

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/319994869>

Univariate and multivariate time series classification with parametric integral dynamic time warping

Article in *Journal of Intelligent and Fuzzy Systems* · September 2017

DOI: 10.3233/JIFS-17523

CITATIONS

0

READS

4

1 author:



Maciej Łuczak

Koszalin University of Technology

16 PUBLICATIONS 62 CITATIONS

SEE PROFILE

Post-print of the final publication:

M. Łuczak, Univariate and multivariate time series classification with parametric integral dynamic time warping, Journal of Intelligent & Fuzzy Systems 33(4) 2017, 2403–2413.

<http://dx.doi.org/10.3233/JIFS-17523>

Univariate and multivariate time series classification with parametric integral dynamic time warping

Maciej Łuczak

Faculty of Civil Engineering, Environmental and Geodetic Sciences, Koszalin University of Technology, Koszalin, Poland

E-mail: mluczak@wilsig.tu.koszalin.pl

Abstract. The dynamic time warping (DTW) distance measure is one of the popular and efficient distance measures used in algorithms of time series classification. It frequently occurs with different kinds of transformations of input data. In this paper we propose a combination of the DTW distance measure with a (discrete) integral transformation. This means that the new distance measure IDTW is simply calculated as the value of DTW on the integrated input time series. However, this design means that the distance cannot in itself give good classification results. We therefore propose to construct a parametric integral dynamic time warping distance measure ID_{DTW} which is a parametric combination of the distances DTW and IDTW. Such a combined distance is used in the nearest neighbor (1NN) classification method in the case of both univariate and multivariate time series analysis. Computational experiments performed on both one-dimensional and multidimensional datasets show that this approach reduces the classification error significantly in comparison with the component methods. The parametric approach allows the new distance to be adapted to each dataset, while showing no significant overfitting effects. The contribution and the main motivation of the paper is to show that the simple transformation as the integral transform can include a bit information about examined time series data and can be used to significantly improve performance of the classification process both for univariate and multivariate time series data. The results are confirmed by graphical and statistical comparisons.

Keywords: time series classification, univariate and multivariate time series data, dynamic time warping, parametric integral distance measure

1. Introduction

Dynamic Time Warping (DTW) is a popular distance measure used in data mining ([4]). The Nearest Neighbor (NN) method with DTW is frequently used in classification and clustering of time series data ([32], [1], [27]). It can be applied to both univariate and multivariate analysis of one-dimensional and multidimensional time series ([31], [28]). The results of classification using the pair NN and DTW are usually very good and hard to surpass by other distance mea-

asures ([13]). The DTW distance often occurs with various kinds of transformations of the input data. For example, computing DTW on the derivative of the raw data is called the Derivative Dynamic Time Warping (DDTW) distance measure ([20]). However, the DTW distance measure on the transformed data rarely gives us a universal distance measure which can compete with DTW on a large number of datasets. Therefore, in order to use the information in both the raw and transformed data, combined methods are used. We

can create a distance measure which is a combination of DTW on raw and transformed time series data. Examples of such combined distances can be found in the literature: a combination of raw and derivative data with the discrete derivative of the first degree for both one-dimensional and multidimensional time series ([14,17]), a combination with a sine/cosine transform ([16]), and others ([30]).

In this paper we examine an integral transformation of time series data. We introduce the **Integral Dynamic Time Warping (IDTW)** distance measure, whose value is computed as the DTW on the integrated input data. This distance measure can be defined for both one-dimensional and multidimensional time series. We then introduce a parametric combination of DTW and IDTW, called the **Parametric Integral Dynamic Time Warping (ID_{DTW})** distance measure. It is a parametric convex combination of the DTW distance measure on the raw and integrated data. The share of each of the distance component is controlled by a single real-valued parameter. Such a parametric distance can be used for both univariate and multivariate time series data. The ID_{DTW} will be used in the classification process with the one nearest neighbor (1NN) method. The parameter of the combination is computed in the learning phase on the training subset by leave-one-out cross-validation.

We perform computational experiments on two large benchmark data bases for one-dimensional and multidimensional time series. The classification error rates obtained on these datasets clearly show that the proposed combined distance ID_{DTW} gives better results than the component distances DTW and IDTW alone. The results are presented by graphical comparison and confirmed statistically. Since the parameter of ID_{DTW} is located outside the distance components, there can be a significant reduction of computation time in the learning phase (cross-validation). We can easily get **lower bounds** for the combined distance, which can also be used to reduce the computational complexity of the method.

The contribution and the main motivation of the paper is to show that the simple transformation as the integral transform can include a bit information about examined time series data and can be used to significantly improve performance of the classification process. On the second hand, we show that only parametrical approach can combine information from raw and transformed data both for univariate and multivariate time series data sets.

The presented method can be used in classification of time series originated from all fields, including price series, complementing other price prediction models ([8,9,10]). It also seems that it may be used in future research with granular computing techniques ([2,26,29,23,33]) to solve univariate and multivariate time series classification problems.

The remainder of the paper is organized as follows. We first (section 2) review the concept of time series and the dynamic time warping algorithm for univariate and multivariate time series data. In the same section we introduce our parametric distance based on integral transformation, and explain the optimization process and properties of the new distance measure. In section 3 the benchmark datasets used in the empirical comparison of methods are described, and we explain the experimental setup. Later in that section we present the results of our experiments on the described time series, as well as statistical analysis of the examined methods. We conclude in section 4 with discussion of possible future extensions of the work.

