Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Classification of time series using combination of DTW and LCSS dissimilarity measures

Tomasz Górecki

To cite this article: Tomasz Górecki (2018) Classification of time series using combination of DTW and LCSS dissimilarity measures, Communications in Statistics - Simulation and Computation, 47:1, 263-276, DOI: 10.1080/03610918.2017.1280829

To link to this article: https://doi.org/10.1080/03610918.2017.1280829

Published online: 02 Jun 2017.

COMMUNICATIONS IN STATISTICS—SIMULATION AND COMPUTATION®
2018, VOL. 47, NO. 1, 263–276
https://doi.org/10.1080/03610918.2017.1280829

Classification of time series using combination of DTW and LCSS dissimilarity measures

Tomasz Górecki

Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Umultowska, Poznań, Poland

ABSTRACT
In the domain of time series, different dissimilarity measures are applied for comparing sequences, the most successful ones being based on dynamic programming. Such measures include Longest Common Sub-Sequence (LCSS) and Dynamic Time Warping (DTW). In this article, a novel method is proposed to measure the dissimilarity of time series. We propose a parametric combination of LCSS and derivative DTW. The new dissimilarity measure is used in classification with the 1NN rule. We empirically compare our new approach to LCSS and DTW used separately and demonstrate its superiority in classification accuracy. In addition, our method is statistically comparable with method proposed by Lines and Bagnall considered so far the most accurate method of time series classification.

ARTICLE HISTORY
Received 3 May 2016
Accepted 7 January 2017

KEYWORDS
Derivative dynamic time warping; Dynamic time warping; Longest common subsequence; Time series classification

MATHEMATICS SUBJECT CLASSIFICATION
Primary 62M99; 62H30; Secondary 68T10

1. Introduction

Many time-series analysis methods require dissimilarity measures. A lot of measures have been proposed for this purpose to date. The simplicity of computing (a linear complexity) and efficiency of Euclidean distance (ED; Faloutsos et al., 1994) makes it probably the most popular dissimilarity measure for time series data mining (Agrawal et al., 1993; Keogh et al., 2001). However, ED is known to be quite insufficient in dealing with noise and local time shifting, even worse it could not deal with series that are of unequal length (Euachongprasit and Ratanamahatana, 2008; Ratanamahatana and Keogh, 2005; Yi et al., 1998). These problems could be solved by elastic dissimilarity measures such as DTW (Berndt and Clifford, 1994; Rabiner and Juang, 1993) and LCSS (Agrawal et al., 1995; Paterson and Dancik, 1994; Vlachos et al., 2002). Compared with ED, DTW and LCSS are more elastic, supporting local time shifts and variations in lengths of time series, but they are also costlier to compute (time and space complexity is $O(nm)$, where n and m are the time series lengths). Of the three measures, LCSS is more robust to noise and outliers (Vlachos et al., 2002). On the other hand, recent studies suggest that DTW is a competitive dissimilarity measure in the time series classification context (Ding et al., 2008; Wang et al., 2013).

Known problem of DTW is "singularity" (Keogh and Pazzani, 2001)—a single point in one time series maps onto a large subsection of another. The reason of this is that DTW only considers the point value as features. Keogh and Pazzani (2001) proposed a derivative dynamic time warping (DDTW) to cope with this serious problem. However, DDTW only

CONTACT Tomasz Górecki tomasz.gorecki@amu.edu.pl Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Umultowska 87, 61-614 Poznań, Poland.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lssp.

considers the derivatives of time series and neglects the point values feature, which is not sufficient to compare time series (Górecki and Łuczak, 2013). Benedikt et al. (2008) also showed that DDTW is very noise sensitive. Górecki and Łuczak (2013) proposed a weighted DTW that use additional parameters to achieve better performance. Similarly, the use of derivatives (first and second) for LCSS was proposed by Górecki (2014). It is worth to mention that to solve the "singularity" problem, we can also use global constraints (Kurbalija et al., 2011).

We decided to combine these measures into a single weighted measure of dissimilarity. Our objective was to extend previous works combining the point-to-point dissimilarities (for DTW and LCSS) and dissimilarities of derivatives (for DTW). We decided not to add LCSS distance on derivatives to not complicate the model already with three parameters, and thus does not extend the operation time of the proposed algorithm. For the same reason, we decided to limit to the first derivative only and skip the second derivative (Górecki and Łuczak, 2014). The first component (DTW) adds information about point-to-point distances between time series (Keogh and Pazzani, 2001). The second component (DD_{DTW}) adds information about the shapes of the considered time series (Górecki and Łuczak, 2013). The last component (LCSS) helps us when there are outliers in the data (Górecki, 2014; Vlachos et al., 2002). These ideas were our main motivation and we suspected that such a combination could work.

In the current work, there is also a comparison with the method proposed by Lines and Bagnall (2014)—so far considered as the best method of time series classification.

The method works well on almost all considered time series. Moreover, the computational complexity is actually like for a method with only one parameter. For all these reasons, proposed classifier seems to be a quite universal method for the classification of time series.

In the article, we first review the concept of dynamic time warping and the longest common subsequence dissimilarity measures (Section 2). At the end of that section, we introduce our dissimilarity measure. All used datasets and the experimental methodology are described in Section 3. Following section contains the results of our experiments and statistical analysis of the obtained results. In the same section, we compare our method with Lines and Bagnall (2014) method. Finally, in the last section of our article, we conclude and discuss possible future work.

