UDK 519.24

A. V. Goncharov[1], **V. V. Strijov**[2]

ANALYSIS OF DISSIMILARITY SET BETWEEN TIME SERIES *

This paper investigates the metric time series classification problem. Distance functions between time series are constructed using the dynamic time warping method. This method aligns two time series and builds a dissimilarity set. The vector-function of distance between the time series is a set of statistics. It describes the distribution of the dissimilarity set. The object feature description in the classification problem is the set of selected statistics values of the dissimilarity set. It is built between the object and all the reference objects. The additional information about the dissimilarity distribution improves the classification quality. We propose a classification method and demonstrate its result on the classification problem of the human physical activity time series from the mobile phone accelerometer.

Keywords: time series; metric classification; dynamic time warping; distance function.

1. Introduction. This paper investigates the metric time series classification problem [1]. The classification quality depends on the chosen distance function between time series. It constructs feature space for the object description. The dimensionality of metric feature space must be much smaller than the dimensionality of original object space.

We assume the similarity for objects from the same class in the metric classification problem. The vector-function of distance between the time series and other objects is the time series feature description [2]. Its calculations for the time series and all other objects from the sample are computationally expensive. It is calculated only for the time series and reference objects: centroids [3] or cluster centers.

The dynamic time warping DTW method [4, 2] and derivative distance functions based on it deliver better classification quality compared to the other distance functions [5, 6, 7].

[1]Moscow Institute of Physics and Technology. E–mail: alex.goncharov@phystech.edu

[2]Dorodnicyn Computing Center, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences. E–mail: strijov@ccas.ru

*The research was made possible by the Government of the Russian Federation (Agreement No. 05.Y09.21.0018) and RFBR project 16-07-01163.

This method is more stable than the Euclidean distance in the case of time series stretched, compressed, or displaced along the time axis. The best alignment between objects [8] is constructed. Then the Euclidean distance is calculated.

Kernel-DTW [9] improves the quality of the DTW method [8]. This is a method modification that uses different kernels, like Gaussian [10, 11], to calculate the distance between time series values instead of the Euclidean distance. Recent papers [9, 11, 12] propose to calculate the distance as a weighted cost of all possible paths between time series instead of a single warping path.

In the current work the distance function between time series and the feature space construction is a function. It maps from space of object pairs into space of real numbers with dimension different from unity. This allows one to improve the classification quality. The proposed method is applicable to modified versions of the dynamic time warping [8] where dissimilarities are analyzed.

The computational experiment holds on the WISDM data set [13]. The time series were normalized, and centroids were built for each class. The algorithm constructs a warping path for each pair of time series and centroids according to selected constraints and a given kernel. It builds a dissimilarities set using the warping path. Its average value corresponds to the DTW distance, as well as the quantiles of the dissimilarities set distribution, and the mean values of its distribution tail corresponds to the proposed vector-function. Classification methods use the obtained feature matrix. The classification quality while using additional statistics improved compared to the classification quality while using only the mean value.

2. The Problem Formulation. Let $\mathfrak{D} = \{(\mathbf{s}_i, y_i)\}, \quad i \in \mathcal{I} = \{1, \ldots, N\}, \quad \mathbf{s}_i \in \mathbb{R}^n$ be a set of time series and let $y_i \in \mathbb{Y}$ be class labels; $\mathbb{Y}$ is a finite set of class labels. By $C = \{\mathbf{c}_z\}, \quad k = 1, \ldots, |\mathbb{Y}|, \quad \mathbf{c}_k \in \mathbb{R}^n$ denote a set of reference objects. The number of reference objects is equal to the number of different class labels. Each class corresponds to a single reference object. This paper deals with the metric time series classification problem.