2. Parametric integral dynamic time warping (ID_{DTW})

A one-dimensional (univariate) time series is a sequence of observations aligned in time or space ([6]). In this paper we shall assume that time series are discrete i.e. they are finite sequences of real numbers:

$$x = \{x(i) \in \mathbb{R} : i = 1, 2, \dots, n\},$$

where $n \in \mathbb{N}$ is the length of the time series.

Multidimensional (multivariate) time series are defined as finite sequence of one-dimensional time series:

$$X = (x_1, x_2, \dots, x_m),$$

where $m \in \mathbb{N}$ is the dimensionality of the multi-series X i.e. the number of variables of the multi-series. In this paper we shall assume that all time series (dimensions, variables) of a multi-series have the same length n for all elements of the dataset.

2.1. Dynamic time warping

The dynamic time warping distance measure (DTW) is a popular distance measure used to calculate the similarity/dissimilarity of time series ([4]). To calculate

the DTW for the two one-dimensional time series with length $n \in \mathbb{N}$

$$\begin{aligned} x &= \{x(i) \in \mathbb{R} : i = 1, 2, \dots, n\}, \\ y &= \{y(i) \in \mathbb{R} : i = 1, 2, \dots, n\} \end{aligned}$$

we proceed as follows. We define a local cost function d — a real-valued function of two real variables which computes a distance between two different points of the time series x and y . For the standard DTW it is usually defined as:

$$d(x(i), y(j)) = (x(i) - y(j))^2. \quad (1)$$

Then we construct a square matrix D with dimensionality $n \times n$ consisting of the local cost function values $D(i, j) = d(x(i), y(j))$. The matrix element $D(i, j)$ corresponds to the alignment between values $x(i)$ and $y(j)$ of the time series. Then we create a warping path $W = \{w_1, w_2, \dots, w_K\}$ ($K \in \mathbb{N}$) with elements of the matrix D . In standard DTW the warping path is required to satisfy three conditions:

1. $w_1 = d(1, 1)$ and $w_K = D(n, n)$ (boundary conditions);
2. if $w_k = D(i_k, j_k)$ and $w_{k+1} = D(i_{k+1}, j_{k+1})$ then $i_{k+1} - i_k \leq 1$ and $j_{k+1} - j_k \leq 1$ (continuity);
3. $i_{k+1} - i_k \geq 0$ and $j_{k+1} - j_k \geq 0$ (monotonicity).

To get a warping path we start at the element $D(1, 1)$ and shifting at most one index forward we finish at the element $D(n, n)$ (Fig. 1). The path which minimizes the warping cost gives the value of the DTW distance measure:

$$\text{DTW}(x, y) = \min_W \left\{ \sum_{k=1}^{K} w_k \right\}. \quad (2)$$

Sometimes the DTW is defined as the square root of (2).

In practice, we calculate the value of DTW by building a cumulative distance matrix Γ by dynamic programming with the following recursion:

$$\begin{aligned} \Gamma(i, j) &= D(i, j) + \\ &\quad \min\{\Gamma(i-1, j-1), \Gamma(i-1, j), \Gamma(i, j-1)\} \end{aligned}$$

with initial conditions:

$$\begin{aligned} \Gamma(0, 0) &= 0, \\ \Gamma(0, i) &= \infty, \quad \Gamma(i, 0) = \infty \quad (i = 1, 2, \dots, n). \end{aligned}$$

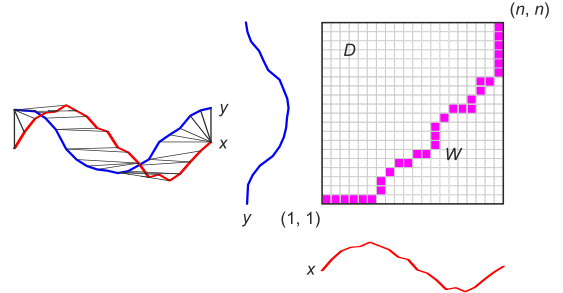


Fig. 1. Time series alignment and the corresponding warping path.

The value of DTW is found at position (n, n) of the matrix Γ :

$$\begin{aligned} \text{DTW}(x, y) &= \Gamma(n, n) \\ \left(\text{DTW}(x, y) &= \sqrt{\Gamma(n, n)} \right). \end{aligned}$$

The DTW distance measure does not satisfy the triangle inequality hence it is not a metric. However, it holds that $\text{DTW}(x, x) = 0$ and $\text{DTW}(x, y) = \text{DTW}(y, x)$ for the cost function (1).

2.2. Integral dynamic time warping

The distance measure computed as DTW on an integrated input time series (indefinite integral of a function) will be called the Integral Dynamic Time Warping distance measure (IDTW):

$$\text{IDTW}(x) = \text{DTW}(I(x)), \quad (3)$$

where $I(x)$ is an integral of the time series x . In the case of discrete time series, integration is simply the cumulative sum. For $y = I(x)$:

$$\begin{aligned} y(1) &= x(1), \\ y(i) &= y(i-1) + x(i), \quad i = 2, 3, \dots, n \end{aligned}$$

or

$$y(i) = \sum_{j=1}^i x(j), \quad i = 1, 2, \dots, n. \quad (4)$$

The metric conditions for IDTW are the same as for DTW.