2. Methods

2.1. Dynamic time warping

Suppose we have two time series: $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_m)$. To find the DTW distance, we first construct a cost (distance) matrix (which represents all pairwise distances) $\boldsymbol{D}$ where the (i, j)th element corresponds to $d(x_i, y_j) = (x_i - y_j)^2$. Then we find a path through the matrix with minimal overall cost (Fig. 1). The DTW dissimilarity measure corresponds to the path with minimal warping cost:

$$DTW(x, y) = \min \sqrt{\sum_{k=1}^{K} w_k},$$

where w_k is the element of matrix $\boldsymbol{D}$ that is also the kth element of a warping path W. To determine an optimal path, one could test every possible warping path. Such a procedure would lead to a computational complexity that is exponential in the length n and m. Fortunately,

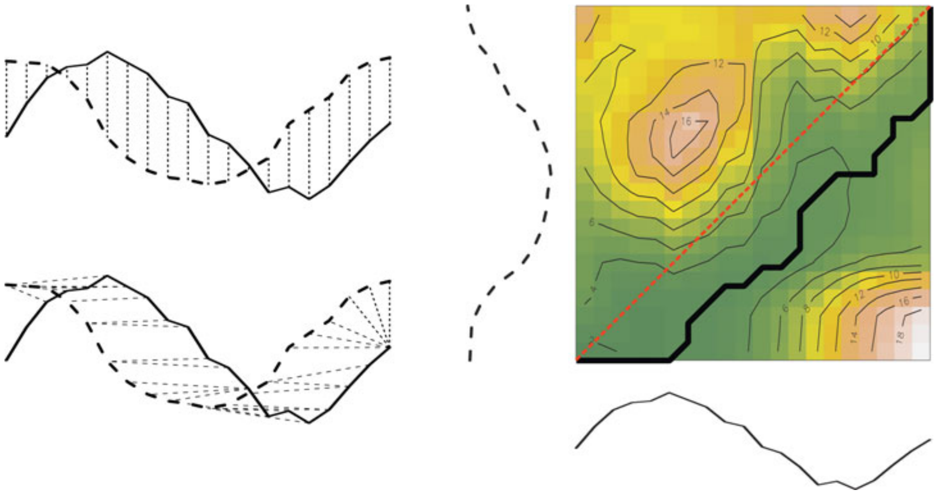

Figure 1. Top left: Two time series that are similar but shifted produce a large Euclidean distance. Bottom left: This can be efficiently corrected by DTW's nonlinear alignment. Right: To align the signals, we construct a cost matrix, and search for the optimal warping path (the bold black line on the heat map).

we can use an $O(nm)$ algorithm that is based on dynamic programming to find the desired warping path.

Unfortunately, described distance is not a metric (does not satisfy the triangle inequality from definition of metric) even in case the local distance measure is a metric (Micó, 1998).

Euclidean distance is a special case of DTW where only the diagonal elements of matrix $\boldsymbol{D}$ are taken into account (red crossed line in Fig. 1), and which is defined if the time series are of the same length.

2.2. Longest common subsequence

The LCSS is a variant of the edit measure usually used in speech recognition. The idea is to match two series by allowing them to stretch, but without rearranging the sequence of the components. However, we allow some components to be unmatched or left out (e.g., outliers)—while in ED and DTW, all components from both series must be used. The LCSS measure has two constant parameters, δ and ε (Fig. 2). The δ parameter, which is commonly

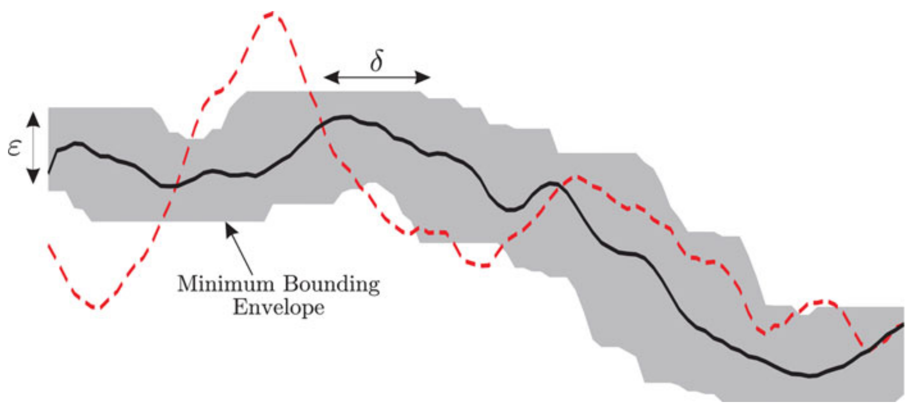

Figure 2. Matching within δ in time and ε in space. Everything outside the bounding envelope can never be matched.

set to a percentage of the series length, is the warping threshold and controls the window size for matching a given point from one series to a point in the other series. It controls how far in time we could go in order to match a given point from one trajectory to a point in another trajectory. The parameter $0 < \varepsilon < 1$ is the matching threshold: two points from two series are considered to match if their distance is less than ε.

Let,

$$L(i, j) = \begin{cases} 0 & \text{for } i = 0 \\ 0 & \text{for } j = 0 \\ 1 + L[i-1, j-1] & \text{for } |x_i - y_j| < \varepsilon \\ & \text{and } |i - j| \leq \delta \\ \max(L[i-1, j], L[i, j-1]) & \text{in other cases} \end{cases}.$$

Now, to define the dissimilarity between x and y, we can compute (Ratanamahatana et al., 2010):

$$\mathrm{LCSS}(x, y) = \frac{n + m - 2L(n, m)}{n + m}.$$

For two signals of equal length, this expression takes values from 0 to 1.

By taking into account only sufficiently similar points, LCSS solves the problem of the presence of noise, but it is not a metric (Vlachos et al., 2002). LCSS is robust to noise and we expect that it should be more accurate than DTW in the presence of outliers and noise. Moreover, let us notice that local distance measure d used with DTW is squared Euclidean, whereas for LCSS it is Manhattan.

2.3. A dissimilarity measure based on a combination of DTW and LCSS distances

A dissimilarity measure that uses the function values of time series for both DTW and LCSS can be defined as

$$D_{DTW}^{\mathrm{LCSS}}(x, y) := aDTW(x, y) + b\,\mathrm{LCSS}(x, y), \tag{1}$$

where $a, b \in [0, 1]$ are parameters.