Definition 2.1. *We say that a vector distance function between two time series* $\boldsymbol{s}_1$ *and*

$\boldsymbol{s}_2$ *is a map* $\boldsymbol{\rho}$ *from space of time series pairs to space of real numbers with dimension* l*:*

$$\boldsymbol{\rho} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^l.$$

Assign to each object $\mathbf{s}_i$ a feature description $\mathbf{x}_i$. It consists of distances between an original time series $\mathbf{s}_i$ and each of reference objects $\mathbf{c}_z$:

$$\mathbf{x}_i = \Big[\boldsymbol{\rho}(\mathbf{s}_i, \mathbf{c}_1) \quad \vdots \quad \ldots \quad \vdots \quad \boldsymbol{\rho}(\mathbf{s}_i, \mathbf{c}_{|\mathbb{Y}|})\Big]^\mathsf{T}, \quad \mathbf{x}_i \in \mathbb{R}^{l|\mathbb{Y}|},$$

here $\mathbf{a}\vdots\mathbf{b}$ is a vector concatenation.

The classification model parameters are optimized for the object feature description set $\mathfrak{X} = \{(\mathbf{x}_i, y_i)\}, \quad i \in \mathcal{I} = \{1, \ldots, N\}, \quad \mathbf{x}_i \in \mathbb{R}^{l|\mathbb{Y}|}$.

Definition 2.2. *A map from space of object feature description* $\mathbf{x} \in \mathbb{R}^{l|\mathbb{Y}|}$ *and model parameters* $\boldsymbol{w} \in \mathbb{W}$ *to space of class labels is called a classification model and is denoted by* f*:*

$$f : \mathbb{R}^{l|\mathbb{Y}|} \times \mathbb{W} \to \mathbb{Y}.$$

Suppose f is a model from the set $\mathfrak{F} = \{\text{Random Forest}, \text{KNN}, \text{SVM}\}$. Model parameters $\mathbf{w}$ optimize the quality function S, and optimal parameters are denoted by $\mathbf{w}^*$:

$$\mathbf{w}^* = \underset{\mathbf{w} \in \mathbb{W}}{\operatorname{argmin}} S(\mathbf{w}|\mathfrak{X}, f). \tag{1}$$

Using (1) with fixed f and S, we get different classification results for various distance functions $\boldsymbol{\rho}$. By $\mathfrak{R} \ni \boldsymbol{\rho}$ denote the set of distance functions such that $\mathfrak{R}$ includes modifications of the dynamic time warping method: DTW, Kernel-DTW with Gaussian Kernel, and further proposed distance function.

By $\mathfrak{X}_{\boldsymbol{\rho}}$ denote the object feature description such that the distance function $\boldsymbol{\rho}$ is fixed. We have to find the distance function $\boldsymbol{\rho}$ and the classification model f such that they optimize

the quality function S:

$$(\boldsymbol{\rho}^*, f^*) = \underset{(\boldsymbol{\rho},f)\in\mathfrak{R}\times\mathfrak{F}}{\operatorname{argmin}} S(\mathbf{w}^*_{\boldsymbol{\rho}}|\mathfrak{X}_{\boldsymbol{\rho}}, f), \tag{2}$$

where the optimal parameters $\mathbf{w}^*_{\boldsymbol{\rho}}$ are obtained from (1) for fixed $\boldsymbol{\rho}$ and f.

3. DTW Distance Function. Distance functions based on Dynamic Time Warping compared to the other distance functions deliver better quality in the metric classification problem. We propose to improve the classification quality by using the additional information on the basis of dynamic time warping method.

3.1. The Warping Path Construction Definition 3.1.1. *By $\boldsymbol{\Omega}$ denote the dissimilarity matrix between time series $\boldsymbol{s}$ and $\boldsymbol{c}$ such that $\boldsymbol{\Omega}_{ij}$ is equal to $\phi(\boldsymbol{s}_i, \boldsymbol{c}_j), \quad i, j \in \{1, \ldots, n\}$; ϕ is the alignment kernel.*

The DTW method uses the Euclidean kernel as ϕ:

$$\phi(\mathbf{s}_i, \mathbf{c}_j) = (\mathbf{s}_i - \mathbf{c}_j)^2. \tag{3}$$

The Kernel-DTW method uses the Gaussian kernel as ϕ instead of the Euclidean one:

$$\phi(\mathbf{s}_i, \mathbf{c}_j) = e^{\frac{1}{2\pi}(\mathbf{s}_i - \mathbf{c}_j)^2}. \tag{4}$$

