Continuous Time Series Alignment in Human Actions Recognition.

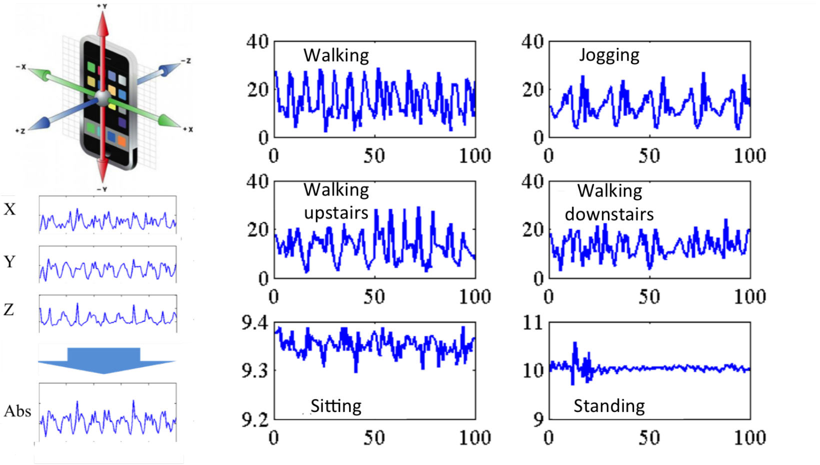
Goncharov Alexey Vladimirovich

Moscow institue of Physics and Technology
Department of intellectual systems, DCAM.
Scientific supervisor: V. V. Strijov.

Conference IOI-2016

Example of scientific tasks

Time series of acceleration from phone accelerometer for different types of human activity



ad

References

1. *Petitjean F., Chen Y., Keogh E.*, ICDE, 2014. — DBA method for centrois calculating.
2. *Goncharov A.V., Strijov V.V.* Systems and applications of informatics, 2015. — reasons for using the DTW distance in the classification task.
3. *Kwapisz J. R.* 2010. — Data using for human activity recognition http://sourceforge.net/p/mlalgorithms/TSLearning/data /preprocessedlarge.csv

Main points

Suggest an efficient way for large multiscale time series analyzing.

Suggestions

- Continuous versions of time series.
- Similar to Dynamic time warping properties.

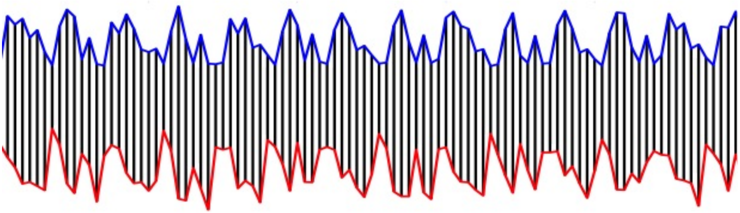
Difficulties

- What the continuous time series are?
- What the distance function between continuous objects is?
- Warping path search can't be solved like in the discrete case.

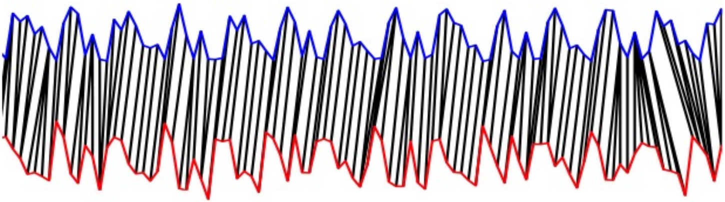
Benefits

- Less memory space for keeping data.
- Ability of working with multiscale time series.
- Theoretical computing of DTW cost.

What the Dynamic time warping is?

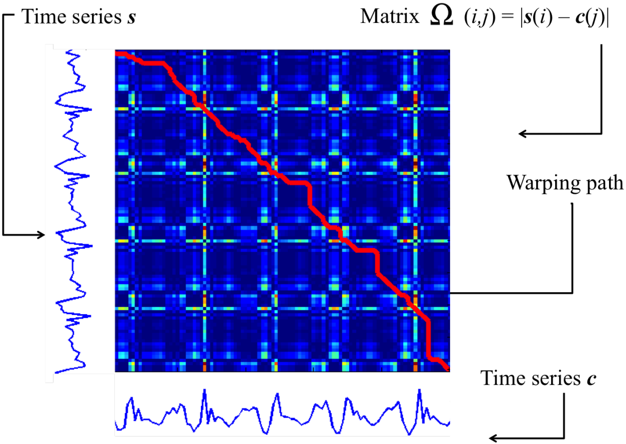


Euclidean distance between time series



Alignment distance between time series

What the alignment path is?