A dissimilarity measure that uses both the function values of time series and values of the first derivative (for DTW) can be defined as (Górecki and Łuczak, 2013)

$$DD_{DTW}(x, y) := aDTW(x, y) + bDTW(\nabla x, \nabla y), \tag{2}$$

where ∇x and ∇y are the first discrete derivatives of x, y, and $a, b \in [0, 1]$ are parameters. The discrete derivative of a time series x with length n can be defined as

$$\nabla x(i) = x(i+1) - x(i), \ \ i = 1, 2, \ldots, n-1.$$

A distance measure that uses both function values of time series (DTW and LCSS) and values of the derivative (DTW) can be defined as

$$\begin{aligned} DD_{DTW}^{\mathrm{LCSS}}(x, y) &:= aDTW(x, y) + bDTW(\nabla x, \nabla y) + \\ &+ c\,\mathrm{LCSS}(x, y), \end{aligned} \tag{3}$$

where $a, b, c \in [0, 1]$ are parameters.

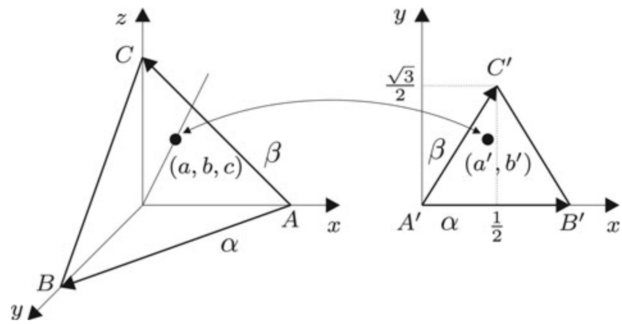

Figure 3. Relationship between the 3D parameters a, b, c and 2D parameters α, β.

2.4. Parameterization

We do not have to check all values of $a, b, c \in [0, 1]$. If $a_1 = ka_2$, $b_1 = kb_2$, $c_1 = kc_2$ where $k > 0$ is a constant (i.e., points (a_1, b_1, c_1), (a_2, b_2, c_2) are linearly dependent), we have

$$DD_{DTW}^{\text{LCSS}}(x_1, y_1; a_1, b_1, c_1) \overset{=}{\lesseqgtr} DD_{DTW}^{\text{LCSS}}(x_2, y_2; a_1, b_1, c_1) \iff$$
$$\iff DD_{DTW}^{\text{LCSS}}(x_1, y_1; a_2, b_2, c_2) \overset{=}{\lesseqgtr} DD_{DTW}^{\text{LCSS}}(x_2, y_2; a_2, b_2, c_2),$$

hence, we can choose parameters (a, b, c) on any continuous surface between the points $A = (1, 0, 0)$, $B = (0, 1, 0)$, and $C = (0, 0, 1)$. For example, this may be the area of the triangle with vertices at those points. This 3D triangle is mapped to the 2D triangle with vertices at the points $A' = (0, 0)$, $B' = (1, 0)$, and $C' = (\frac{1}{2}, \frac{\sqrt{3}}{2})$. Both triangles can be defined in parametrical way:

$$\begin{aligned}(a, b, c) &= A + \alpha\overrightarrow{AB} + \beta\overrightarrow{AC},\\ (a', b') &= A' + \alpha\overrightarrow{A'B'} + \beta\overrightarrow{A'C'},\end{aligned}$$

where (a, b, c) are points of the 3D triangle, (a', b') are points of the 2D triangle, and $\alpha \in [0, 1]$, $\beta \in [0, 1 - \alpha]$ are parameters (Fig. 3).

In the training phase, we have to tune both parameters. We should compute the cross-validation (leave-one-out) error rate on the training data for every pair of parameters and select the pair for the smallest value of the error. However, we can do this in several ways, differing in computational and memory complexity. We decided to use the approach proposed by Górecki and Łuczak (2013). Due to this algorithm, the computation time depends to a small degree on the number of parameters (especially for large values of n). Due to this fact, we can choose a large subset of parameters in the cross-validation process without increasing the computation time of the parameter tuning phase. If the minimal error rate is the same for more than one pair of parameters, we choose the smallest pair—minimizing β first, then α. Finite subsets of parameters α and β are chosen, from 0 to 1 with fixed step 0.01.

3. Experimental setup

We performed experiments on 47 real time series that originate from the UCR Time Series Classification/Clustering Homepage (Keogh et al., 2011). Information on the time series used is presented in Table 1.

The proposed measures are used with the 1NN classifier. We selected this classifier because despite its simplicity, it achieves results that are one of the best compared with other more (sometimes even much more) complex classifiers for time series (Batista et al., 2011;

Table 1. Summary of all datasets used in the experiments. The highest values for each column are marked with • .

Dataset	Number of classes	Size of training	Size of testing	Length
50Words	•50	450	455	270
Adiac	37	390	391	176
Beef	5	30	30	470
Car	4	60	60	577
CBF	3	30	900	128
ChlorineConcentration	3	467	3,840	166
CinC_ECG_torso	4	40	1,380	1,639
Coffee	2	28	28	286
Cricket_X	12	390	390	300
Cricket_Y	12	390	390	300
Cricket_Z	12	390	390	300
DiatomSizeReduction	4	16	306	345
ECG200	2	100	100	96
ECGFiveDays	2	23	861	136
Face (all)	14	560	1,690	131
Face (four)	4	24	88	350
FacesUCR	14	200	2,050	131
Fish	7	175	175	463
Gun-Point	2	50	150	150
Haptics	5	155	308	1,092
InlineSkate	7	100	550	•1,882
ItalyPowerDemand	2	67	1,029	24
Lightning-2	2	60	61	637
Lightning-7	7	70	73	319
Mallat	8	55	2,345	1,024
MedicalImages	10	381	760	99
MoteStrain	2	20	1,252	84
Non-Invasive Thorax1	42	•1,800	1,965	750
Non-Invasive Thorax2	42	•1,800	1,965	750
OliveOil	4	30	30	570
OSU Leaf	6	200	242	427
Plane	7	105	105	144
SonyAIBORobot Surface	2	20	601	70
SonyAIBORobot SurfaceII	2	27	953	65
StarLightCurves	3	1,000	•8,236	1,024
Swedish Leaf	15	500	625	128
Symbols	6	25	995	398
Synthetic Control	6	300	300	60
Trace	4	100	100	275
Two Patterns	4	1,000	4,000	128
TwoLeadECG	2	23	1,139	82
uWaveGestureLibrary_X	8	896	3,582	315
uWaveGestureLibrary_Y	8	896	3,582	315
uWaveGestureLibrary_Z	8	896	3,582	315
Wafer	2	1,000	6,174	152
WordsSynonyms	25	267	638	270
Yoga	2	300	3,000	426