Definition 3.1.2. *A path $\boldsymbol{\pi}$ between $\boldsymbol{s}$ and $\boldsymbol{c}$ is an ordered set such that its elements are pairs of the matrix $\boldsymbol{\Omega}$ indexes; $\boldsymbol{\pi}$ belongs to space $(\mathfrak{I} \times \mathfrak{I})^{|\boldsymbol{\pi}|}$:*

$$\boldsymbol{\pi} = \{\pi_r\} = \{(i_r, j_r)\}, \quad r = 1, \ldots, |\boldsymbol{\pi}|, \quad i, j \in \{1, ..., n\},$$

where $|\boldsymbol{\pi}|$ is a chosen path length.

For the path $\boldsymbol{\pi}$ to respond to restrictions on time continuity and physical properties of solving the problem, it is necessary to satisfy the following constraints:

Boundary constraints. *Hold $\pi_1 = (1, 1)$ and $\pi_{|\boldsymbol{\pi}|} = (n, n)$. This means that $\boldsymbol{\pi}$ edges are on the matrix $\boldsymbol{\Omega}$ diagonal on opposite sides.*

Time continuity constraints. *For $\pi_r = (i_1, j_1)$ and $\pi_{r-1} = (i_2, j_2), \quad r = 2, ..., |\boldsymbol{\pi}|$*

$i_1 - i_2 \leq 1, \quad j_1 - j_2 \leq 1$ *hold. In other words* $\boldsymbol{\pi}$ *steps consist of nearby matrix elements.*

Time monotony constraints. *For* $\pi_r = (i_1, j_1)$ *and* $\pi_{r-1} = (i_2, j_2), \quad r = 2, ..., |\boldsymbol{\pi}|$, *at least one of the conditions* $i_1 - i_2 \geq 1, \quad j_1 - j_2 \geq 1$ *hold. This means* $\boldsymbol{\pi}$ *is monotonous.*

Definition 3.1.3. *The path* $\hat{\boldsymbol{\pi}}$ *between time series* $\boldsymbol{s}$ *and* $\boldsymbol{c}$ *is called the optimal path if the constraints defined above and the following condition hold:*

$$\boldsymbol{\pi} = \underset{\boldsymbol{\pi}}{\operatorname{argmin}} \sum_{(i,j)\in\boldsymbol{\pi}} \Omega_{ij}. \tag{5}$$

By $\mathbf{a} : (\mathbf{s}, \mathbf{c}) \mapsto \hat{\boldsymbol{\pi}}$ denote an alignment method between $\mathbf{s}$ and $\mathbf{c}$. It is a map acting into space with non-fixed dimensionality:

$$\mathbf{a} : \quad \mathbb{R}^n \times \mathbb{R}^n \to (\mathfrak{I} \times \mathfrak{I})^{|\boldsymbol{\pi}|}.$$

In the computational experiment time series normalization, additional constraints and edge cutting improve the classification quality.

3.2. Dissimilarities Analysis. Definition 3.2.1. *Suppose* $\boldsymbol{\pi}$ *(3) is an optimal path between the time series* $\boldsymbol{s}$ *and the reference object* $\boldsymbol{c}$*; then* $\{|\Omega_{i_r j_r}|\}, (i_r, j_r) \in \boldsymbol{\pi}$ *is called a dissimilarity set and is denoted by* $\boldsymbol{\delta}$*:*

$$\boldsymbol{\delta} = \{|\Omega_{i_r j_r}|\}, \quad (i_r, j_r) \in \boldsymbol{\pi}.$$