2.3. Multivariate dynamic time warping

With the assumption that the univariate time series of all dimensions of a multivariate time series have the same length we can view a multi-series as a one-dimensional trajectory in an m -dimensional Euclidean space:

$$X = \{X(i) = (x_1(i), x_2(i), \dots, x_m(i)) \in \mathbb{R}^m : i = 1, 2, \dots, n\}. \quad (5)$$

Then we can define the DTW distance measure between two multi-series X and Y ([17]) in the same way as for the one-dimensional case, but with a local cost function d defined as

$$d(X(i), Y(j)) = \sum_{k=1}^m (x_k(i) - y_k(j))^2, \quad (6)$$

i.e. as the squared Euclidean distance of two m -dimensional vectors formed by the values along the dimensions of the multi-series at positions i and j (Fig. 2).

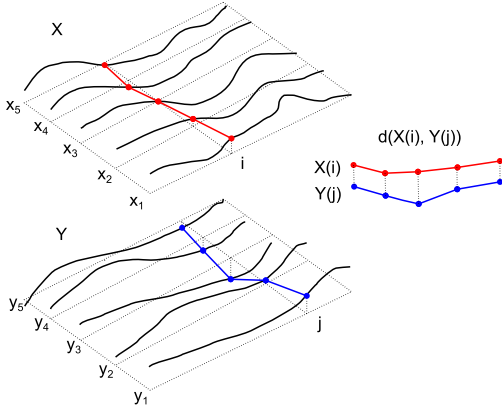


Fig. 2. Multivariate time series alignment and the cost function d .

Similarly to the one-dimensional case, we define the Integral Dynamic Time Warping distance measure as DTW computed on the integrated multi-series $X = (x_1, x_2, \dots, x_m)$, i.e. we integrate each component one-dimensional time series (dimension, variable) separately by formula (4):

$$I(X) = (I(x_1), I(x_2), \dots, I(x_m)),$$

$$\text{IDTW}(X, Y) = \text{DTW}(I(X), I(Y)). \quad (7)$$

The multidimensional IDTW has similar properties as the multidimensional DTW.

2.4. Combining raw and integrated data

Before combining into one distance measure, both raw and integrated data are normalized. The normalized time series will be a new time series denoted $N(x)$:

$$N(x) = \text{z-norm}(x),$$

where z-norm denotes z-normalization by:

$$\text{z-norm}(x) = \frac{x - \mu(x)}{\sigma(x)},$$

where $\mu(x)$ is the mean of the (univariate) time series x , and $\sigma(x)$ is the standard deviation of x .

The normalized multivariate time series will be the series for which every dimensionality is normalized separately:

$$N(X) = (N(x_1), N(x_2), \dots, N(x_m)).$$

Therefore the strict definitions of the distances are:

$$\text{NDTW}(a, b) = \text{DTW}(N(a), N(b)),$$

$$\text{NIDTW}(a, b) = \text{NDTW}(I(N(a)), I(N(b))),$$

where a and b are univariate and multivariate time series respectively. However, to simplify notation, in the rest of the work we shall always use the symbols DTW and IDTW (instead of NDTW and NIDTW) with the assumption that the data is always normalized before computation of the distance measure.

We can now define the parametric integral dynamic time warping distance measure (for both univariate and multivariate time series) as a convex combination of the distances DTW and IDTW:

$$\begin{aligned} \text{ID}_{\text{DTW}}(a, b) = \\ (1 - \alpha) \text{DTW}(a, b) + \alpha \text{IDTW}(a, b), \end{aligned}$$

where a and b are univariate and multivariate time series respectively and α is a real-valued parameter ($\alpha \in [0, 1]$).

The distance function ID_{DTW} can be used in the nearest neighbor classification method, where the pa-

parameter α is tuned in the learning phase. In this work the parameter α will be chosen by leave-one-out cross-validation on the training dataset.

Since α is outside the distances DTW and IDTW, to compute values of ID_{DTW} for all parameters from the interval $[0, 1]$ we need to calculate DTW and IDTW only once. This allows some optimization of the computation time in the learning phase of the NN method. The optimized algorithm for leave-one-out cross-validation on the training dataset is shown in Fig. 3.

2.5. Metric conditions and lower bounds

Since the DTW distance measure is not a metric (it does not satisfy the triangle inequality), the new combined distance measure ID_{DTW} is also not a metric. However, similarly to DTW, the following identities hold:

$$ID_{DTW}(X, X) = 0,$$

$$ID_{DTW}(X, Y) = ID_{DTW}(Y, X),$$

for every fixed parameter $\alpha \in [0, 1]$.

To shorten the calculation time of the NN method with the distance ID_{DTW} we can use a lower bound. If LB is a lower bound for DTW and LBI is a lower bound for IDTW then

$$LB_{\alpha}(a, b) = (1 - \alpha) LB(a, b) + \alpha LBI(a, b) \quad (8)$$

is a lower bound for the distance ID_{DTW} (for every fixed $\alpha \in [0, 1]$ in both the univariate and multivariate cases). There are many good lower bounds of the DTW for one-dimensional time series, such as LB_Keogh ([18]) and LB_Improved ([22]). By our definition of DTW for multidimensional time series (5), (6) we can easily transform those univariate lower bounds to multivariate DTW lower bounds. Then, by (8), we can also find a good lower bounds for ID_{DTW} .