Ding et al., 2008; Radovanović et al., 2010; Tomašev & Mladenić, 2012). Additionally, the 1NN classifier is probably one of the most popular algorithms in data mining (Wu et al., 2008).

For each dataset, we calculated the classification error rate on a test subset. We found all parameters using the training subset. We used the leave-one-out cross-validation (CV) method to find the best parameters. Finite subsets of values of parameters are chosen, ranging from 0 to 1 with fixed step size 0.01.

We set δ to 100%. We selected this value because LCSS is very resistant to changes in δ (Kurbalija et al., 2011). Additionally such choice of δ is consistent with using unconstrained DTW. The calculation of ε is application-dependent. We used a value equal to the smallest

Table 2. Testing error rates (in %) and values of parameters *a*, *b*, and *c* (Eq. (3)). Clearly the best (single winner) results are marked with • .

Dataset	DTW	DD_{DTW}	LCSS	D_{DTW}^{LCSS}	DD_{DTW}^{LCSS}	*a*	*b*	*c*
50Words	30.99	24.62	31.65	23.96	•22.86	0.05	0.44	0.51
Adiac	39.64	29.92	97.19	39.64	29.92	0.17	0.83	0.00
Beef	50.00	43.33	56.67	50.00	43.33	0.20	0.80	0.00
Car	26.67	•20.00	56.67	26.67	23.33	0.01	0.62	0.37
CBF	0.33	0.33	4.00	0.33	0.33	1.00	0.00	0.00
ChlorineConcentration	35.16	29.19	61.54	35.16	•28.93	0.01	0.16	0.83
CinC_ECG_torso	34.93	27.54	7.10	7.10	7.10	0.00	0.03	0.97
Coffee	17.86	10.71	50.00	17.86	10.71	0.02	0.98	0.00
Cricket_X	22.31	21.79	25.90	19.23	19.23	0.04	0.00	0.96
Cricket_Y	20.77	20.77	21.28	•16.67	16.92	0.05	0.10	0.85
Cricket_Z	20.77	21.54	24.36	18.46	•18.21	0.02	0.03	0.95
DiatomSizeReduction	3.27	3.27	69.93	3.27	3.27	1.00	0.00	0.00
ECG200	23.00	17.00	12.00	12.00	15.00	0.01	0.20	0.79
ECGFiveDays	23.23	23.11	•5.69	14.40	14.40	0.02	0.00	0.98
Face (all)	19.23	•9.82	24.38	11.42	11.42	0.03	0.00	0.97
Face (four)	17.05	17.05	18.18	18.18	18.18	0.00	0.00	1.00
FacesUCR	9.51	9.56	10.00	7.37	7.37	0.10	0.00	0.90
Fish	16.57	•5.71	85.14	16.57	6.29	0.02	0.85	0.13
Gun-Point	9.33	2.00	26.67	8.67	2.00	0.02	0.98	0.00
Haptics	62.34	60.06	69.16	60.06	58.12	0.03	0.05	0.92
InlineSkate	61.64	43.82	77.82	62.00	43.82	0.01	0.99	0.00
ItalyPowerDemand	4.96	4.96	20.80	4.96	4.96	1.00	0.00	0.00
Lighting-2	13.11	13.11	18.03	9.84	9.84	0.02	0.00	0.98
Lighting-7	•27.40	32.88	42.47	28.77	28.77	0.03	0.00	0.97
Mallat	6.61	5.12	45.88	5.88	5.12	0.55	0.45	0.00
MedicalImages	26.32	26.32	33.42	26.32	26.32	0.29	0.00	0.71
MoteStrain	16.53	16.69	13.50	•12.38	14.14	0.04	0.05	0.91
Non-Invasive Thorax1	20.97	19.39	85.80	20.51	•17.61	0.17	0.75	0.08
Non-Invasive Thorax2	13.54	•10.69	74.66	13.38	10.89	0.04	0.11	0.85
OliveOil	13.33	13.33	83.33	13.33	13.33	0.64	0.36	0.00
OSU Leaf	40.91	11.98	37.19	38.02	•10.74	0.00	0.78	0.22
Plane	0.00	0.00	19.05	0.00	0.00	1.00	0.00	0.00
SonyAIBORobot Surface	27.45	25.79	29.12	29.12	•24.79	0.00	0.07	0.93
SonyAIBORobot SurfaceII	16.89	10.81	16.37	16.89	10.81	0.01	0.99	0.00
StarLightCurves	9.34	3.78	17.28	9.51	3.67	0.00	0.98	0.02
Swedish Leaf	20.80	•9.92	70.88	20.16	10.08	0.11	0.66	0.23
Symbols	5.03	4.72	20.70	5.03	4.72	0.28	0.72	0.00
Synthetic Control	0.67	0.67	6.00	0.67	0.67	0.19	0.03	0.78
Trace	0.00	0.00	26.00	0.00	0.00	1.00	0.00	0.00
Two Patterns	0.00	0.00	0.08	0.00	0.00	1.00	0.00	0.00
TwoLeadECG	9.57	2.19	48.46	9.57	2.19	0.36	0.64	0.00
uWaveGestureLibrary_X	27.25	22.08	32.69	23.65	•20.27	0.01	0.50	0.49
uWaveGestureLibrary_Y	36.60	28.45	40.82	30.85	•25.63	0.01	0.44	0.55
uWaveGestureLibrary_Z	34.17	30.40	37.52	32.08	•28.31	0.01	0.36	0.63
Wafer	2.01	2.01	0.49	0.49	0.49	0.00	0.00	1.00
WordsSynonyms	35.11	26.96	33.07	25.55	24.61	0.05	0.19	0.76
Yoga	16.37	14.40	40.40	15.10	14.40	0.10	0.90	0.00