Let the map $\mathbf{d} : (\mathbf{s}, \quad \mathbf{c}, \quad \mathbf{a}(\mathbf{s}, \mathbf{c})) \mapsto \boldsymbol{\delta}$ take each pair of time series $\mathbf{s}, \mathbf{c}$ and the optimal path $\boldsymbol{\pi}$ between them to dissimilarity set $\boldsymbol{\delta}$:

$$\mathbf{d} : \mathbb{R}^n \times \mathbb{R}^n \times (\mathfrak{I} \times \mathfrak{I})^{|\boldsymbol{\pi}|} \to \mathbb{R}^{+|\boldsymbol{\pi}|}.$$

The variational series of $\boldsymbol{\delta}$ describes differences between pairs of time series. The distance function f based on the dynamic time warping method take arguments to space with

dimensionality equal to 1 by averaging $\boldsymbol{\delta}$:

$$\rho(\mathbf{s}, \mathbf{c}) = \frac{\sum_{\delta_i \in \boldsymbol{\delta}} \delta_i}{\sum_{\delta_i \in \boldsymbol{\delta}} 1}. \tag{6}$$

We propose statistics instead of the mean value to describe the $\boldsymbol{\delta}$ distribution: α quantiles of $\boldsymbol{\delta}$ and its averaged subsets such that their elements are greater than the predetermined α quantile. This leads to the classification quality (2) improving in new feature space. By $q_\alpha(\boldsymbol{\delta})$ denote α quantiles and by $p_\alpha(\boldsymbol{\delta})$ denote the averaged subset:

$$p_\alpha(\boldsymbol{\delta}) = \frac{\sum_{r \in \boldsymbol{\delta}} r[r > q_\alpha(\boldsymbol{\delta})]}{\sum_{r \in \boldsymbol{\delta}} [r > q_\alpha(\boldsymbol{\delta})]}.$$

Let the resulting distance function $\boldsymbol{\rho}$ be given by the concatenation of $q_\alpha(\boldsymbol{\delta})$ and $p_\alpha(\boldsymbol{\delta})$ for various α.

Definition 3.2.2 *By δDTW denote the distance function between time series $\boldsymbol{s}$ and $\boldsymbol{c}$ such that the alignment method $\boldsymbol{a}$, map $\boldsymbol{d}$ and set of $\boldsymbol{\alpha} = \{\alpha_i\}, i \in \{1, \ldots, a\}$ are fixed:*

$$\boldsymbol{\rho}(\mathbf{s}, \mathbf{c} | \mathbf{a}, \mathbf{d}) = [q_{\alpha_1}(\boldsymbol{\delta}), \ldots, q_{\alpha_a}(\boldsymbol{\delta}), p_{\alpha_1}(\boldsymbol{\delta}), \ldots, p_{\alpha_a}(\boldsymbol{\delta})]^\mathsf{T}, \quad \boldsymbol{\delta} = \mathbf{d}\big(\mathbf{s}, \mathbf{c}, \mathbf{a}(\mathbf{s}, \mathbf{c})\big). \tag{7}$$

Thus we have the distance function between time series such that it acts into space with dimensionality not equal to 1 and fully describes differences between the time series.

4. Reference Objects. Let reference objects $\mathbf{c}_z$ be centroids of classes.

Definition. *Let $\mathfrak{D}_z$ be a set of objects from $\mathfrak{D}$ such that they have similar class z from $\mathbb{Y}$. The time series is called the centroid of $\mathfrak{D}_z = \{\boldsymbol{s}_i | y_i = z\}, \quad i = 1, \ldots, m$ by the distance ρ (4) and is denoted by $\boldsymbol{c}_z \in \mathbb{R}^n$:*

$$\mathbf{c}_z = \operatorname*{argmin}_{\mathbf{c} \in \mathbb{R}^n} \sum_{\mathbf{s}_i \in \mathfrak{D}_z} \rho(\mathbf{s}_i, \mathbf{c}) = \operatorname*{argmin}_{\mathbf{c} \in \mathbb{R}^n} \sum_{\mathbf{s}_i \in \mathfrak{D}_z} \sum_{(t,t') \in \hat{\boldsymbol{\pi}}_i} \big(\mathbf{s}_i(t') - \mathbf{c}(t)\big)^2, \tag{8}$$

where $\hat{\boldsymbol{\pi}}_i$ is the optimal path between time series $\boldsymbol{s}_i$ and $\boldsymbol{c}$.

