}(\mathbf{X}_T, \mathbf{Y}_T)\right] \cdot d(\mathbf{X}_T, \mathbf{Y}_T), \tag{10}$$

where $\phi_k(\cdot)$ is an adaptive tuning function to automatically modulate a conventional data distance according to the temporal correlation. Chouakria and Nagabhushan (2007) proposed $\phi_k(u) = \frac{2}{1+\exp(ku)}, k \geq 0$. The advantage of d_{CORT} distance is that it measures both the proximity of observations and temporal correlation for the behavior proximity estimation (Montero and Vilar, 2014).

Fourier Coefficients Based Distance

The simple idea behind this distance is to compare Discrete Fourier Transform coefficients of the series. The value of these coefficients reflects the associated frequency. As note by Agrawal et al (1993), in case of many time series, most of the information is kept in their first n coefficients, where $n < \frac{T}{2} + 1$.

Given two time series $\mathbf{X}_T$ and $\mathbf{Y}_T$ with Fourier Coefficients (respectively) $(a_0, b_0), \ldots, (a_{\frac{T}{2}}, b_{\frac{T}{2}}), (a'_0, b'_0), \ldots, (a'_{\frac{T}{2}}, b'_{\frac{T}{2}})$ we define mentioned distance as

$$\mathrm{FC}(\mathbf{X}_T, \mathbf{Y}_T) = \sqrt{\sum_{i=0}^{n}((a_i - a'_i)^2 + (b_i - b'_i)^2)}. \tag{11}$$

TQuest Distance

Aßfalg et al (2006) proposed a distance measure based on Threshold Queries, using given τ parameter as a threshold in order to transform a time series into a sequence of time stamps, when the threshold is crossed. It is an interesting feature extraction idea, but – in our opinion – highly dependent on user's specialist knowledge, as the τ parameter must be set. In case of parameter choice, we followed remarks made by Ding et al (2008) with the simplification, that we picked up mean value. The full construction of the distance, the equation and a synthetic description is given by Mori et al (2016).

Periodogram-Based Measures

The periodograms of $\mathbf{X}_T$ and $\mathbf{Y}_T$ are given (respectively) by:$I_{X_T}(\lambda_k)$ and $I_{Y_T}(\lambda_k)$ for $k = 1, \ldots, n$. Based on it, Caiado et al (2006) proposed the Euclidean distance between the periodogram coordinates

$$d_{\mathrm{P}}(\mathbf{X}_T, \mathbf{Y}_T) = \frac{1}{n}\sqrt{\sum_{k=1}^{n}\left(I_{X_T}(\lambda_k) - I_{Y_T}(\lambda_k)\right)^2}. \tag{12}$$

Alternatively, Casado de Lucas (2010) introduced a distance measure based on integrated periodogram, arguing that – due to some properties of integrated periodogram – it presents several adventages over the one based on periodogram. Casado de Lucas proposed the following equation:

$$d_{\mathrm{IP}}(\mathbf{X}_T, \mathbf{Y}_T) = \int_{-\pi}^{\pi} |F_{\mathbf{X}_T}(\lambda) - F_{\mathbf{Y}_T}(\lambda)| \, \mathrm{d}\lambda, \tag{13}$$

where $F_{\mathbf{X}_T}(\lambda_j) = C_{\mathbf{X}_T}^{-1} \sum_{i=1}^{j} I_{\mathbf{X}_T}(\lambda_i)$ and $F_{\mathbf{Y}_T}(\lambda_j) = C_{\mathbf{Y}_T}^{-1} \sum_{i=1}^{j} I_{\mathbf{Y}_T}(\lambda_i)$, with $C_{\mathbf{X}_T} = \sum_i I_{\mathbf{X}_T}(\lambda_i)$, $C_{\mathbf{Y}_T} = \sum_i I_{\mathbf{Y}_T}(\lambda_i)$.

Dissimilarity Measures Based on Nonparametric Spectral Estimators

Kakizawa et al (1998) proposed a general spectral disparity measure between two time series as

$$d_{\mathrm{LLR}}(\mathbf{X}_T, \mathbf{Y}_T) = \int_{-\pi}^{\pi} \tilde{W} \left(\frac{f_{X_T}(\lambda)}{f_{Y_T}(\lambda)} \right) \mathrm{d}\lambda, \tag{14}$$

where f_{X_T} and f_{Y_T} are spectral densities of $\mathbf{X}_T$ and $\mathbf{Y}_T$. $\tilde{W} = W(x) + W(x^{-1})$, $W(x) = \log(\alpha x + (1 - \alpha)) - \alpha \log x$, with $0 < \alpha < 1$. $W(\cdot)$ is a divergence function satisfying regular quasi-distance conditions for d_{LLR}. Alternatively, Díaz and Vilar (2010) proposed two distances – the first one is defined as

$$\begin{aligned} d_{\mathrm{GLK}}(\mathbf{X}_T, \mathbf{Y}_T) = \sum_{k=1}^{n} \Big[Z_k - \hat{\mu}(\lambda_k) - 2\log(1 + e^{Z_k - \hat{\mu}(\lambda_k)}) \Big] - \\ \sum_{k=1}^{n} \Big[Z_k - 2\log(1 + e^{Z_k}) \Big], \end{aligned} \tag{15}$$

where $Z_k = \log(I_{X_T}(\lambda_k)) - \log(I_{Y_T}(\lambda_k))$ and $\hat{\mu}(\lambda_k)$ is the local maximum log-likelihood estimator of $\mu(\lambda_k) = \log(f_{X_T}(\lambda_k) - \log(f_{Y_T}(\lambda_k)$ computed by local linear fitting. The second distance is given by

$$d_{\mathrm{ISD}}(\mathbf{X}_T, \mathbf{Y}_T) = \int_{-\pi}^{\pi} \left(\hat{m}_{X_T}(\lambda) - \hat{m}_{Y_T}(\lambda) \right)^2 \mathrm{d}\lambda, \tag{16}$$

where $\hat{m}_{X_T}(\lambda)$ and $\hat{m}_{Y_T}(\lambda)$ are local linear smoothers of the log-periodograms obtained with the maximum local likelihood criterion.