standard deviation between the two sequences that were examined at any time (Vlachos et al., 2002).

4. Results

The results are presented in Table 2. The absolute error rates on the test subset with the 1NN classifier are shown for each measure. Additionally, we can observe the values of parameters *a*, *b*, and *c* for each dataset. From this table and Fig. 4, we see that the most important parts in the linear combination are LCSS followed by DD_{DTW}. It seems that most of the information is contained in these two distances. We can suppose that DD_{DTW} aggregates information also

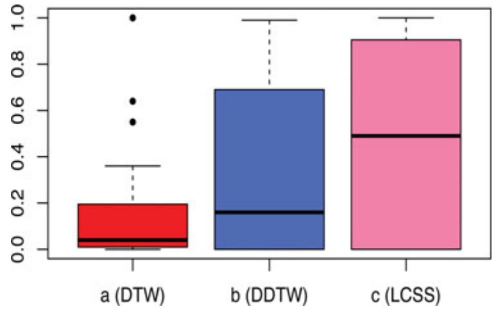

Figure 4. Parameters boxplots.

from pure DTW distance. DTW and LCSS performed clearly the best (single winner, no ties) on 1 of the dataset, DD_{DTW} on 5, D_{DTW}^{LCSS} on 2, and DD_{DTW}^{LCSS} on 12. On 26 datasets no method was clearly better than the others.

Table 3 shows wins, losses, and ties counts (%) for different classifiers (all pairs), across all datasets (in total, we have for each method 188 comparisons). The largest number of wins (125) was produced by DD_{DTW}^{LCSS} method followed by the DD_{DTW} method (102 wins). Note further that the winning method DD_{DTW}^{LCSS} lost only 7% comparisons, while the second DD_{DTW} as much as 23%.

Comparing the new measures with standard DTW and LCSS, we can see a substantial reduction in error rate for most datasets. This is especially clearly seen for the mean of relative errors, presented in Tables 4 and 5. The average relative error reduction for all datasets is equal to 11.52% for D_{DTW}^{LCSS} and 26.41% for DD_{DTW}^{LCSS}, compared with DTW, and, respectively, 41.33% and 48.07% compared with LCSS. A graphical comparison of the methods is presented in Fig. 5. We see that the new method DD_{DTW}^{LCSS} is clearly superior to DD_{DTW} on most of the examined datasets (with a 7.71% average relative error reduction for all datasets). Additionally, we can see that the DD_{DTW}^{LCSS} method outperforms D_{DTW}^{LCSS} (with a 14.98% average relative error reduction for all datasets).

Table 3. Wins, losses, and ties counts (%) for different classifiers, across all datasets. The highest values are marked with symbol •.

	Wins	Losses	Ties
DTW	50 (27%)	98 (52%)	40 (21%)
LCSS	19 (10%)	•161 (86%)	8 (4%)
DD_{DTW}	102 (54%)	44 (23%)	42 (22%)
D_{DTW}^{LCSS}	79 (42%)	58 (31%)	•51 (27%)
DD_{DTW}^{LCSS}	•125 (66%)	14 (7%)	49 (26%)

Table 4. Average relative testing error rates (in%) on all datasets.

	$\frac{DD_{DTW}-DTW}{DTW}$	$\frac{D_{DTW}^{LCSS}-DTW}{DTW}$	$\frac{DD_{DTW}^{LCSS}-DTW}{DTW}$	$\frac{DD_{DTW}-LCSS}{LCSS}$	$\frac{D_{DTW}^{LCSS}-LCSS}{LCSS}$	$\frac{DD_{DTW}^{LCSS}-LCSS}{LCSS}$
MEAN	−19.03	−11.52	−26.41	−28.82	−41.33	−48.07

Table 5. Average relative testing error rates (in%) on all datasets (continued).

	$\frac{D_{DTW}^{LCSS}-DD_{DTW}}{DD_{DTW}}$	$\frac{DD_{DTW}^{LCSS}-DD_{DTW}}{DD_{DTW}}$	$\frac{DD_{DTW}^{LCSS}-D_{DTW}^{LCSS}}{D_{DTW}^{LCSS}}$
MEAN	30.79	−7.71	−14.98