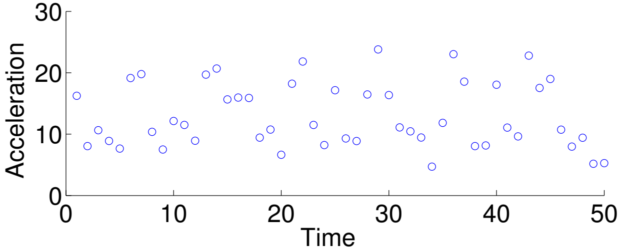


- Dissimilarity matrix for time series elements $\mathbf{\Omega}^{n\times n}: \mathbf{\Omega}(i,j) = |\mathbf{s}(i) - \mathbf{c}(j)|$.
- Path $\boldsymbol{\pi}$ with length K between **s** and **c**: $\boldsymbol{\pi} = \{\pi_k\} = \{(i_k, j_k)\}, \quad k = 1, \dots, K, \quad i, j \in \{1, \dots, n\}$

Time series definition

Definition 1. Discrete case.

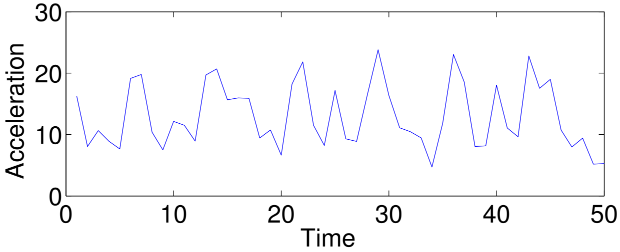
Discrete time series **s** is an oredered in the time sequence $\{s_i\}_{i=1}^{T}$.



Time series definition

Definition 1. Continuous case.

Continuous time series on time plot $\widehat{T} = [0; T]$ is a continuous function $s^c(t) : \widehat{T} \to \mathbb{R}$.



Path definition

Definition 2. Discrete case.

path $\boldsymbol{\pi}$ between two discrete time series $\mathbf{s}_1$ and $\mathbf{s}_2$ is an ordered set of index pairs:

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

and it satisfies the discrete continuity, monotony and the boundary conditions:

$$\pi_r = (p_1, p_2), \quad \pi_{r-1} = (q_1, q_2), \quad r = 2, ..., R, \quad \Rightarrow$$

$$p_1 - q_1 \leq 1, \quad p_2 - q_2 \leq 1,$$

$$\pi_r = (p_1, p_2), \quad \pi_{r-1} = (q_1, q_2), \quad r = 2, ..., R, \quad \Rightarrow$$

$$p_1 - q_1 \geq 1, \quad p_2 - q_2 \geq 1,$$

$$\pi_1 = (1, 1), \quad \pi_R = (n, n).$$

Path definition

Definition 2. Continuous case.

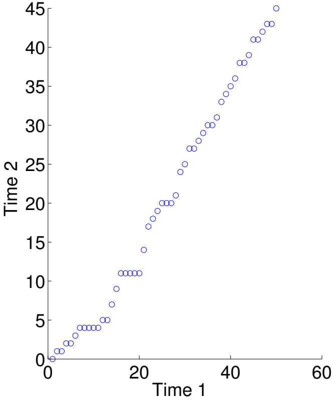
path π^c between two continuous time series $s_1^c(t_1)$ $s_2^c(t_2)$ is a monotonically increasing, continuous function $\pi^c : t_1 \to t_2$ and it satisfies the boundary conditions:

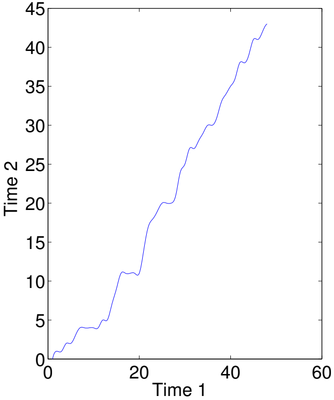
$$\pi^c \in C_{[0;T]},$$

$$t_1 > t_1' \Rightarrow \pi^c(t_1) > \pi^c(t_1'),$$

$$\pi^c(0) = 0, \quad \pi^c(T_1) = T_2.$$

Path definition





Alignment paths in two cases: discrete and continuous

Path cost definition

Definition 3. Discrete case.