Dissimilarity Based on the Symbolic Representation SAX

The symbolic approximation representation (SAX) has been introduced by Lin et al (2003) and became one of the best symbolic representation for most time series problems (Keogh, 2018). The original data are first transformed into the piecewise aggregate approximation (PAA) representation (Yi and Faloutsos, 2000) and then into a discrete string. For the full outline of MINDIST dissimilarity measure based on SAX representation see Lin et al (2007). Concerning the parameter choice, we followed Lin et al (2007), with the simplification that we set $w = \frac{T}{3}$, where T is the length of time series.

2.4 Structure-Based Distances

The last group of distance measures tries to find some higher-level structures and then compare time series on these basis. This category can be subdivided into two further groups: model-based – aiming to fit a model and then to compare coefficients through a certain distance function and compression-based which work by compression ratios.

Piccolo Distance

For the class of invertible ARIMA processes, denoting the vectors of $AR(k_1)$ and $AR(k_2)$ for $\mathbf{X}_T$ and $\mathbf{Y}_T$ respectively by $\hat{\mathbf{\Pi}}_{X_T} = (\hat{\pi}_{1,X_T}, \ldots, \hat{\pi}_{k_1,X_T})$ and $\hat{\mathbf{\Pi}}_{Y_T} = (\hat{\pi}_{1,Y_T}, \ldots, \hat{\pi}_{k_2,Y_T})$, Piccolo (1990) proposed the following dissimilarity measure:

$$d_{\mathrm{PIC}}(\mathbf{X}_T, \mathbf{Y}_T) = \sqrt{\sum_{j=1}^{k} \left(\hat{\pi}'_{j,X_T} - \hat{\pi}'_{j,Y_T} \right)^2}, \tag{17}$$

where $k = \max(k_1, k_2)$, $\hat{\pi}'_{j,X_T} = \hat{\pi}_{j,X_T}$ if $j \leq k_1$ and $\hat{\pi}'_{j,X_T} = 0$ otherwise and analogously $\hat{\pi}'_{j,Y_T} = \hat{\pi}_{j,Y_T}$ if $j \leq k_2$ and $\hat{\pi}'_{j,Y_T} = 0$ otherwise.

Compression-Based Dissimilarity

Keogh et al (2004) proposed compression-based dissimilarity measure defined as

$$d_{\mathrm{CDM}}(\mathbf{X}_T, \mathbf{Y}_T) = \frac{\mathrm{C}(\mathbf{X}_T, \mathbf{Y}_T)}{\mathrm{C}(\mathbf{X}_T)\mathrm{C}(\mathbf{Y}_T)}. \tag{18}$$

The CDM distance is descended from normalized compression distance (NCD) proposed by Lin et al (2004), using the compressed size of $\mathbf{X}_T$ – $C(\mathbf{X}_T)$ – as an approximation of Kolmogorov complexity. $C(\mathbf{X}_T)$ may be computed as the size of $\mathbf{X}_T$ compressed using data compressors for example: bzip2 or gzip.

Permutation Distribution Clustering

Dissimilarity measures based on permutation distribution clustering (PDC) use a permutation $\Pi(\mathbf{X}_T^{'})$ of m-dimensional embedding of $\mathbf{X}_T$. Dissimilarity between two time series $\mathbf{X}_T$ and $\mathbf{Y}_T$ is expressed in terms of the divergence between distribution of these permutations, denoted by $P(\mathbf{X}_T), P(\mathbf{Y}_T)$. Specifically, Brandmaier (2011) proposed the α-divergence between $P(\mathbf{X}_T)$ and $P(\mathbf{Y}_T)$ as a dissimilarity between time series $\mathbf{X}_T$ and $\mathbf{Y}_T$.

3 Experimental Design

We performed experiments on 34 datasets with time series that come from the UCR time series repository (Dau et al, 2018). The datasets originate from a plethora of different domains, including medicine, robotics, astronomy, biology, face recognition, handwriting recognition, etc. Within the data, the number of classes ranges from 2 to 50, the number of time series per dataset ranges from 56 to 9236, and time series lengths range from 60 to 1882 samples. We limited only to smaller datasets from the repository because of computational limitations. In our opinion, this sample is big enough to obtain interesting insight about examined distances.

In our paper, we will follow the methodology proposed by Keogh and Kasetty (2003), which assumes evaluating the efficacy of distance measures by the prism of classification accuracy of the 1NN classifier (error rate on a test subset). While one should be aware that the proposed approach cannot deliver the overall evaluation of a distance measure, there seems to be more pros than cons of the chosen method. For example, Wang et al (2013) pointed out three aspects: simplicity of implementation, performance directly dependent on distance choice, and relatively (to other, often more complex classifiers) good performance. For more information we refer to Batista et al (2011); Ding et al (2008); Tan et al (2005); Tomašev and Mladenić (2012); Xi et al (2006). Additionally, the 1NN classifier is probably one of the most popular algorithms in data mining (Wu et al, 2008).

Table 2: Error rates (in %) of all considered distance measures on 1NN classifier. Best classifier for each dataset was bold. In the last row we computed number of wins for each distance.