Theorem 1. [3] *Let $\mathfrak{D}_z = \{\boldsymbol{s}_i | y_i = z\}_{i=1}^m$ be a set of objects such that they have similar*

class, $\boldsymbol{c}_z$ *be an initial centroid and* $\{\tilde{\boldsymbol{\pi}}_i\}_{i=1}^m$ *be a set of optimal paths between all time series and* $\boldsymbol{c}_z$*; then the local minimum of the optimization task (8) is obtained with*

$$\mathbf{c}_z(t) = \frac{1}{N} \sum_{\mathbf{s}_i \in \mathbb{S}_z} \sum_{t':(t,t') \in \tilde{\boldsymbol{\pi}}_i} \mathbf{s}_i(t'),$$

$$N = \sum_{\mathbf{s}_i \in \mathfrak{D}_z} \sum_{t':(t,t') \in \tilde{\boldsymbol{\pi}}_i} 1.$$

This theorem is the basis of the DBA method [3]: each centroid's element is the averaged elements of time series such that they are aligned with this centroid's element due to the set of optimal paths.

Using Theorem 1, we do not get the optimal centroid. The set of the optimal paths is changed after solving the optimization problem (8). The DBA method is an iteration method such that it calculates the set of optimal paths and the new centroid iteratevely.

5. Computational Experiment. This paper deals with the metric multiclass classification problem. Data is taken from the open dataset WISDM. The time series demonstrate human physical activity. They are collected from the mobile phone accelerometer. The experiment analyzes the proposed distance function δDTW and compares the classification quality for different distance functions $\mathfrak{R} = \{$ DTW, Kernel DTW, δ DTW$\}$.

There are six types of human activity: walking, jogging, skipping, staying, and walking upstairs and downstairs. Time series length is 6 s with 10^{-2}s time step. Each class $z \ni \mathbb{Y}$ consists of 650 time series.

The set of indices $\mathcal{I} = \mathcal{I}^l \cup \mathcal{I}^t$ divides the sample into the learning sample $\mathcal{I}^l$ and the testing one $\mathcal{I}^t$.

Let $\mathbf{c}_z$ be built for each class z using the learning sample. Figure 1 demonstrates these centroids. The object's feature describtion $\mathfrak{X}_{\boldsymbol{\rho}}$ is built with fixed distance function $\boldsymbol{\rho} \in \mathfrak{R}$ for learning and testing samples.

The classification model's parameters $\mathbf{w}$ are optimized (1) on the learning sample. The quality function S is the classification accuracy. The model and the distance function mini-