the cost $\text{Cost}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\pi})$ of path $\boldsymbol{\pi}$ with length R between two discrete time series $\mathbf{s}_1$ and $\mathbf{s}_2$ is:

$$\text{Cost}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\pi}) = \frac{1}{R} \sum_{(i,j)\in\boldsymbol{\pi}} |\mathbf{s}_1(i) - \mathbf{s}_2(j)|.$$

Definition 3. Continuous case.

the cost $\text{Cost}(s_1^c(t_1), s_2^c(t_2), \pi^c)$ of path π^c between two continuous time series $s_1^c(t_1)$ and $s_2^c(t_2)$ is:

$$\text{Cost}(s_1^c(t_1), s_2^c(t_2), \pi^c) = \frac{1}{L} \int_{t_1} |s_1^c(t_1) - s_2^c(\pi^c(t_1))| dt_1,$$

where L is length of the curve that is given by the graph of the function $\pi^c(t)$, $t \in [0, T]$.

Warping path definition

Definition 4. Discrete case.

warping path $\widehat{\boldsymbol{\pi}}$ between two discrete time series $\mathbf{s}_1$ and $\mathbf{s}_2$ is a path that has the smallest cost among all possible paths:

$$\widehat{\boldsymbol{\pi}} = \operatorname*{argmin}_{\boldsymbol{\pi}} \mathrm{Cost}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\pi}).$$

Definition 4. Continuous case.

warping path $\widehat{\pi}^c$ between two continuous time series $s_1^c(t_1)$ and $s_2^c(t_2)$ is a function $\widehat{\pi}^c$ that has the smallest value of cost from the 3rd definition:

$$\widehat{\pi}^c = \operatorname*{argmin}_{\pi^c} \mathrm{Cost}(s_1^c(t_1), s_2^c(t_2), \pi^c).$$

The cost of warping path

Definition 5. Discrete case.

the cost of warping path or DTW distance between two discrete time series is:

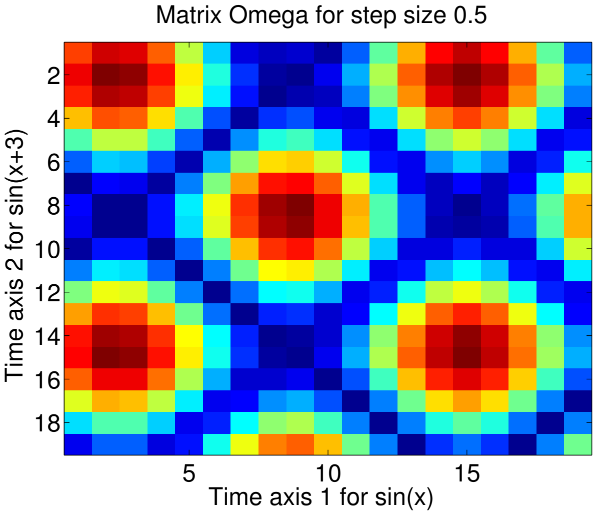
$$\mathrm{DTW}(\mathbf{s}_1, \mathbf{s}_2) = \mathrm{Cost}(\mathbf{s}_1, \mathbf{s}_2, \widehat{\boldsymbol{\pi}}).$$

Definition 5. Continuous case.

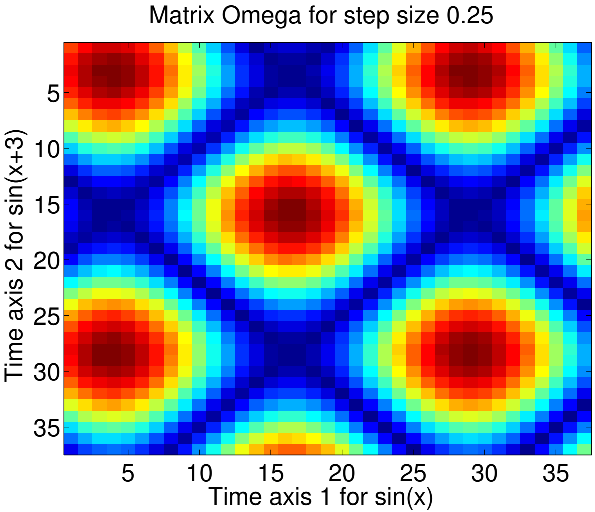
the cost of warping path or DTW distance between two continuous time series is:

$$\mathrm{DTW}(s_1^c(t_1), s_2^c(t_2)) = \mathrm{Cost}(s_1^c(t_1), s_2^c(t_2), \widehat{\pi}^c).$$