Dataset	ED	MAN	MIN	INF	CC	STS	DTW	DTWc	LCSS	FC	TQ	ACF	PACF	PIC	CDM	CID	PC2	CORT	IP	P	SAX	NCD	GLK	ISD	LLR	PDC
50ww	37	33	40	47	40	48	31	**24**	32	37	43	73	67	74	95	34	37	39	67	60	43	93	80	74	80	90
Adiac	39	40	38	38	38	42	40	39	97	39	75	44	45	84	88	**37**	39	40	44	51	95	88	81	76	63	89
Beef	33	37	33	**30**	37	**30**	50	33	57	33	57	57	47	53	80	37	33	33	53	57	63	77	63	63	60	67
Car	27	28	27	28	27	33	27	**23**	57	27	35	62	67	70	77	27	27	25	32	45	40	75	52	42	57	55
CBF	15	11	20	58	26	66	**0**	0	4	15	32	32	59	50	44	2	15	21	38	31	21	43	45	43	42	61
CinC	10	6	13	20	13	21	35	7	7	10	31	49	15	38	48	8	10	9	36	19	31	47	13	11	27	**1**
Coffee	**0**	4	**0**	**0**	**0**	7	18	**0**	50	**0**	39	7	**0**	14	46	**0**	**0**	**0**	**0**	7	43	50	21	18	14	50
CrickX	42	37	50	67	58	86	**22**	23	26	43	50	53	68	75	84	37	42	47	50	41	48	83	69	68	64	87
CrickY	43	37	47	62	56	84	**21**	24	21	43	53	61	77	80	83	42	43	41	54	47	54	85	78	75	71	87
CrickZ	41	34	45	65	58	79	**21**	25	24	41	48	50	68	79	84	41	41	44	52	39	46	83	71	65	65	87
Diatom	6	7	6	4	7	11	**3**	6	70	7	12	8	12	39	61	7	7	6	7	7	69	62	26	22	44	34
ECG2	12	**11**	**11**	13	**11**	15	23	12	12	12	22	26	20	24	33	**11**	12	12	20	22	16	32	20	15	16	38
ECGF	20	21	21	29	17	38	23	20	6	20	23	1	23	32	40	22	20	24	3	**0**	46	39	28	26	17	38
FaceF	22	16	27	45	31	48	17	**11**	18	22	12	47	56	69	69	19	22	23	41	34	22	57	58	51	52	56
FacesU	23	20	27	41	27	35	10	**9**	10	23	25	26	41	65	88	23	23	26	41	23	38	87	55	52	51	85
FISH	22	21	20	21	21	19	17	**15**	85	22	60	60	45	76	81	22	22	21	29	37	80	82	70	65	70	72
GunP	9	**5**	12	15	9	9	9	8	27	9	19	8	**5**	33	33	7	9	7	9	11	29	31	22	23	10	34
Haptics	63	64	61	61	62	69	62	59	69	63	62	75	69	73	85	**58**	63	**58**	62	65	69	83	74	75	66	72
InlineS	66	65	67	71	67	77	62	61	78	66	75	80	53	66	83	63	66	65	66	70	80	83	25	**18**	65	53
Light2	25	18	34	31	39	49	13	**13**	18	25	39	23	36	48	39	25	25	30	25	28	30	38	36	31	38	49
Light7	42	29	53	55	49	68	**27**	29	42	42	48	48	63	78	66	40	42	47	67	41	44	63	68	62	53	79
Mallat	9	8	9	12	8	15	**7**	9	46	9	31	53	29	39	80	7	9	9	15	13	55	78	35	31	16	41
MedI	32	29	32	33	32	39	26	**25**	33	32	49	35	39	46	61	31	32	31	44	44	55	61	42	40	45	57
MoteS	12	13	18	26	19	28	17	13	14	**12**	17	28	43	46	39	21	12	17	45	42	28	39	44	43	43	17
OliveO	**13**	17	13	17	17	17	13	**13**	83	13	80	23	20	47	20	13	13	13	17	17	80	20	60	63	50	50
OSUL	48	45	48	47	47	61	41	39	37	48	50	52	48	66	70	44	48	46	36	48	50	71	36	**32**	35	57
Plane	4	4	4	4	4	3	**0**	**0**	19	4	5	4	3	2	66	4	4	4	2	4	20	66	11	5	1	50
Sony	30	31	29	31	50	32	27	30	29	30	39	27	32	27	44	18	30	21	22	39	37	41	15	**12**	17	49
SonyII	14	13	16	22	24	17	17	14	16	14	17	23	23	16	44	**12**	14	14	13	21	20	45	21	14	**12**	37
SwedL	21	21	22	29	25	44	21	15	71	21	44	25	39	45	82	12	21	19	**12**	23	76	80	39	27	27	64
SynthC	12	12	12	19	48	65	**1**	2	6	11	15	37	45	48	78	5	12	26	51	36	18	78	39	34	34	57
Trace	24	24	23	31	21	41	**0**	1	26	24	46	20	7	18	50	14	24	28	**0**	16	54	44	**0**	**0**	**0**	47
TwoL	25	27	26	27	24	15	10	13	48	25	27	15	27	36	41	23	25	22	**7**	21	45	36	27	31	23	38
WordsS	38	37	41	51	41	50	35	**25**	33	38	47	70	69	76	88	36	38	41	63	63	46	87	77	73	77	85
# wins	2	2	2	2	2	1	10	11	0	2	0	0	2	0	0	5	1	2	3	1	0	0	1	4	2	1

For each dataset we calculated the classification error rate on a test subset. An appropriate distribution of the training and test sets was proposed by the authors of the repository (each dataset is divided into a training and testing subset). Specifically, for each dataset, we computed the classification error rate on a test subset. When a parameter to train the 1NN classifier was needed, we tried to use values proposed already in the literature (referred to in Section 2). The computations were carried out using Intel Xeon E5-2697 architecture provided by Poznań Supercomputing and Networking Center.

4 Results

The results are presented in Table 2. We present the absolute error rates on the test subset with the 1NN classifier for each of 26 distance measures. In Table 3 we computed average ranks for all considered distances.

Table 3: Average ranks in ascending order for considered distances across all datasets.

DTWc	CID	DTW	MAN	FC	ED	PC2	CORT	MIN	CC	IP	LCSS	P
4.0	6.0	6.2	7.4	8.3	8.4	8.4	8.5	10.2	11.8	12.0	13.0	13.0
INF	LLR	ACF	PACF	ISD	TQ	STS	GLK	SAX	PIC	PDC	NCD	CDM
13.6	15.0	15.2	15.8	15.9	16.1	16.3	18.0	18.4	19.9	21.9	23.4	24.1

Note: An overview of the used abbreviations can be found in the appendix on page 19.