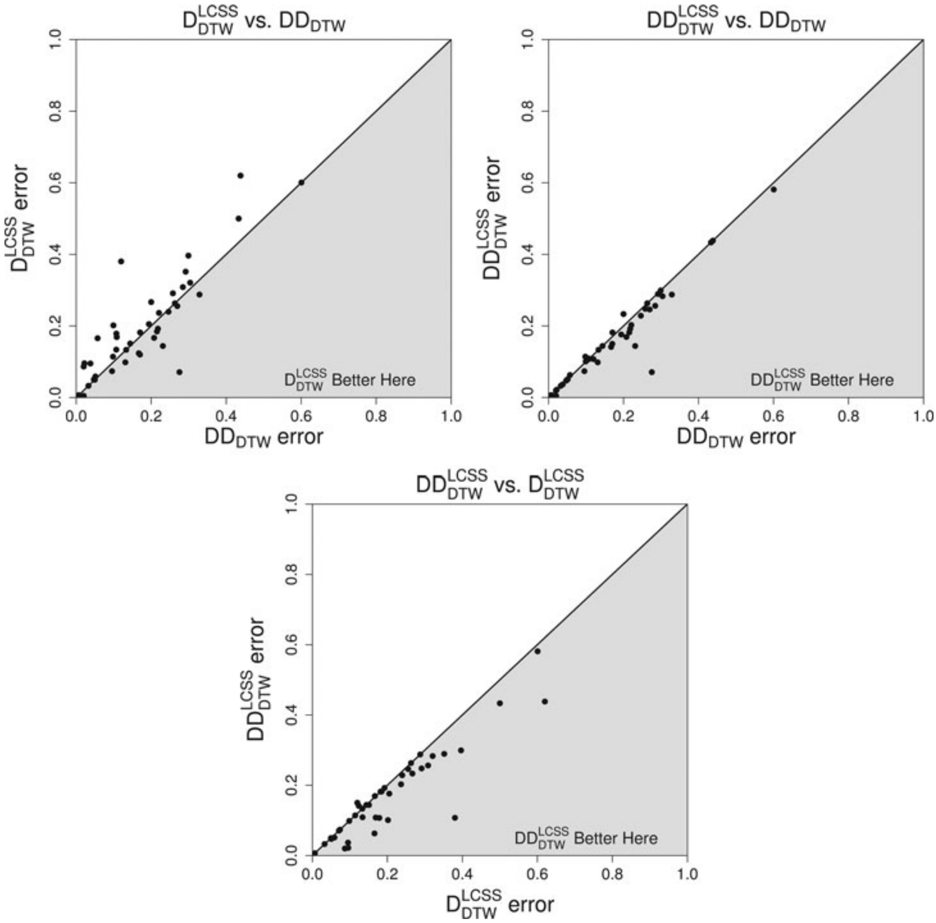

Figure 5. Comparison of test errors.

We can observe that both DTW and LCSS win on single datasets. This could be surprising because they are equivalent to D_{DTW}^{LCSS} with the specific choice of parameters a and b. Such behavior is caused by method of choosing parameters. In these situations, cross-validation method fails to choose suitable parameters.

Of course, the interesting question is that of what components contribution in the final distance measure is perfect. Could we obtain some optimal quantity that determines for all cases that such-and-such participation will give us the best result of classification with 1NN classifier? Figure 6 presents the contribution of each component of the new distance, that is, the value of DTW distance DD_{DTW} distance, and LCSS distance. We can see that the minimum error can be obtained at different points of the parameters triangle, that is, for different values of the parameters a, b, c. Thus, we see that for each dataset and each method, we should select the optimal parameters values independently of the others.

4.1. Statistical comparison of classifiers

To distinguish the methods, we performed a detailed statistical comparison. We tested the hypotheses that are no differences between classifiers. We used the Iman and Davenport's (1980) test, which is a less conservative variant of Friedman's ANOVA test. The p-value from this test is equal to 0. We can therefore proceed with the post hoc tests to detect

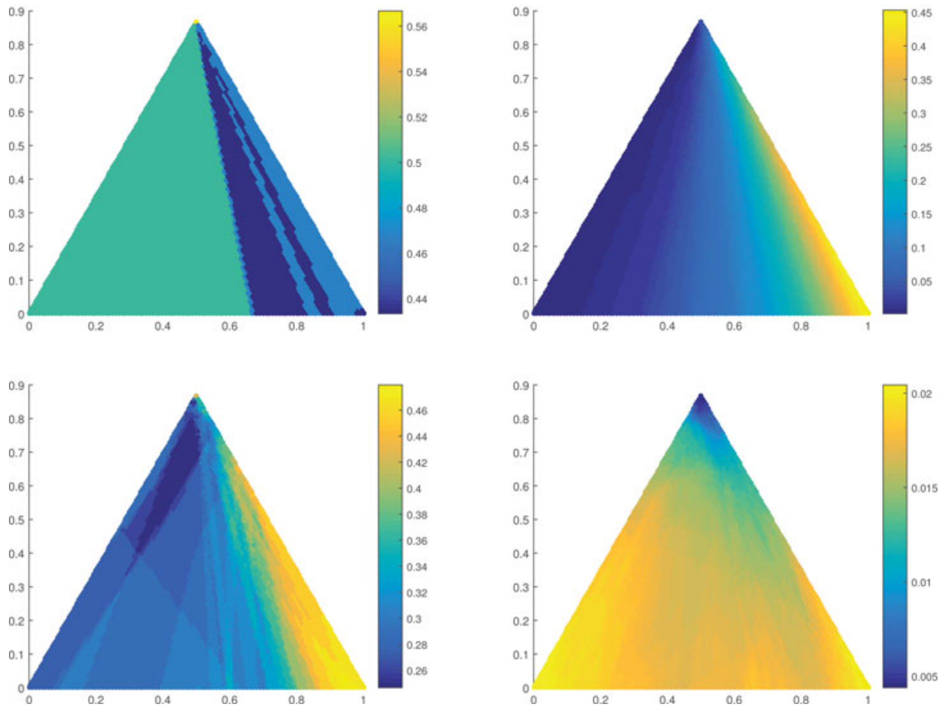

Figure 6. Dependence of classification test error on the participation of DTW distance (bottom-left vertex), DD_{DTW} (bottom-right vertex), and LCSS distance (top vertex) on example datasets. From top left: *beef*, *cbf*, *lighting7*, and *wafer* dataset.

which classifiers are significantly different from each other. Garcia and Herrera (2008) proved that the procedure of Bergmann and Hommel (1988) is the most powerful post hoc comparison test. The results of multiple comparisons are given in Tables 6 and 7. We

Table 6. *p*-Values in the Bergmann–Hommel post hoc test.

i	Hypothesis	*p*-Value
1	LCSS vs. DD_{DTW}^{LCSS}	1.56E-15
2	LCSS vs. DD_{DTW}	4.13E-10
3	LCSS vs. D_{DTW}^{LCSS}	4.23E-7
4	DTW vs. DD_{DTW}^{LCSS}	1.29E-6
5	DTW vs. DD_{DTW}	1.63E-3
6	DTW vs. LCSS	8.67E-3
7	D_{DTW}^{LCSS} vs. DD_{DTW}^{LCSS}	9.99E-3
8	DTW vs. D_{DTW}^{LCSS}	4.88E-2
9	DD_{DTW} vs. DD_{DTW}^{LCSS}	8.38E-2
10	DD_{DTW} vs. D_{DTW}^{LCSS}	2.27E-1

Table 7. Results of the Bergmann–Hommel post hoc test.