3. Results

3.1. Experimental setup

We conducted computational experiments for both one-dimensional and multidimensional time series data ([19]).

In the univariate case, we performed experiments on 85 data sets from the UCR Time Series Classification

Archive ([7]). This is a database with labeled time series data from a very broad range of fields, including medicine, finance, multimedia and engineering. Each dataset from the database is split into training and testing subsets. All data is z-normalized. Information on the time series used is presented in Table 1.

For the classification process the nearest neighbor method (1NN) is used for all compared distances: DTW, IDTW and ID_{DTW} . We use the cross-validation (leave-one-out) method to find the best parameter α in our classifier ID_{DTW} on a training subset. If the minimal error rate is the same for more than one value of α , we choose the median of those values. A finite subset of the parameter α is chosen, ranging from 0 to 1 with a fixed step size of 0.01.

In the multivariate case, the experiments are carried out on 16 data sets, all of which have labels given. The data sets originate from different domains, including medicine, robotics, handwriting recognition, etc. Information on the time series used is presented in Table 2 (UCI — [3], CMU MOCAP — [11], [5], [21], [25]). Most data is not z-normalized. There is no split into training and testing subsets.

The multivariate time series samples in each dataset are of different lengths. For each dataset, the samples are extended to the length of the longest time series in the data set. We extend all variables of the multidimensional series to the same length. For a short time series instance x with length n , we extend it to a long instance y with length n_{\max} by

$$y(j) = x(i), \quad \text{for } i = \left\lceil \frac{j-1}{n_{\max}-1}(n-1) + 0.5 \right\rceil$$

$$j = 1, 2, \dots, n_{\max}.$$

Some of the values in the sample are duplicated in order to extend a time series. In this way, all the values in the original multivariate time series sample appear in the extended sample.

For the classification process the nearest neighbor method (1NN) is used for all compared distances: DTW, IDTW and ID_{DTW} . For each dataset we calculated the classification error rate using the 10-fold cross-validation method (1NN classifier). We use the cross-validation (leave-one-out) method to find the best parameter α for our classifier ID_{DTW} on a training subset for each split of the 10-fold CV. If the minimal error rate is the same for more than one value of α , we choose the median of those values. A finite subset of the parameter α is chosen, ranging from 0 to 1 with a fixed step size 0.01.

```

% E - time series in the train dataset (cell vector)
% labels - labels of time series (vector)

alpha = 0 : 0.01 : 1;           % alpha parameter subset
n = length(N);                  % number of cross-validation iterations
k = length(alpha);              % number of alpha parameters
mistakes(1 : k) = 0;            % number of false classifications

for i = 1 : n                    % cross-validation iteration loop
    D(1 : k) = inf;              % vector of minimal distances
    L(1 : k) = 0;                % vector of 'minimal' labels
    for j = [1 : i-1, i+1 : n] % leave-one-out
        d = (1 - alpha) * DTW(E{j}, E{i}) + alpha * IDTW(E{j}, E{i}); % combined distance measure
        D(d < D) = d(d < D);    % minimal distances computation
        L(d < D) = labels(j);   % minimal labels computation
    end
    mistakes = mistakes + (L ~= labels(i)); % updating number of false classifications
end
errors = mistakes / n;          % error computation

```

Fig. 3. Implementation of the optimized algorithm (for univariate and multivariate time series) for the leave-one-out cross-validation routine (Matlab code).

3.2. Experimental results

The results for one-dimensional time series are presented in Table 3. The columns of the table (from left) show: the testing error rate (as a percentage) for the component methods DTW and IDTW and for the examined parametric method ID_{DTW} . In Fig. 4 a graphical comparison of the combined method with the component methods is presented.

For statistical confirmation of the performance of ID_{DTW} , p -values were calculated for the Wilcoxon test. To statistically compare two classifiers over multiple data sets, [12] recommends the Wilcoxon signed-ranks test. The Wilcoxon signed-ranks test is a non-parametric alternative to the paired t-test, which ranks the differences in the performances of two classifiers for each dataset, ignoring the signs, and compares the ranks for the positive and the negative differences. The p -values are: 0.0272 for the pair DTW vs. ID_{DTW} and 5.6043×10^{-14} for the pair IDTW vs. ID_{DTW} . This means that the parametric combined distance ID_{DTW} is better than the compared component distances DTW and IDTW at a 97% significance level.

For multidimensional time series data, the results are shown in Table 4. The columns (from left) show the testing error rate (as a percentage) for the component distances DTW and IDTW and for the new parametric combined distance ID_{DTW} . In Fig. 5 a graphical comparison of the examined method with the component methods is shown.