If we look at the overall result, we can observe that none of the compared distances achieves the best performance for all, or even the most of datasets. In fact, the lowest error rate is computed for DTWc (11 wins) and DTW (10 wins), far ahead of other distances. On the other hand, looking at average ranks, one may be surprised by the good performance of L_p norms: MAN – 7.4, ED – 8.4, MIN – 10.2. It is also worth mentioning that Complexity-invariant distance (CID) achieved the second best average rank (equal 6.0), while in fact it only improves the Euclidean distance by a simple complexity correction factor.

Looking at certain datasets, we see, that some of them are almost perfectly classified (e.g. *CinC ECGF, Coffee, ECG2, Plane*), which could mean that their classes are relatively easy to recognize by the algorithm. Another interesting fact is, that there are datasets, which are better classified by some group of distances. For example, performance of L_p norms is relatively good for *Coffee, Mallat, SynthetC*, while clearly worse for *CrickX, CrickY, Haptics*, which may indicate cases, where we should pay attention to shape (without editing) or not. Correlation-based distances (e.g. ACF, PACF, CC) may be considered as a good choice for datasets *ECGF* and *Trace*.

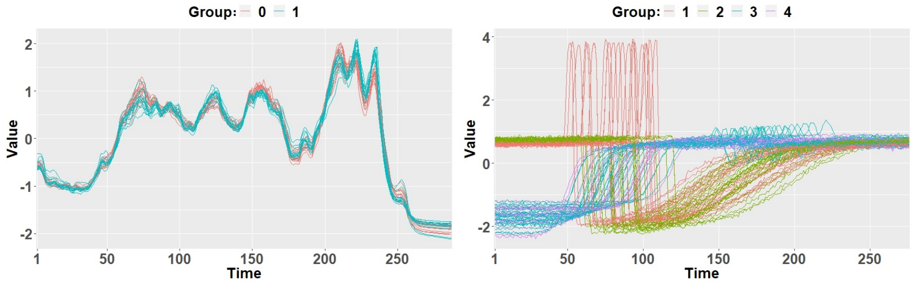

Figure 1: Datasets *Coffee* (left) and *Trace* (right).

In Figure 1, we present plots of two datasets, which show a comparison between a good performance of L_P norms and weak of DTW (dataset *Coffee*) and the opposite situation (dataset *Trace*) – a weak performance of L_p norms and good of DTW. It happens, because sometimes elastic measures may tray to force matching, when it is not desirable (e.g. dataset *Coffee*). On the other hand, sometimes peaks are slightly shifted and using elastic measures can bring significant increase of accuracy (e.g. dataset *Trace*).

To assess the differences between the examined methods, we performed a statistical comparison. Firstly, we employed the test proposed by Iman and Davenport (1980), which is a less conservative variant of Friedman's ANOVA (Friedman, 1940). If the hypotheses, that there is no significant difference between classifiers, is rejected, we can proceed with the post hoc test to provide all pairwise comparisons. In this way we can detect the statistically significant differences between certain classifiers. Garcia and Herrera (2008) proved that the procedure presented in Bergmann and Hommel (1988) is the most powerful post hoc comparison test.

The p-value from the Iman and Davenport's test is equal to 0. We can therefore proceed with the post hoc tests. The results of multiple comparisons are given in Table 4. We have chosen for the comparison 7 distance measures, which achieved the best average ranks. Based on it, we see that there is a group of distances distinct from others: DTW, DTWc and CID. They are in most cases statistically significantly different from the others, which is also in line with the interpretation of average ranks.

Table 4: p-values in the Bergmann-Hommenl post hoc test for best 7 measures (taking into account average ranks). Statistically significant differences ($p < 0.05$) are in bold.

	ED	MAN	DTW	DTWc	FC	CID	PC2
ED	n/a	1.000	**0.005**	**0.000**	1.000	**0.040**	1.000
MAN	1.000	n/a	0.181	**0.001**	1.000	0.679	1.000
DTW	**0.005**	0.181	n/a	0.679	**0.011**	1.000	**0.005**
DTWc	**0.000**	**0.001**	0.679	n/a	**0.000**	0.181	**0.000**
FC	1.000	1.000	**0.011**	**0.000**	n/a	0.064	1.000
CID	**0.040**	0.679	1.000	0.181	0.064	n/a	**0.040**
PC2	1.000	1.000	**0.005**	**0.000**	1.000	**0.040**	n/a

Note: An overview of the used abbreviations can be found in the appendix on page 19.

Regarding the computation time, we present figures 2, 3 and 4. In Figure 2 we can see, that most of measures are quite fast, with medians around several minutes, while there are some distances taking even several hours for one dataset. Having measured computations times and efficiency, it is reasonable to see, if we can distinguish groups of measures based on these features. Figure 3 presents results of k-means clustering. The optimal number of clusters is equal 3. In Figure 4, mean ranks and mean times of each measure are shown. Additionally, measures are grouped according to clusters. These results suggest, that we may distinguish the 3 following groups: weak and slow (grey), highly efficient, but with various speed (green), moderately powerful, but quite fast (red). This plot may be interpreted also as a "price to value" plot. It is worth noting, that in some cases it may be enough to use CID or ED distances to ensure fast computations. On the other hand, when we expect high efficacy, DTW or DTWc may be a reasonable choice.