Procedure	Ranks mean	Group			
DD_{DTW}^{LCSS}	1.82	a			
D_{DTW}^{LCSS}	2.38		b		
DD_{DTW}	2.77		b		
DTW	3.51			c	
LCSS	4.51				d

finally obtained, at the significance level $\alpha = 0.1$, four homogenous disjoint groups of classifiers: $\{DD_{DTW}^{LCSS}\}$, $\{DD_{DTW}, D_{DTW}^{LCSS}\}$, $\{DTW\}$, and {LCSS}. The best classifier is in the first group.

4.2. A comparison with a state-of-the-art time series classifier

Lines and Bagnall (2014) argue that combined classifier (proportional ensemble) is the best so far classifier for times series. They have compared results to other 11 approaches that they found in the literature, and their method statistically outperforms them. They concluded that such ensemble classifier is significantly more accurate than DTW with window size set through cross-validation (the best so far) or any other elastic distance measure.

We decided to compare this ensemble classifier with our classifier. In all of the analysis below, we discount ECG200 dataset as well as Lines and Bagnall (2014). The dataset ECG200 is removed because it can be perfectly classified with a single rule on the sum of squared values for each series (Lines and Bagnall, 2014). Our method wins on 21 datasets, is tied on 3, and is worse on 22 (Fig. 7). Next we tested the difference between algorithms using a Wilcoxon signed rank test, which one was recommended by Demšar as the best test to compare two classifiers. This test gives a p-value of 0.957. Which means that methods are not significantly different. However, our method has much simpler structure and does not require so many parameters as the proportional ensemble method.

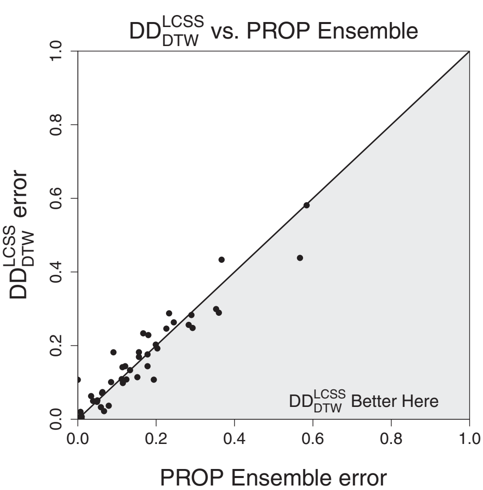

Figure 7. Scatterplot of test error rates of DD_{DTW}^{LCSS} classifier against the proportional ensemble classifier.

5. Conclusions and future work

In this article, we have introduced and studied new time series dissimilarity measures based on derivatives and we used them in classification. We have conducted extensive experiments on the large set of time series classification problems to demonstrate that our methods give very good results. Our measures are superior to LCSS and DTW. Our experiments confirm the power and usefulness of these methods, especially DD_{DTW}^{LCSS}, which is statistically significantly better than other examined methods. Thus, we see that the combination of information coming from different dissimilarity measures leads to a significant improvement in classification results. Proposed measure is also able to cope well with outliers and noise.

Additionally, we have shown that the proposed method is comparable to Lines and Bagnall (2014) method, which was so far considered to be the best classifier for time series. In addition, our method has much simpler structure and does not require so many parameters as the proportional ensemble method.

Since the elastic measures generally provide more accurate classification accuracies compared to non-elastic measures, it would be interesting to examine other elastic measures like Edit Distance with Real Penalty (ERP; Chen and Ng, 2004), Edit Distance on Real sequence (EDR; Chen et al., 2005), which is robust against noise, time shifts, and scaling, Sequence Weighted Alignment model (Swale; Morse and Patel, 2007), and Time Warp Edit Distance (TWED; Marteau, 2009). We would also examine the quality of our methods on constrained versions of elastic measures. The introduction of global constraints significantly speeds-up the computation (Kurbalija et al., 2011, 2014) and in some cases even improves the classification accuracies (Ratanamahatana and Keogh, 2005; Xi et al., 2006). Finally, we can transfer presented methodology to multivariate time series (Górecki and Łuczak, 2015).

Acknowledgments

The authors wish to thank the editor and the two referees for their comments that have helped to improve the presentation of this article.

References

Agrawal, R., Faloutsos, C., Swami, A. (1993). Efficient similarity search in sequence databases. In: Lomet D. B. (ed.,). *4th International Conferenceon Foundations of Data Organization and Algorithms*, Heidelberg: Springer, pp. 69–84.

Agrawal, R., Lin, K., Sawhney, S., Shim, K. (1995). Fast similarity search in the presence of noise, scaling and translation in time-series databases. In: *Proceedings of International Conference on Very Large Databases (VLDB'95)*, Zurich, Switzerland, pp. 490–501.

Batista, G., Wang, X., Keogh, E. (2011) A complexity-invariant distance measure for time series. In: *Proceedings of the Eleventh SIAM Conference on Data Mining (SDM)*. SIAM / Omnipress, pp. 699–710.