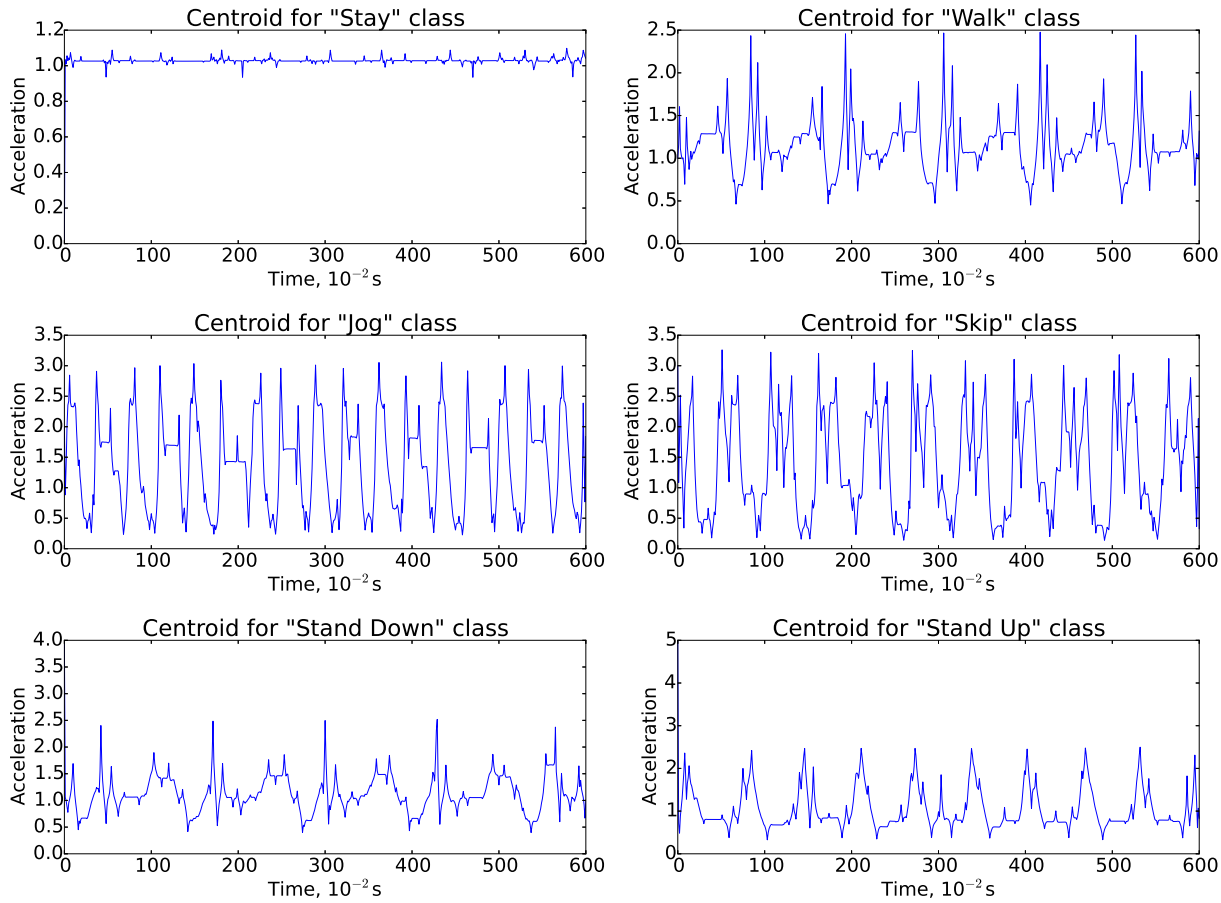

Fig. 1. The centroids for six types of physical human activity.

mize (2) on the testing sample:

$$S(\mathbf{w}_{\boldsymbol{\rho}}|\mathfrak{X}_{\boldsymbol{\rho}}, f) = \frac{\sum_{\mathbf{x}_i}[y_i = f(\mathbf{x}_i, \mathbf{w}_{\boldsymbol{\rho}})][i \in \mathcal{I}^t]}{\sum_{\mathbf{s}_i}[i \in \mathcal{I}^t]}.$$

The experiment compares the results for four different alignment procedures $\mathbf{a}$ between two time series:

1 "Euclidean Kernel." This procedure uses the Euclidean kernel (3) and deals with non-normalized time series.

2 "Gaussian Kernel." This is similar to the "Euclidean kernel" except that it uses the Gaussian kernel (4) instead of the Euclidean one (3). Figures 2 and 3 demonstrate the dissimilarity set $\boldsymbol{\delta}$ between time series and centroid for both kernels.

3 "Gaussian Kernel, Normalized Time Series." This repeats the "Gaussian kernel" but deals with Z-normalized time series. The Z-normalized time series is derived from the original one by applying the following procedure to each element of the sequence:

$$\mathbf{s}'_i = \frac{\mathbf{s}_i - \overline{\mathbf{s}}}{\sqrt{\frac{1}{n}\sum_j (\mathbf{s}_j - \overline{\mathbf{s}})^2}}, \quad \overline{\mathbf{s}} = \frac{1}{n}\sum_j \mathbf{s}_j, \quad i \in \{1, \ldots, n\}.$$

4 "Gaussian Kernel, Normalized Time Series, Cutting Edges." In addition to the "Gaussian kernel, normalized time series," the edges of the constructed path are truncated. The edges of the optimal path add large values because of boundary constraints and influence the dissimilarity set distribution.