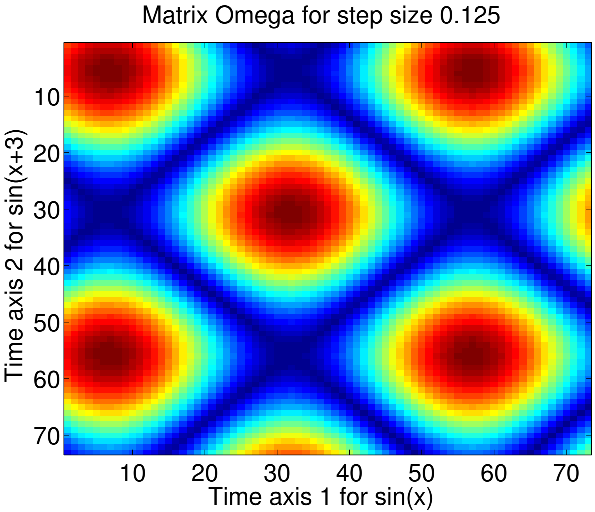
Dissimilarity matrix Omega



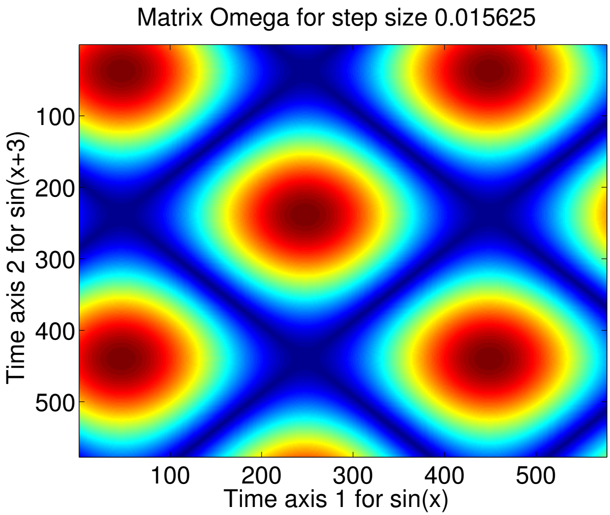
Dissimilarity matrix Omega



Dissimilarity matrix Omega



Dissimilarity matrix Omega



The alignment path properties

Lemma 1.

$s_1(t)$ and $s_2(t)$ are two time series with Lipschitz constant L, $\widehat{\pi}^c : t_1 \to t_2$ is the warping path between them. Its cost does not vary greatly while there are small changes in this path:

$$\| \widehat{\pi}^c - \pi^c \|_C \leq \epsilon \quad \Rightarrow \quad |Cost(s_1, s_2, \widehat{\pi}^c) - Cost(s_1, s_2, \pi^c)| \leq \epsilon TL,$$

where T determines the time boundary for time series, $\epsilon > 0$.

Lemma 2.

$s_1(t)$ and $s_2(t)$ are two time series with Lipschitz constant L, $\widehat{\pi}^c : t_1 \to t_2$ is the warping path between them. Its cost does not vary greatly while there are small changes in one of time series:

$$\| \widehat{s}_2 - s_2 \|_C \leq \epsilon \quad \Rightarrow \quad |Cost(s_1, \widehat{s}_2, \widehat{\pi}^c) - Cost(s_1, s_2, \widehat{\pi}^c)| \leq \epsilon TL,$$

where T determines the time boundary for time series, $\epsilon > 0$.

Assumption 1.

$s_1(t)$ and $s_2(t)$ are two time series with Lipschitz constant L; $\widehat{s}_2(t)$ is a small variation of $s_1(t)$. Then:

$$\text{for all} \quad \epsilon_1 > 0 \quad \text{holds} \quad \epsilon_2(\epsilon_1), \quad \text{for all} \quad \widehat{s}_2(t):$$

$$\| \, \widehat{s}_2(t) - s_2(t) \, \|_c \leq \epsilon_2 \quad \mapsto \quad \| \, \pi^c - \widehat{\pi^c} \, \|_c \quad \leq \quad \epsilon_1,$$

where π^c and $\widehat{\pi^c}$ are the warping paths between $s_1(t), s_2(t)$ and $s_1(t), \widehat{s}_2(t)$ respectively.