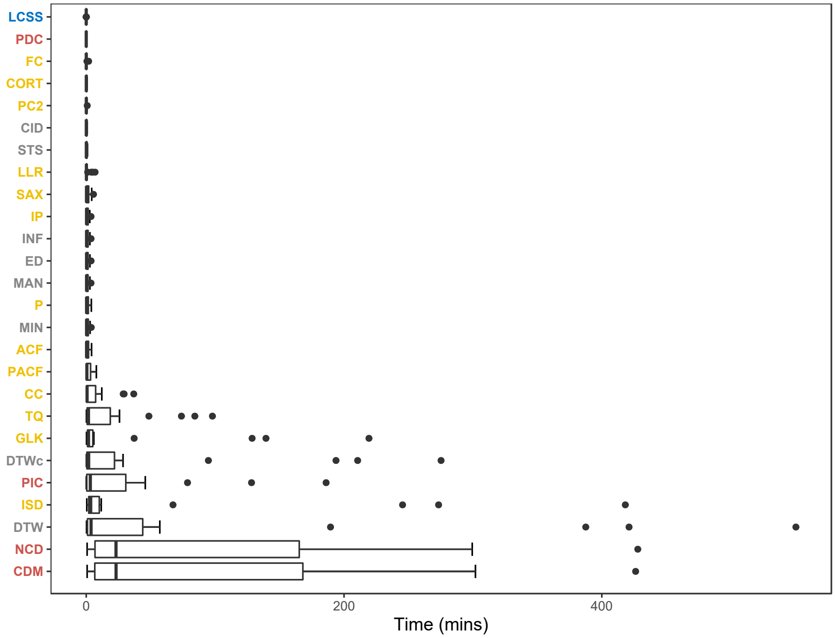

Figure 2: Box plot of time in minutes needed for computations for each measure across all datasets. Y-axis labels are colored according to the category of a measure: shape-based (grey), edit-based (blue), feature-based (yellow), structure-based (red).

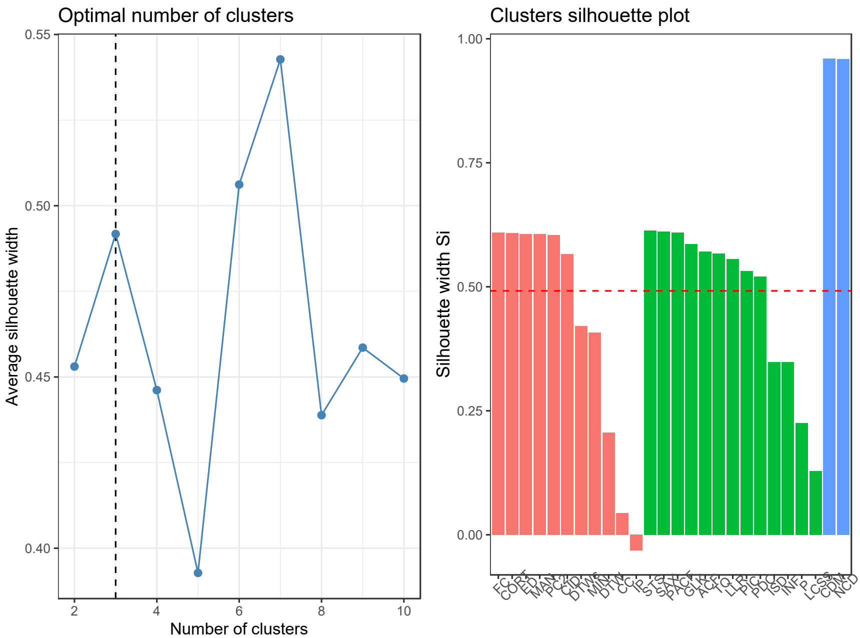

Figure 3: Results of clustering using k-means algorithm.

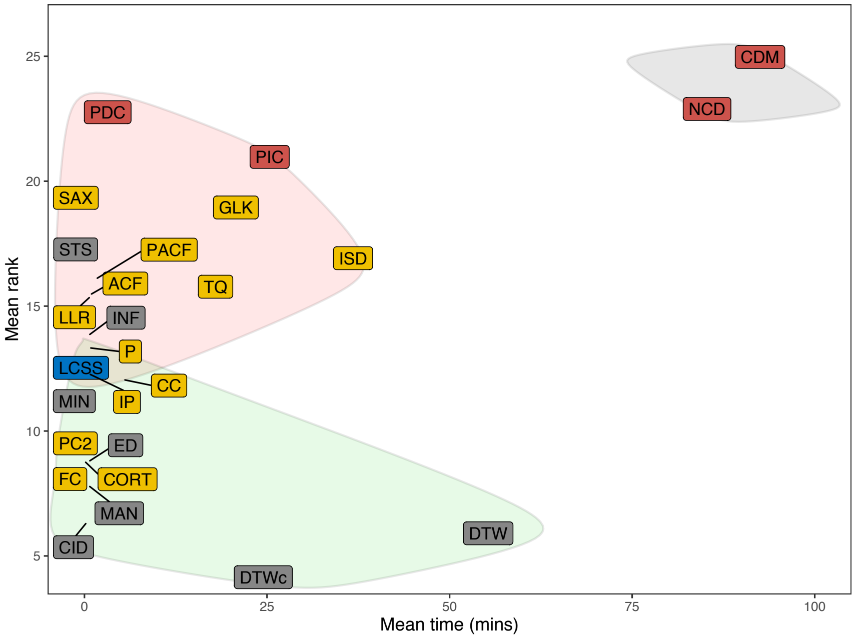

Figure 4: Mean time and mean rank for each measure. Boxes are colored according to the category of a measure: shape-based (grey), edit-based (blue), feature-based (yellow), structure-based (red). Measures are grouped in 3 clusters based on the k-means algorithm.

5 Conclusion

In this article, we have compared the efficacy of 26 distance measures on 34 datasets, based on the prism of 1NN classifier accuracy. Similarly to Serrà et al (2012); Wang et al (2013), we have observed, that there is no measure distinctly better than the others or appropriate for a majority of datasets. Thus, there is still a place for new ones, maybe connecting some properties of already existing measures. On the other hand, best average ranks were achieved by DTWc and DTW which shows that processing the shape of time series in a smart way may be a direction for future researches. We have also observed that there are some datasets that are classified better with some groups of measures. It would be highly desirable to find a set of metadata, which could help us to choose the most appropriate measure. We have also shown, that – taking into consideration computation times and efficiency of measures – we can distinguish 3 groups of distances: weak and slow, highly efficient, but with various speed, moderately powerful, but quite fast.

Since this study only discussed 26 of about 40 available distance measures, there is still potential to extend the presented comparison. We plan to cover all available distance measures in the nearest future and, as well, extend the number of datasets for testing them. It would be also interesting to confront conclusions made during these analyses with different time series mining tasks, e.g. with clustering.