As in the univariate case, we confirm statistically the superior performance of ID_{DTW} by the Wilcoxon

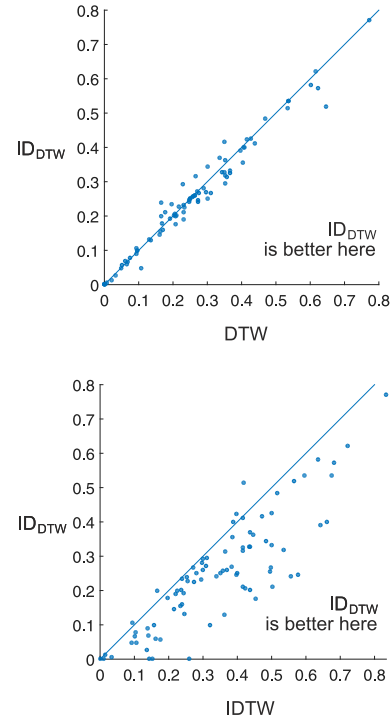


Fig. 4. Graphical comparison of error rates: DTW vs ID_{DTW} and IDTW vs ID_{DTW} for one-dimensional time series data.

test. The calculated p -values are: 0.0703 for the comparison of DTW vs. ID_{DTW} and 0.0012 for IDTW vs. ID_{DTW} . We can see that the combined method ID_{DTW} outperforms the component methods DTW and IDTW at a significance level of 93

Table 1

Characteristics of the one-dimensional datasets used in the experiments (#cl – number of classes, #el – number of elements (size), len – time series length).

dataset	#cl	#el	len	dataset	#cl	#el	len
50Words	50	905	270	MedicalImages	10	1141	99
Adiac	37	781	176	MiddlePhalanxOutlineAgeGroup	3	554	80
ArrowHead	3	211	251	MiddlePhalanxOutlineCorrect	2	891	80
Beef	5	60	470	MiddlePhalanxTW	6	553	80
BeetleFly	2	40	512	MoteStrain	2	1272	84
BirdChicken	2	40	512	NonInvasiveFatalECG Thorax1	42	3765	750
Car	4	120	577	NonInvasiveFatalECG Thorax2	42	3765	750
CBF	3	930	128	OliveOil	4	60	570
ChlorineConcentration	3	4307	166	OSU Leaf	6	442	427
CinC ECG torso	4	1420	1639	PhalangesOutlinesCorrect	2	2658	80
Coffee	2	56	286	Phoneme	39	2110	1024
Computers	2	500	720	ProximalPhalanxOutlineAgeGroup	3	605	80
Cricket X	12	780	300	ProximalPhalanxOutlineCorrect	2	891	80
Cricket Y	12	780	300	ProximalPhalanxTW	6	605	80
Cricket Z	12	780	300	RefrigerationDevices	3	750	720
DiatomSizeReduction	4	322	345	ScreenType	3	750	720
DistalPhalanxOutlineAgeGroup	3	539	80	ShapeletSim	2	200	500
DistalPhalanxOutlineCorrect	2	876	80	ShapesAll	60	1200	512
DistalPhalanxTW	6	539	80	SmallKitchenAppliances	3	750	720
Earthquakes	2	461	512	SonyAIBORobot Surface	2	621	70
ECG200	2	200	96	SonyAIBORobot SurfaceII	2	980	65
ECG5000	5	5000	140	StarLightCurves	3	9236	1024
ECGFiveDays	2	884	136	Strawberry	2	983	235
ElectricDevices	7	16637	96	Swedish Leaf	15	1125	128
FaceAll	14	2250	131	Symbols	6	1020	398
FaceFour	4	112	350	Synthetic Control	6	600	60
Fish	7	350	463	ToeSegmentation1	2	268	277
FordA	3	4921	500	ToeSegmentation2	2	166	343
FordB	3	4446	500	Trace	4	200	275
Gun Point	2	200	150	Two Patterns	4	5000	128
Ham	2	214	431	TwoLeadECG	2	1162	82
HandOutlines	2	1370	2709	uWaveGestureLibrary X	8	4478	315
Haptics	5	463	1092	uWaveGestureLibrary Y	8	4478	315
Herring	2	128	512	uWaveGestureLibrary Z	8	4478	315
InlineSkate	7	650	1882	UWaveGestureLibraryAll	8	4478	945
InsectWingbeatSound	11	2200	256	Wafer	2	7174	152
ItalyPowerDemand	2	1096	24	Wine	2	111	234
LargeKitchenAppliances	3	750	720	WordsSynonyms	25	905	270
Lightning-2	2	121	637	Worms	5	258	900
Lightning-7	7	143	319	WormsTwoClass	2	258	900
MALLAT	8	2400	1024	Yoga	2	3300	426
Meat	3	120	448				

We can present the contribution of the component distances to the combined distance ID_{DTW} . Fig. 6 shows graphs of the parameter α for example univariate time series. The α corresponding to the minimal error rate is different for each data set but we can see that the minimum of the error is well-positioned — there is only one minimum for each error curve. The testing error rate generally corresponds to the cross-validation training error rate, so we can predict quite

well the best value of the parameter α . The combined method adapts well to different data sets without showing signs of overfitting. Similar behavior can be observed for multivariate time series.

Table 2
Summary of multidimensional datasets used in the experiments.