Benedikt, L., Kajic, V., Cosker, D., Rosin, P. L., Marshall, D. (2008). Facial dynamics in biometric identification. In: *Proceedings of the British Machine Vision Conference (BMVC'08)*, Leeds, UK, pp. 107.1–107.10.

Bergmann, G., Hommel, G. (1988). Improvements of general multiple test procedures for redundant systems of hypotheses. In: Bauer, P., Hommel, G., Sonnemann, E., eds. *Multiple Hypotheses Testing.* New York: Springer, pp. 110–115.

Berndt, D. J., Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In: *AAAI Workshop on Knowledge Discovery in Databases (KDD'94)*, Seattle, Washington, USA, pp. 229–248.

Chen, L., Ng, R. (2004). On the marriage of Lp-norms and edit distance. In: *Proceedings of the Thirtieth International Conference on Very Large Data Bases* (Vol. 30), pp. 792–803.

Chen, L., Özsu, M. T., Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In: *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, pp. 491–502.

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. In: *Proceedings of 34th International Conference on Very Large Data Bases*, pp. 1542–1552.

Euachongprasit, W., Ratanamahatana, C. A. (2008). Accurate and efficient retrieval of multimedia time series data under uniform scaling and time warping. In: *Proceedings of the 2008 PAKDD (Lecture Notes in Computer Science*, Vol. 5012), pp. 100–111.

Faloutsos, C., Ranganathan, M., Manolopoulos, Y. (1994). Fast subsequence matching in timeseries databases. *ACM SIGMOD Record* 23:419–429.

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.

Górecki, T. (2014). Using derivatives in a longest common subsequence dissimilarity measure for time series classification. *Pattern Recognition Letters* 45C:99–105.

Górecki, T., Łuczak, M. (2013). Using derivatives in time series classification. *Data Mining and Knowledge Discovery* 26(2):310–331.

Górecki, T., Łuczak, M. (2014). First and Second Derivatives in Time Series Classification Using DTW. *Communications in Statistics - Simulation and Computation* 43(9):2081–2092.

Górecki, T., Łuczak, M. (2015). Multivariate time series classification with parametric derivative dynamic time warping. *Expert Systems with Applications* 42(5):2305–2312.

Iman, R., Davenport, J. (1980). Approximations of the critical region of the Friedman statistic. *Communications in Statistics - Theory and Methods* 9(6):571–595.

Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S. (2001). Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems* 3:263–286.

Keogh, E., Pazzani, M. (2001). Dynamic time warping with higher order features. In: *First SIAM International Conferenceon Data Mining (SDM'2001)*, Chicago, USA.

Keogh, E., Zhu, Q., Hu, B., Hao, Y., Xi, X., Wei, L., Ratanamahatana, C. A. (2011). The UCR Time Series Classification/Clustering. Available at: http://www.cs.ucr.edu/eamonn/time_series_data.

Kurbalija, V., Radovanović, M., Geler, Z., Ivanović, M. (2011). The influence of global constraints on DTW and LCS similarity measures for time-series databases. In: *Third International Conference on Software, Services and Semantic Technologies*. Berlin Heidelberg: Springer, pp. 67–74.

Kurbalija, V., Radovanović, M., Geler, Z., Ivanović, M. (2014). The influence of global constraints on similarity measures for time-series databases. *Knowledge-Based Systems* 56:49–67.

Lines, J., Bagnall, A. (2014). Time series classification with ensembles of elastic distance measures. *Data Mining and Knowledge Discovery* 29(3):565–592.

Marteau, P. F. (2009). Time warp edit distance with stiffness adjustment for time series matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(2):306–318.

Micó, L. (1998). Comparison of fast nearest neighbour classifiers for handwritten character recognition. *Pattern Recognition Letters* 19(3–4):351–356.

Morse, M. D., Patel, J. M. (2007). An efficient and accurate method for evaluating time series similarity. *In: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, pp. 569–580.

Paterson, M., Dancik, V. (1994). Longest common subsequences. In: *Proceedings of the 19th MFCS (Lecture Notes in Computer Science*, Vol. 841), pp. 127–142.

Rabiner, L., Juang, B. H. (1993). *Fundamentals of Speech Recognition*. Englewood Cliffs, New Jersey: Prentice Hall.

Radovanović, M., Nanopoulos, A., Ivanović, M. (2010). Time-series classification in many intrinsic dimensions. In: *10th SIAM International Conference on Data Mining (SDM'2010)*, Columbus, USA, pp. 677–688.

Ratanamahatana, C. A., Keogh, E. (2005). Three myths about dynamic time warping data mining. In: *5th SIAM International Conference on Data Mining (SDM'05)*, Newport, USA, pp. 506–510.

Ratanamahatana, C. A., Lin, J., Gunopulos, D., Keogh, E., Vlachos, M., Das, G. (2010). Mining time series data. In: Maimon, O., Rokach, L. (ed.,). New York: Springer-Verlag. *Data Mining and Knowledge Discovery Handbook*, pp. 1049–1077.

Tomašev, N., Mladenić, D. (2012). Nearest neighbor voting in high dimensional data: Learning from past occurrences. *Computer Science and Information Systems* 9:691–712.

Vlachos, M., Gunopoulos, D., Kollios, G. (2002). Discovering similar multidimensional trajectories. In: *Proceedings of the 18th International Conference on Data Engineering*, pp. 673–684.

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.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z., Steinbach, M., Hand, D. J., Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge Information Systems* 14(1):1–37.

Xi, X., Keogh, E., Shelton, C., Wei, L., Ratanamahatana, C. A. (2006). Fast time series classification using numerosity reduction. In: *Proceedings of the 23rd International Conference on Machine Learning*. New York, USA: ACM, pp. 1033–1040.

Yi, B. K. Jagadish, H., Faloutsos, C. (1998). Efficient retrieval of similar time sequences under time warping. In: *Proceedings of the 14th International Conference on Data Engineering (ICDE'98)*, Orlando, USA, pp. 23–27.