Acknowledgements Resources supporting this work were provided by the Poznań Supercomputing and Networking Center.

Appendix

ACF	Autocorrelation based distance
CC	Distance based on the Cross Correlation
CDM	Compression based Dissimilarity
CID	Complexity Invariant Distance
CORT	Dissimilarity Index Combining Temporal Correlation and Raw Value Behaviors
DTW	Dynamic Time Warping distance
DTW_c	Dynamic Time Warping distance with Sakoe-Chiba band
ED	Euclidean Distance
FC	Fourier Coefficients based distance
INF	Infinite Norm Distance
IP	Integrated Periodogram based distance
LCSS	Longest Common Subsequence distance
LLR, GLK, ISD	Dissimilarity measures based on nonparametric spectral estimator
MAN	Manhattan distance
MIN	Minkowski distance
NCD	Normalized Compression Distance
P	Periodogram based distance
PACF	Partial Autocorrelation based distance
PC1, PC2	Distances based on Pearson's Correlation
PDC	Dissimilarity measure based on Permutation Distribution Clustering
PIC	Piccolo Distance
SAX	Dissimilarity based on the Symbolic Representation SAX
STS	Short Time Series distance
TQ	Threshold Queries (TQUEST) distance

References

Agrawal R, Faloutsos C, Swami AN (1993) Efficient Similarity Search in Sequence Databases. In: Foundations of Data Organization and Algorithms (FODO 1993), Lomet DB (ed), Springer, Berlin, Heidelberg (Germany), Lecture Notes in Computer Science (LNCS), vol. 730, pp. 69–84. DOI: 10.1007/3-540-57301-1_5.

Antunes C, Oliveira A (2001) Temporal Data Mining: An Overview. KDD Workshop on Temporal Data Mining, pp. 1–13.

Aßfalg J, Kriegel HP, Kröger P, Kunath P, Pryakhin A, Renz M (2006) Similarity Search on Time Series Based on Threshold Queries. In: Advances in Database Technology – Proceedings of the International Conference on Extending Database Technology (EDBT 2006), Ioannidis Y, Scholl MH, Schmidt JW, Matthes F, Hatzopoulos M, Boehm K, Kemper A, Grust T, Boehm C (eds), Springer, Berlin, Heidelberg (Germany), Lecture Notes in Computer Science (LNCS), vol. 3896, pp. 276–294. DOI: 10.1007/11687238_19.

Aßfalg J, Kriegel HP, Kröger P, Kunath P, Pryakhin A, Renz M (2008) Similarity Search in Multimedia Time Series Data Using Amplitude-Level Features. In: Advances in Multimedia Modeling – Proceedings of the International Conference on Multimedia Modeling (MMM 2008), Satoh S, Nack F, Etoh M (eds), Springer, Berlin, Heidelberg (Germany), Lecture Notes in Computer Science (LNCS), vol. 4903, pp. 123–133. DOI: 10.1007/978-3-540-77409-9_12.

Bagnall A, Lines J, Bostrom A, Large J, Keogh E (2017) The Great Time Series Classification Bake Off: A Review and Experimental Evaluation of Recent Algorithmic Advances. Data Mining and Knowledge Discovery 31(3):606–660, Springer US. DOI: 10.1007/s10618-016-0483-9.

Batista G, Wang X, Keogh EJ (2011) A Complexity-Invariant Distance Measure for Time Series. In: Proceedings of the SIAM International Conference on Data Mining (SDM), Liu B, Liu H, Clifton C, Washio T, Kamath C (eds), Society for Industrial and Applied Mathematics (SIAM), Philadelphia (USA), pp. 699–710. DOI: 10.1137/1.9781611972818.60.

Bergmann B, Hommel G (1988) Improvements of General Multiple Test Procedures for Redundant Systems of Hypotheses. In: Multiple Hypotheses Testing, Bauer P, Hommel G, Sonnemann E (eds), Springer, Berlin, Heidelberg, Medical Informatics and Statistics (MEDINFO), vol. 70, pp. 100–115. DOI: 10.1007/978-3-642-52307-6_8.

Berndt DJ, Clifford J (1994) Using Dynamic Time Warping to Find Patterns in Time Series. In: Workshop on Knowledge Discovery in Databases (KDD'94), AAAI Press, Seattle (USA), pp. 359–370. URL: `https://www.aaai.org/Papers/Workshops/1994/WS-94-03/WS94-03-031.pdf`.

Brandmaier A (2011) Permutation Distribution Clustering and Structural Equation Model Trees. PhD thesis, Saarland University. DOI: 10.22028/D291-26289.

Caiado J, Crato N, Peña D (2006) A Periodogram-based Metric for Time Series Classification. Computational Statistics & Data Analysis 50(10):2668–2684, Elsevier Science Publishers B. V., Amsterdam (The Netherlands). DOI: 10.1016/j.csda.2005.04.012.

Casado de Lucas D (2010) Classification Techniques for Time Series and Functional Data. PhD thesis, Universidad Carlos III de Madrid. URL: http://www.casado-d.org/edu/Casado-Thesis.pdf.

Chen L, Ng R (2004) On the Marriage of Lp-norms and Edit Distance. VLDB Endowment, Toronto (Canada), Proceedings of the 30th International Conference on Very large Data Bases (VLDB ’04), vol. 30, pp. 792–803.

Chen L, Özsu MT, Oria V (2005) Robust and Fast Similarity Search for Moving Object Trajectories. Association for Computing Machinery (ACM), New York (USA), Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data (SIGMOD ’05), pp. 491–502. DOI: 10.1145/1066157.1066213.

Chen Y, Nascimento MA, Ooi BC, Tung AKH (2007) SpADe: On Shape-based Pattern Detection in Streaming Time Series. In: Proceedings of the IEEE 23rd International Conference on Data Engineering, Institute of Electrical and Electronics Engineers (IEEE), Istanbul (Turkey), pp. 786–795. DOI: 10.1109/ICDE.2007.367924.