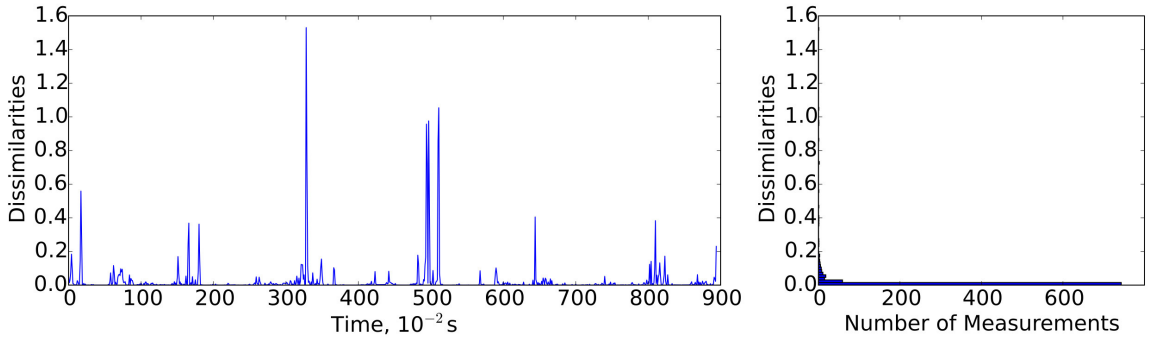

Fig. 2. The dissimilarity set for Euclidean kernel. Left: ordered in time set. Right: distribution histogram.

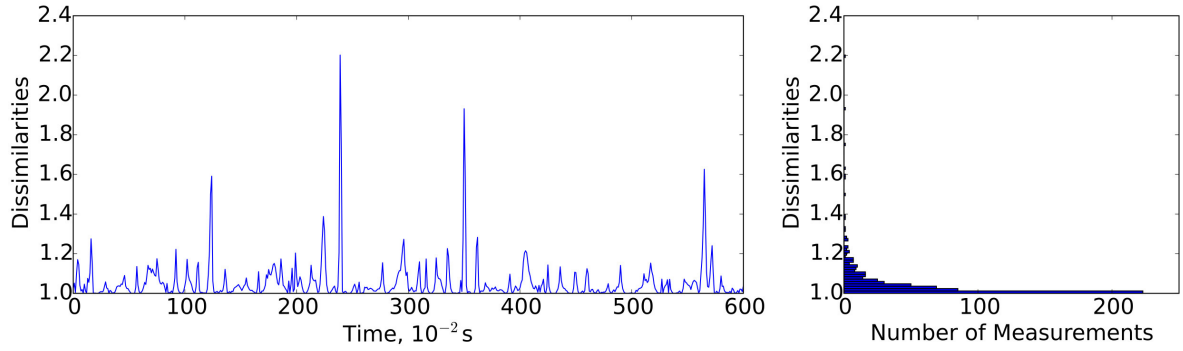

Fig. 3. The dissimilarity set for Gaussian kernel. Left: ordered in time set. Right: distribution histogram.

The Sakoe-Chiba band is further constrained to the global path structure. This reduces

the computational complexity and improves the classification accuracy. For any element (i, j) from path $\boldsymbol{\pi}$ and constant k, $\quad i-k \leq j \leq i+k \quad$ holds. Let the constant k be fixed; then the computational complexity of the construction of the warping path decreases from $O(n^2)$ to $O(n)$.

The experiment compares the classification results for two different ways of dissimilarity set $\boldsymbol{\delta}$ processing:

1 "Mean." The mean value (6) of $\boldsymbol{\delta}$ is the distance function $\boldsymbol{\rho}$ for fixed $\mathbf{a}$ and $\mathbf{d}$.