Data sets	Variables	Max length	Min length	Classes	Size	Source
Arabic Digits	13	93	4	10	8800	UCI
Australian Language	22	136	45	95	2565	UCI
BCI	28	500	500	2	416	Blankertz
Character Trajectories	3	205	109	20	2858	UCI
CMU Subject 16	62	580	127	2	58	CMUMC
ECG	2	152	39	2	200	Olszewski
Graz	3	1152	1152	3	140	Leeb
Japanese Vowels	12	29	7	9	640	UCI
Libras	2	45	45	15	360	UCI
Non-Invasive Fetal ECG	2	750	750	42	3765	UCR
Pen Digits	2	8	8	10	10992	UCI
Robot Failure LP1	6	15	15	4	88	UCI
Robot Failure LP2	6	15	15	5	47	UCI
Robot Failure LP3	6	15	15	4	47	UCI
Robot Failure LP4	6	15	15	3	117	UCI
Robot Failure LP5	6	15	15	5	164	UCI
uWaveGestureLibrary	3	315	315	8	4478	UCR
Wafer	6	198	104	2	1194	Olszewski

Table 4
Test errors of the compared methods (in %) for multivariate time series.

Dataset	DTW	IDTW	ID _{DTW}
ArabicDigits	0.22	2.23	0.22
AUSLAN	23.20	47.72	22.11
BCI	46.54	45.22	51.23
CharacterTrajectories	1.50	0.87	0.74
CMUsubject16	0.00	1.67	0.00
ECG	16.00	28.00	17.00
Graz	34.29	46.43	40.71
JapaneseVowels	36.09	22.50	19.84
Libras	18.61	17.50	11.67
NonInvasiveFetalECG Thorax	9.99	21.04	10.01
PenDigits	0.63	1.70	0.62
RobotFailure LP1	28.06	22.64	14.58
RobotFailure LP2	34.00	45.50	28.00
RobotFailure LP3	48.00	33.50	25.50
RobotFailure LP4	16.21	22.88	13.64
RobotFailure LP5	36.54	34.19	33.53
uWaveGestureLibrary	1.90	4.20	1.92
Wafer	3.85	4.18	3.68

4. Conclusions and future work

In this paper we have proposed and examined a parametrical distance measure based on a convex combination of the dynamic time warping distance measure (DTW) and the integral dynamic time warping distance measure (IDTW). The distance IDTW, computed as DTW on the integrated data, when used separately does not give good results in the process of time series classification. However, there exist quite a large group of datasets for which using the IDTW

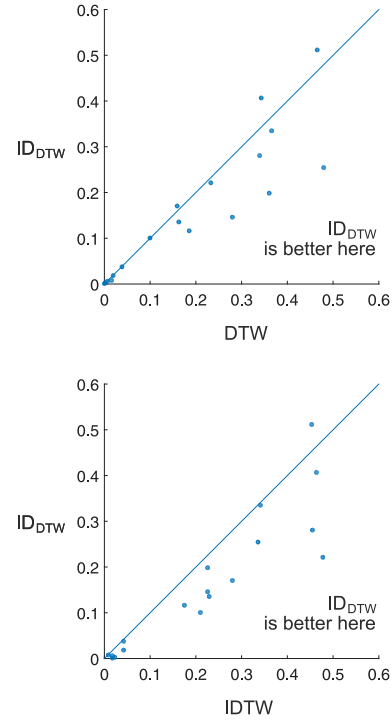


Fig. 5. Graphical comparison of error rates: DTW vs ID_{DTW} and IDTW vs ID_{DTW} for multidimensional time series data.

distance is very favorable. The experiments performed show that the parametrical combination ID_{DTW} outperforms the component distance measures DTW and IDTW in the case of both univariate and multivariate

Table 3
Test errors of the compared methods (in %) for univariate time series.