Algorithm of building the warping path

The warping path search

- The warping path is a solution of the optimization task from definition:

$$\widehat{\pi}^c = \underset{\pi^c}{\operatorname{argmin}} \operatorname{Cost}(s_1^c(t_1), s_2^c(t_2), \pi^c).$$

- Suggest to search the approximation of this solution among the parametric functions.
- Formulate the problem in the following form:

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}}\, Cost(s_1, s_2, \theta) = \underset{\theta}{\operatorname{argmin}} \int_{t_1} |s_1(t_1) - s_2(F(\theta)(t_1))| dt_1$$

where $F(\theta)$ is a mapping from the parameters to the parametric functions.

Experimental part

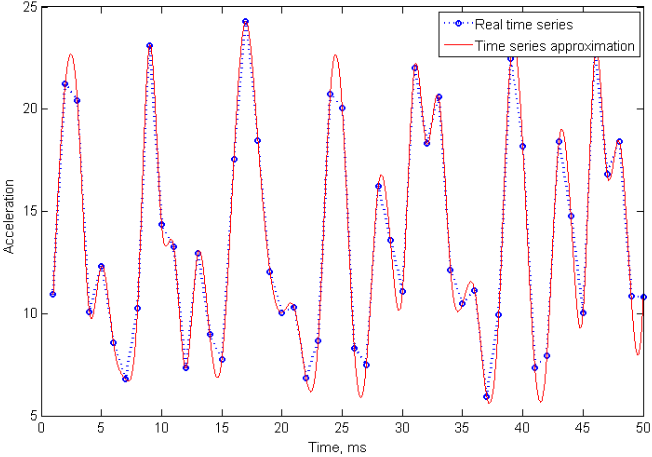
The data is collected the set of time series describing human activity. This set consists of 600 time series, 200 acceleration measurements each. There are six different human activity types.

The experiment plan

- Building centroids with DBA method for each class.
- Building the continuous version for all time series and centroids.
- The DTW distance between all centroids and time series for both cases.

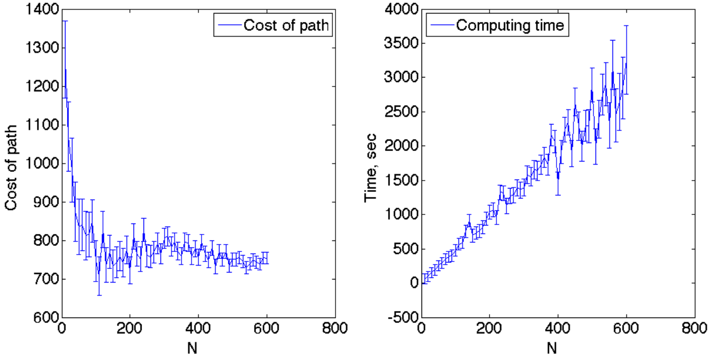
Continuous version of time series

Cubic splines interpolation. One can apply any interpolation or approximation type for getting the continuous object if more accurate method exists.



Distance and computation complexity

The convergence of the method to the real path cost is shown on the left.



The dependencies between number of nodes N, path cost and calculation time

Averaged distance matrix

Table: The mean intraclass values.

	Walk	Run	Up	Down	Sit	Lie
Run	**693**	803	811	733	1165	1143
Walk	676	**498**	696	610	946	927
Up	714	739	**696**	701	1038	1021
Down	591	601	653	**464**	836	804
Sit	516	465	434	400	**6**	42
Lie	508	441	454	366	105	**79**

The results for both algorithms, DTW and DTW in the continious space, gave following results: 85% and 83% relatively. These results don't vary greatly.

Conclusion

- The continuous space usage solves the problem of resampling the data.
- Continuous DTW function has the same properties as the discrete one.

Future plans:

- Use better approximation methods.
- Explore the dependence between approximation method and distance quality.
- Use more effective optimization methods for searching the best path approximation.
- Adapt DTW bounds for continuous space.