2 "Heuristic." The δDTW (7) is the distance function $\boldsymbol{\rho}$ for fixed $\mathbf{a}$ and $\mathbf{d}$. The feature space builds for vector $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha} = [0, \ldots, 1.0], \quad \text{with step } 0.1.$$

Table 1 shows the classification results for three different classification models f: random forest, SVM, and KNN.

The combination of the model f random forest, method $\mathbf{a}$ "Gaussian kernel, normalized time series, cutting edges" and the proposed distance function δDTW achieves the best classification accuracy. The linear classifier does not show a high quality. The nearest neighbors classification model with the "mean" method shows better quality. This is explained by the interpretability of such space. Random forest shows high quality. Using the proposed distance function improves the quality of classification: it beats other methods by 7% on average.

6. Conclusion. This work proposes a new approach for building a distance function and dissimilarity analysis to solve the classification problem. The additional information about distribution of this set improves significantly the classification accuracy.

REFERENCES

	Random forest		SVM		KNN	
	Mean	Heuristic	Mean	Heuristic	Mean	Heuristic
Euclidean kernel	0.71	**0.82**	0.55	**0.60**	**0.68**	0.62
Gaussian kernel	0.77	**0.84**	0.39	**0.60**	**0.76**	0.65
Gaussian kernel, Normalized time series	0.78	**0.85**	0.56	**0.60**	**0.78**	0.60
Gaussian kernel, Normalized time series, Cutting edges	0.82	**0.86**	0.56	**0.63**	**0.81**	0.70

Table 1: The classification results

1. A. V. Goncharov, M. S. Popova and V. V. Strijov , "Metric time series classification using dynamic warping relative to centroids of classes," *Systems and Means of Informatics,* **25**. No. 4, 52–64 (2015).

2. E. J. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowl. Inform. Syst.,* **7**. No 3, 358–386 (2005).

3. F. Petitjean, G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen and E. Keogh , "Dynamic time warping averaging of time series allows faster and more accurate classification," *IEEE International Conference on Data Engineering (ICDE),* Chicago, IEEE Computer Society, 470–479 (2014).

4. D. J. Berndt and J. Clifford , "Using dynamic time warping to find patterns in time series," *Workshop on Knowledge Discovery in Databases,* 12th Int'l Conference on Artificial Intelligence, Seattle, 359–370 (1994).

5. E. Frentzos, K. Gratsias and Y. Theodoridis , "Index-based most similar trajectory search," *IEEE International Conference on Data Engineering (ICDE),* Istanbul, IEEE Computer Society, 816–825 (2007).

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

7. Y. Chen, M. A. Nascimento, B. C. Ooi and A. K. H. Tung , "SpADe: On shape-based pattern detection in streaming time series," *IEEE International Conference on Data*

Engineering (ICDE), Istanbul, IEEE Computer Society, 786–795 (2007).

8. S. Salvador and P. Chan , "Fastdtw: Toward accurate dynamic time warping in linear time and space," *Workshop on Mining Temporal and Sequential Data,* Seattle, 70–80 (2004).
9. P.-F. Marteau and S. Gibet , "On recursive edit distance kernels with application to time series classification," *IEEE Transactions on Neural Networks and Learning Systems,* 1–14 (2014).
10. D. Haussler , "Convolution kernels on discrete structures," *Technical Report UCS-CRL-99-10,* University of California at Santa Cruz, Santa Cruz, (1999).
11. B. Scholkopf, A. Smola and K.-R. Muller , "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.,* **10**, No 5, 1299–1319 (1998).
12. M. Cuturi , J.-P. Vert, O. Birkenes and T. Matsui , "A kernel for time series based on global alignments," *In Acoustics, Speech and Signal Processing,* ICASSP 2007, IEEE International Conference on, **2**, 413—416 (2007).
13. Data from accelerometer. Available at: http://sourceforge.net/p/mlalgorithms/TSLearning/data/preprocessed_large.csv (accessed November 15, 2016).