Dataset	DTW	IDTW	ID _{DTW}	Dataset	DTW	IDTW	ID _{DTW}
50words	30.99	49.67	26.59	MedicalImages	26.32	36.84	26.05
Adiac	39.64	64.19	39.13	MiddlePhalanxOutlineAgeGroup	25.00	28.00	25.00
ArrowHead	29.71	38.29	26.86	MiddlePhalanxOutlineCorrect	35.17	31.00	29.50
Beef	36.67	50.00	33.33	MiddlePhalanxTW	41.60	39.85	42.36
BeetleFly	30.00	40.00	25.00	MoteStrain	16.53	25.24	23.80
BirdChicken	25.00	35.00	25.00	NonInvasiveFatalECG Thorax1	20.97	43.66	20.15
Car	26.67	41.67	31.67	NonInvasiveFatalECG Thorax2	13.54	36.28	12.98
CBF	0.33	1.11	0.11	OliveOil	16.67	16.67	20.00
ChlorineConcentration	35.16	44.69	36.33	OSULeaf	40.91	66.12	40.08
CinC ECG torso	34.93	43.62	32.68	PhalangesOutlinesCorrect	27.16	30.77	27.27
Coffee	0.00	14.29	0.00	Phoneme	77.16	83.44	77.00
Computers	30.00	48.40	34.40	Plane	0.00	15.24	0.00
Cricket X	24.62	57.69	24.62	ProximalPhalanxOutlineAgeGroup	19.51	23.90	23.41
Cricket Y	25.64	49.49	25.64	ProximalPhalanxOutlineCorrect	21.65	25.43	22.68
Cricket Z	24.62	55.64	24.10	ProximalPhalanxTW	26.25	30.00	26.00
DiatomSizeReduction	3.27	13.73	2.61	RefrigerationDevices	53.60	59.47	53.60
DistalPhalanxOutlineAgeGroup	20.75	22.50	20.00	ScreenType	60.27	63.47	58.13
DistalPhalanxOutlineCorrect	23.17	29.50	23.17	ShapeletSim	35.00	47.22	41.67
DistalPhalanxTW	29.00	29.75	28.00	ShapesAll	23.17	41.67	21.17
Earthquakes	25.78	35.71	25.78	SmallKitchenAppliances	35.73	36.53	31.47
ECG200	23.00	22.00	19.00	SonyAIBORobotSurface	27.45	27.12	26.79
ECG5000	7.56	10.29	7.78	SonyAIBORobotSurfaceII	16.89	19.73	17.94
ECGFiveDays	23.23	27.29	22.42	StarLightCurves	9.34	13.82	9.06
ElectricDevices	40.42	38.65	35.52	Strawberry	6.04	14.19	6.85
FaceAll	19.23	24.38	19.29	SwedishLeaf	20.80	45.44	17.60
FaceFour	17.05	23.86	15.91	Symbols	5.03	17.69	5.73
FacesUCR	9.51	15.76	9.90	synthetic control	0.67	3.33	0.67
FISH	17.71	50.29	21.14	ToeSegmentation1	22.81	29.82	29.39
FordA	43.79	41.66	41.21	ToeSegmentation2	16.15	21.54	14.62
FordB	40.59	38.81	40.04	Trace	0.00	26.00	0.00
Gun Point	9.33	9.33	10.67	Two Patterns	0.00	0.00	0.00
Ham	53.33	41.90	51.43	TwoLeadECG	9.57	32.13	9.83
HandOutlines	20.20	23.70	20.10	uWaveGestureLibrary X	27.25	33.84	24.09
Haptics	62.34	68.18	57.14	uWaveGestureLibrary Y	36.60	41.71	32.52
Herring	46.88	51.56	48.44	uWaveGestureLibrary Z	34.17	43.44	32.72
InlineSkate	61.64	72.18	62.18	UWaveGestureLibraryAll	10.83	10.47	4.83
InsectWingbeatSound	64.49	56.46	51.92	wafer	2.01	1.56	1.23
ItalyPowerDemand	4.96	9.14	4.86	Wine	42.59	50.00	42.59
LargeKitchenAppliances	20.53	42.40	20.53	WordsSynonyms	35.11	53.61	31.82
Lighting2	13.11	24.59	13.11	Worms	53.59	67.40	53.59
Lighting7	27.40	39.73	24.66	WormsTwoClass	33.70	43.65	37.02
MALLAT	6.61	16.12	5.97	yoga	16.33	23.33	15.57
Meat	6.67	10.00	6.67				

time series data and significantly reduces the classification error rate. This is confirmed by graphical and statistical comparison. The parametric approach used in the ID_{DTW} method makes it possible to combine the advantages and avoid the disadvantages of the component methods. The new distance can be adapted to individual datasets giving the best classification performance without showing signs of overfitting. We can see a very good correspondence between the values of the distance parameter on the training and testing data

subsets. The method seems to be easily implemented and interpreted. There exist methods of computational complexity reduction resulting both from the existence of lower bounds and directly from the structure of the distance measure ID_{DTW}. The main motivation of the paper is to show that the simple transformation as the integral transform can include a bit information about examined time series data and can be used to significantly improve performance of the classification process.

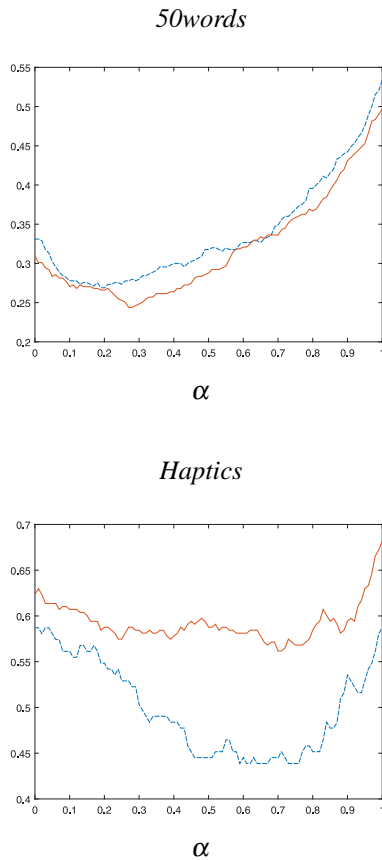


Fig. 6. Correspondence of the parameter α and error rates for the ID_{DTW} method on some univariate datasets. Dashed line: training subset (CV) error; solid line: test subset error.

A disadvantage of the examined method is that its computational complexity is higher than that of the component methods DTW and IDTW. Tuning the distance parameter requires some additional computation in the learning phase. However, in the testing phase, the computational complexity is the same as for the component distances. The new parametric combined method seems to be especially interesting in cases where we have precomputed integral data of the examined time series or where the computation time of integrals is negligible, for example in systems with fast software or even hardware integrators.

Future investigation of the parametric integral dynamic time warping distance measure ID_{DTW} may take several directions. We may try to adapt the method to the unsupervised classification case. Clustering methods often require a different approach (to address certain specific problems) than for supervised methods ([24]). We can also construct a parametrical distance

measure for higher degrees of integrals, similarly as for derivatives in the paper [15]. It may be interesting to examine different definitions of the discrete integral and their influence on the performance of the distance measure. We may also seek further methods of reducing computational complexity in the training phase of classification.