Chouakria AD, Nagabhushan PN (2007) Adaptive Dissimilarity Index for Measuring Time Series Proximity. Advances in Data Analysis and Classification (ADAC’07) 1(1):5–21, Springer. DOI: 10.1007/s11634-006-0004-6.

Cuturi M (2011) Fast Global Alignment Kernels. In: Proceedings of the 28th International Conference on Machine Learning (ICML’11), Bellevue (USA), pp. 929–936. URL: http://www.icml-2011.org/papers/489_icmlpaper.pdf.

Dau HA, Keogh E, Kamgar K, Yeh CCM, Zhu Y, Gharghabi S, Ratanamahatana CA, Yanping, Hu B, Begum N, Bagnall A, Mueen A, Batista G (2018) The UCR Time Series Classification Archive. URL: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

Díaz SP, Vilar JA (2010) Comparing Several Parametric and Nonparametric Approaches to Time Series Clustering: A Simulation Study. Journal of Classification 27(3):333–362, Springer, Berlin, Heidelberg (Germany). DOI: 10.1007/s00357-010-9064-6.

Ding H, Trajcevski G, Scheuermann P, Wang X, Keogh E (2008) Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures. Proceedings of the VLDB Endowment 1(2):1542–1552, VLDB Endowment. DOI: 10.14778/1454159.1454226.

Esling P, Agon C (2012) Time-Series Data Mining. ACM Computing Surveys (CSUR) 45(1), New York (USA). DOI: 10.1145/2379776.2379788.

Frentzos E, Gratsias K, Theodoridis Y (2007) Index-based Most Similar Trajectory Search. In: Proceedings of the IEEE 23rd International Conference on Data Engineering, Institute of Electrical and Electronics Engineers (IEEE), Istanbul (Turkey), pp. 816–825. DOI: 10.1109/ICDE.2007.367927.

Friedman M (1940) A Comparison of Alternative Test of Significance for the Problem of M Rankings. The Annals of Mathematical Statistics 11(1):86–92.

Garcia S, Herrera F (2008) An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. Journal of Machine Learning Research 9:2677–2694. URL: http://www.jmlr.org/papers/v9/garcia08a.html.

Golay X, Kollias S, Stoll G, Meier D, Valavanis A, Boesiger P (2005) A New Correlation-based Fuzzy Logic Clustering Algorithm for FMRI. Magnetic Resonance in Medicine 40(2):249–60, Bernstein MA (ed), International Society for Magnetic Resonance in Medicine. DOI: 10.1002/mrm.1910400211.

Górecki T (2018) Classification of Time Series Using Combination of DTW and LCSS Dissimilarity Measures. Communications in Statistics – Simulation and Computation 47(1):263–276, Taylor & Francis. DOI: 10.1080/03610918.2017.1280829.

Górecki T, Łuczak M (2013) Using Derivatives in Time Series Classification. Data Mining and Knowledge Discovery 26(2):310–331, Springer US. DOI: 10.1007/s10618-012-0251-4.

Iman RL, Davenport JM (1980) Approximations of the Critical Region of the Fbietkan Statistic. Communications in Statistics – Theory and Methods 9(6):571–595, Taylor & Francis. DOI: 10.1080/03610928008827904.

Kakizawa Y, Shumway RH, Taniguchi M (1998) Discrimination and Clustering for Multivariate Time Series. Journal of the American Statistical Association 93(441):328–340, Taylor & Francis, American Statistical Association (ASA). URL: https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1998.10474114.

Kalpakis K, Gada D, Puttagunta V (2001) Distance Measures for Effective Clustering of ARIMA Time-Series. In: Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM'01), Institute of Electrical and Electronics Engineers (IEEE), San Jose (USA), pp. 273–280. DOI: 10.1109/ICDM.2001.989529.

Keogh E (2006) A Decade of Progress in Indexing and Mining Large Time Series Databases. Proceedings of the 32nd International Conference on Very large Data Bases (VLDB'06), pp. 1268–1268, VLDB Endowment, Seoul (Korea).

Keogh E (2018) The SAX (Symbolic Aggregate aproXimation). URL: http://www.cs.ucr.edu/~eamonn/SAX.htm [accessed 2018/10/26].

Keogh E, Kasetty S (2003) On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. Data Mining and Knowledge Discovery 7(4):349–371, Kluwer Academic Publishers. DOI: 10.1023/A:1024988512476.

Keogh E, Lin J (2005) Clustering of Time-Series Subsequences is Meaningless: Implications for Previous and Future Research. Knowledge and Information Systems 8(2):154–177, Springer, London. DOI: 10.1007/s10115-004-0172-7.

Keogh E, Ratanamahatana CA (2005) Exact Indexing of Dynamic Time Warping. Knowledge and Information Systems 7(3):358–386, Springer, London. DOI: 10.1007/s10115-004-0154-9.

Keogh E, Lonardi S, Ratanamahatana CA (2004) Towards Parameter-free Data Mining. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04), Association for Computing Machinery (ACM), New York (USA), pp. 206–215. DOI: 10.1145/1014052.1014077.

Keogh EJ, Pazzani MJ (2001) Derivative Dynamic Time Warping. In: Proceedings of the 1st SIAM International Conference on Data Mining (SDM'2001), Chicago (USA), vol. 1, pp. 1–11. ISBN: 978-0-898714-95-1, DOI: 10.1137/1.9781611972719.1

Lin J, Keogh E, Lonardi S, Chiu B (2003) A Symbolic Representation of Time Series, With Implications for Streaming Algorithms. In: Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD '03), Association for Computing Machinery (ACM), New York (USA), pp. 2–11. DOI: 10.1145/882082.882086.

Lin J, Keogh E, Lonardi S, Lankford JP, Nystrom DM (2004) Visually Mining and Monitoring Massive Time Series. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04), Association for Computing Machinery (ACM), New York (USA), pp. 460–469. DOI: 10.1145/1014052.1014104.

Lin J, Keogh E, Wei L, Lonardi S (2007) Experiencing SAX: A Novel Symbolic Representation of Time Series. Data Mining and Knowledge Discovery 15(2):107–144, Springer US. DOI: 10.1007/s10618-007-0064-z.