References

- [1] Aghabozorgi, S., Shirkhorshidi, A. S., Wah, T. Y. (2015). Time-series clustering—A decade review, *Information Systems* 53:16–38.
- [2] Ahmad, S. S. S., Pedrycz, W. (2017), The development of granular rule-based systems: a study in structural model compression, *Granular Computing* 2 (1), 1–12.
- [3] Bache, K., Lichman, M. (2013). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- [4] Berndt, D. J., Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *AAAI Workshop on Knowledge Discovery in Databases*: 229–248.
- [5] Blankertz, B., Curio, G., Müller, K. R. (2002). Classifying single trial EEG: towards brain computer interfacing. In: T.G. Diettrich, S. Becker, Z. Ghahramani (Eds.), *Advances in Neural Inf. Proc. Systems* 14 (NIPS 01). Available from: <http://www.bbci.de/competition/ii/>.
- [6] Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (2008). *Time series analysis: Forecasting and control*. Wiley.
- [7] Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., Batista, G. (2015). The UCR Time Series Classification Archive. www.cs.ucr.edu/~eamonn/time_series_data/
- [8] Chen, M. Y., Chen, B. T. (2015). A hybrid fuzzy time series model based on granular computing for stock price forecasting, *Information Sciences* 294, 227–241.
- [9] Chen, M. Y., Chen, B. T. (2014), Online Fuzzy Time Series Analysis Based on Entropy Discretization and a Fast Fourier Transform, *Applied Soft Computing* 14, 156–166.
- [10] Chen, M. Y. (2014) A high-order fuzzy time series forecasting model for Internet stock trading, *Future Generation Computer Systems - The International Journal of eScience* 37, 461–467.
- [11] Carnegie Mellon University Motion Capture Database (2014). Available from: <http://mocap.cs.cmu.edu/>.
- [12] Demšar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research* 7:1–30.

- [13] 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 Proc. 34th Int. Conf. on Very Large Data Bases, 1542–1552.
- [14] Górecki, T., Łuczak, M. (2013). Using derivatives in time series classification. *Data Mining and Knowledge Discovery* 26(2):310–331.
- [15] Górecki, T., Łuczak, M. (2014a). First and second derivative in time series classification using DTW. *Communications in Statistics-Simulation and Computation* 43(9):2081–2092.
- [16] Górecki, T., Łuczak, M. (2014b). Non-isometric transforms in time series classification using DTW. *Knowledge-Based Systems* 61:98–108.
- [17] Górecki, T., Łuczak, M. (2015). Multivariate time series classification with parametric derivative dynamic time warping. *Expert Systems with Applications* 42(5):2305–2312.
- [18] Keogh, E. (2002). Exact indexing of dynamic time warping. In 28th International Conference on Very Large Data Bases, 406–417.
- [19] 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.
- [20] Keogh, E., Pazzani, M. (2001). Dynamic Time Warping with Higher Order Features. In First SIAM International Conference on Data Mining (SDM'2001), Chicago, USA.
- [21] Leeb, R., Lee, F., Keinrath, C., Scherer, R., Bischof, H., Pfurtscheller, G. (2007). Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 15, 473–482. Available from: <http://www.bbci.de/competition/iv/>.
- [22] Lemire, D. (2009). Faster retrieval with a two-pass dynamic-time-warping lower bound. *Pattern Recognition*, 42(9):2169–2180.
- [23] Liu, H., Gegov, A., Cocca, M. (2016). Rule-based systems: A granular computing perspective, *Granular Computing* 1 (4) 259–274.
- [24] Łuczak, M. (2016). Hierarchical clustering of time series data with parametric derivative dynamic time warping. *Expert Systems with Applications*, 62, 116–130.
- [25] Olszewski, R. T. (2001). Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh. Available from: <http://www.cs.cmu.edu/~bobski>.
- [26] Peters, G., Weber, R. (2016) DCC: A framework for dynamic granular clustering, *Granular Computing* 1 (1), 1–11.
- [27] Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., Keogh, E. (2016) Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm. *Knowledge and Information Systems* 47(1), 1–26.
- [28] Shokoohi-Yekta, M., Wang, J., Keogh, E. (2015) On the Non-Trivial Generalization of Dynamic Time Warping to the Multi-Dimensional Case. *Proceedings of the 2015 SIAM International Conference on Data Mining*, pp. 9.
- [29] Skowron, A., Jankowski, A., Dutta, S. (2016) Interactive granular computing, *Granular Computing* 1 (2), 95–113.
- [30] Taylor, J., Zhou, X., Roupail, N. M., Porter, R. J. (2015) Method for investigating intradriver heterogeneity using vehicle trajectory data: A Dynamic Time Warping approach. *Transportation Research Part B: Methodological* 73, 59–80.
- [31] Warrenliao, T. (2005) Clustering of time series data—a survey, *Pattern Recognition* 38(11):1857–1874.
- [32] Xing, Z., Pei, J., Keogh, E. (2010) A brief survey on sequence classification. *ACM SIGKDD Explorations Newsletter* 12 (1), 40–48.
- [33] Yao, Y. (2016). A triarchic theory of granular computing, *Granular Computing* 1 (2) (2016) 145–157.