Maharaj E (2000) Cluster of Time Series. Journal of Classification 17(2):297–314, Springer US. DOI: 10.1007/s003570000023.

Marteau PF (2009) Time Warp Edit Distance with Stiffness Adjustment for Time Series Matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2):306–318, Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/TPAMI.2008.76.

Möller-Levet CS, Klawonn F, Cho KH, Wolkenhauer O (2003) Fuzzy Clustering of Short Time-Series and Unevenly Distributed Sampling Points. In: International Symposium on Intelligent Data Analysis, Advances in Intelligent Data Analysis (IDA'2003), R. Berthold M, Lenz HJ, Bradley E, Kruse R, Borgelt C (eds), Springer, Berlin, Heidelberg (Germany), pp. 330–340. DOI: 10.1007/978-3-540-45231-7_31.

Montero P, Vilar JA (2014) TSclust: An R Package for Time Series Clustering. Journal of Statistical Software 62(1):1–43. DOI: 10.18637/jss.v062.i01.

Mori U, Mendiburu A, Lozano JA (2016) Distance Measures for Time Series in R: The TSdist Package. R journal 8(2):451–459. ISSN: 2073-4859, URL: https://journal.r-project.org/archive/2016/RJ-2016-058/RJ-2016-058.pdf.

Papadopoulos AN (2008) Trajectory Retrieval with Latent Semantic Analysis. In: Proceedings of the 2008 ACM Symposium on Applied computing (SAC'08), Association for Computing Machinery (ACM), Fortaleza (Brazil), pp. 1089–1094. ISBN: 978-1-595937-53-7.

Peña D, Galeano P (2000) Multivariate Analysis in Vector Time Series. DES Working Papers, Statistics and Econometrics ws012415. URL: http://halweb.uc3m.es/esp/Personal/personas/dpena/publications/ingles/2000RESENHAS_galeano.pdf.

Piccolo D (1990) A Distance Measure for Classifying ARIMA Models. Journal of Time Series Analysis 11(2):153–164. DOI: 10.1111/j.1467-9892.1990.tb00048.x

Radovanović M, Nanopoulos A, Ivanović M (2010) Time-Series Classification in Many Intrinsic Dimensions. Proceedings of the 2010 SIAM International Conference on Data Mining, pp. 677–688, Srinivasan Parthasarathy BGJPCK Bing Liu (ed), Columbus (USA). DOI: 10.1137/1.9781611972801.59.

Serrà J, Zanin M, Herrera P, Serra X (2012) Characterization and Exploitation of Community Structure in Cover Song Networks. Pattern Recognition Letters 33(1):1032–1041. DOI: 10.1016/j.patrec.2012.02.013.

Spiegel S, Jain B, Albayrak S (2014) Fast Time Series Classification under Lucky Time Warping Distance. In: Proceedings of the 29th Annual ACM Symposium on Applied Computing, Association for Computing Machinery (ACM), Gyeongju (Korea), pp. 71–78. DOI: 10.1145/2554850.2554885.

Tan PN, Steinbach M, Kumar V (2005) Introduction to Data Mining, 1st edn. Addison-Wesley Longman Publishing Co., Inc., Boston (USA). ISBN: 03-2132-136-7.

Tomašev N, Mladenić D (2012) Nearest Neighbor Voting in High-Dimensional Data: Learning from Past Occurrences. Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops 9(2):691–712, Institute of Electrical and Electronics Engineers (IEEE), Vancouver (Canada). DOI: 10.1109/ICDMW.2011.127.

Vilar JA, Alonso AM, Vilar JM (2010) Non-linear Time Series Clustering Based on Non-parametric Forecast Densities. Computational Statistics & Data Analysis 54(11):2850–2865, Elsevier Science Publishers B. V. DOI: 10.1016/j.csda.2009.02.015.

Vlachos M, Kollios G, Gunopulos D (2002) Discovering Similar Multidimensional Trajectories. In: Proceedings of the 18th International Conference on Data Engineering, Institute of Electrical and Electronics Engineers (IEEE), San Jose (USA), pp. 673–684. DOI: 10.1109/ICDE.2002.994784.

Wang X, Mueen A, Ding H, Trajcevski G, Scheuermann P, Keogh E (2013) Experimental Comparison of Representation Methods and Distance Measures for Time Series Data. Data Mining and Knowledge Discovery 26(2):275–309, Springer US. DOI: 10.1007/s10618-012-0250-5.

Warren Liao T (2005) Clustering of Time Series Data – A Survey. Pattern Recognition 38(11):1857–1874, Elsevier Science Inc., New York (USA). DOI: 10.1016/j.patcog.2005.01.025.

Weigend A, Gershenfeld N (1994) Time Series Prediction: Forecasting the Future and Understanding the Past. Addison-Wesley, Reading (USA), pp. 161–163.

Weiss GM (2004) Mining with Rarity: A Unifying Framework. SIGKDD Explor Newsl 6(1):7–19, Association for Computing Machinery (ACM), New York (USA). ISSN: 1931-0145.

Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou ZH, Steinbach M, Hand DJ, Steinberg D (2008) Top 10 Algorithms in Data Mining. Knowledge and Information Systems 14(1):1–37, Springer, London (United Kingdom). DOI: 10.1007/s10115-007-0114-2.

Xi X, Keogh E, Shelton C, Wei L, Ratanamahatana CA (2006) Fast Time Series Classification Using Numerosity Reduction. In: Proceedings of the 23rd International Conference on Machine Learning (ICML'06), Pittsburgh (USA), pp. 1033–1040. DOI: 10.1145/1143844.1143974.

Yi BK, Faloutsos C (2000) Fast Time Sequence Indexing for Arbitrary Lp Norms. In: Proceedings of the 26th International Conference on Very Large Data Bases (VLDB'00), Morgan Kaufmann Publishers Inc., San Francisco (USA), pp. 385–394. ISBN: 15-5860-715-3.