Continuous Time Series Alignment in Human Actions Recognition

Alexey V. Goncharov

Moscow Institute of Physics and Technology

Moscow, Russia

Alex.goncharov@phystech.edu

Vadim V. Strijov

Computing Centre of the Russian Academy of Sciences

Moscow, Russia

strijov@ccas.ru

Abstract—**Human physical activity monitoring with wearable devices imposes significant restrictions on the processing power and the amount of memory available to the algorithm. Proposed to move from discrete time series representation to its analytical description and analyze them using mathematical models for satisfying these constraints. The work deals with physical activity classification. It uses metric classification algorithm, where the object's class determined by the distance from this object to the nearest centroid. Paper proposed to approximate all time series with splines and find the distance to the nearest centroid using continuous alignment path. The calculation of distance is performed using analytical transformations.**

I. Introduction

Representing discrete time series with continuous objects is a useful technique in the multiscale time series analysis. Multiscale time series are common, for example, in the medical, industry or even financial applications: human health indications [1], [2], EEG signals [3], human activity [4] or financial data[5] are collected with different frequencies that differ by tens of times [6]. Therefore, such measurements are difficult to analyze using only their discrete representation.

In this paper we present the metric method of time series analysis in continuous space, based on DTW (Dynamic Time Warping) distance measure which performs the dynamic alignment between two time series. This approach lets to simplify the restrictions on the amount of memory by representing the digit objects with their models and on computing time by analytical evaluations.

The DTW distance between two time series has several benefits [7]. It finds the best alignment between two time series if they are nonlinearly deformed relative to each other — they can be stretched, compressed or shifted along the time axis [8].

The DTW defines the distance between discrete time series. Dynamic programming helps to find warping path in discrete case [9]. It is impossible to use the standard DTW method for the continuous space. We introduce the concept of DTW distance function between continuous time series, warping path between them and its cost. We solve the path searching problem by approximating the real path with parametric functions. B-splines or cubic splines [11], for example, can be applied for this approximation. Searching the warping path is equal to searching the best approximation, or the most suitable parameters. The versatility of this approach gives the ability to use any type to approximate the warping path.

This paper introduces a novel approach for analyzing continuous time series. The properties of warping path and its cost are investigated in the experimental part. This part uses discrete time series from accelerometer of mobile phone [12]. Continuous objects are their parametric interpolation. We use cubic spline interpolation. Metric classification task is solved for describing the merging and splitting ability of new distance function in the space of the continuous functions. Centroids of classes are build [13] and classification is made according to the nearest centroid method.

Nowadays DTW distance is the most common distance function and it has many varieties: DTW distance with Sakoe-Chiba [14] band allows to build bounds on the warping path and speed up the computational time, a large number of different upper and lower bounds [15] help to bring the computational complexity of DTW to the Euclidean distance. These techniques make results better but can not be used in the continuous space.

The rest of this paper is organized as follows. Section 2 briefly discusses conversion from discrete objects to continuous. Standard DTW method is described in the 3 section. Section 4 introduces our novel approach, and discusses its properties. The method for searching the warping path is introduced in section 5. Section 6 contains the experimental evaluation of new approach.

II. From discrete object to continuous

Discrete time series $\mathbf{s}$ is an ordered in time sequence. Introduce its definition for continuous time:

Definition 1 *Continuous time series on time plot $\widehat{T} = [0; T]$ is a continuous function $s^c(t) : \widehat{T} \to \mathbb{R}$.*

Let the $\mathcal{S}$ be a space of all discrete time series. And $\mathcal{S}_c$ is a space of all continuous time series. Each discrete object $s \in \mathcal{S}$ associates with continuous analogue $s^c(t) \in \mathcal{S}^c$.

The discrete time series interpolation or approximation with the parametric functions defines the mapping $f : \mathcal{S} \to \mathcal{S}^c$. The algorithm consists of three main parts: choosing the parametric function type, solving the optimization problem to search the optimal parameters. We use cubic spline interpolation in this work [10].

A. Cubic spline interpolation

A set of nodes $\{(x_i, y_i)\}_{i=1}^n$ in $\mathbb{R}^2$ is generated by a continuous and smooth function $f(x)$. In our study x is a time and y

is a measurement. Cubic polynomial interpolates the function between two adjacent dots: (x_i, y_i) and (x_{i+1}, y_{i+1}). Adjacent polynomial values, their derivatives and second derivatives are coincided. The given set of points $\{(x_i, y_i)\}_{i=1}^n$ are spline nodes.

III. DTW DISTANCE

The DTW defines the distance between discrete objects. Define the wapring path, its cost and DTW distance for continuous case.

Suppose, there are two discrete time series $\mathbf{s}_1$ and $\mathbf{s}_2$, and also their continuous analogues, two functions: $s_1^c(t_1)$, and $s_2^c(t_2)$, $t_1, t_2 \in [0; T]$.

Definition 2 (discrete case): *path* π *between two discrete time series $\mathbf{s}_1$ and $\mathbf{s}_2$ is an ordered set of index pairs:*

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

and it satisfies the continuity, monotony and the boundary conditions.

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:*

$$\pi^c \in C_{[0;T]}, \quad \pi^c(0) = 0, \quad \pi^c(T_1) = T_2.$$

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

Definition 3 (discrete case): *the cost $Cost(\mathbf{s}_1, \mathbf{s}_2, \pi)$ of path π with length R between two discrete time series $\mathbf{s}_1$ and $\mathbf{s}_2$ is:*

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

Definition 3 (continuous case): *the cost $Cost(s_1^c(t_1), s_2^c(t_2), \pi^c)$ of path π^c between two continuous time series $s_1^c(t_1)$ and $s_2^c(t_2)$ is:*

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

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

Definition 4 (discrete case): *warping path $\widehat{\pi}$ between two discrete time series $\mathbf{s}_1$ and $\mathbf{s}_2$ is a path that has the smallest cost among all possible paths:*

$$\widehat{\pi} = \underset{\pi}{\arg\min}\, Cost(\mathbf{s}_1, \mathbf{s}_2, \pi).$$

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:*

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

Definition 5 (discrete case): *the cost of warping path or DTW distance between two discrete time series is:*

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

Definition 5 (continuous case): *the cost of warping path or DTW distance between two continuous time series is:*

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

IV. WARPING PATH AND ITS PROPERTIES

This section contains two lemmas. They demonstrate properties of the defined distance function and warping path that can't be observed in the discrete case.

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:*

$$\| \widehat{\pi}^c - \pi^c \|_C \le \epsilon \quad \Rightarrow$$

$$|Cost(s_1, s_2, \widehat{\pi}^c) - Cost(s_1, s_2, \pi^c)| \le \epsilon T L,$$

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

Proof. Write down the chain of inequalities that proves our assumption:

$$|Cost(s_1, s_2, \widehat{\pi}^c) - Cost(s_1, s_2, \pi^c)| =$$

$$| \int_{t_1} (|s_1(t_1) - s_2(\widehat{\pi}^c(t_1))|) dt_1 - \int_{t_1} (|s_1(t_1) - s_2(\pi^c(t_1))|) dt_1 | \le$$

$$\int_{t_1} |s_2(\widehat{\pi}^c(t_1)) - s_2(\pi^c(t_1))| dt_1 \quad \le$$

$$\int_{t_1} L |\widehat{\pi}^c(t_1) - \pi^c(t_1)| dt_1 \quad \le$$

$$\int_{t_1} L \max_{t_1 \in [0,T]} (\widehat{\pi}^c(t_1) - \pi^c(t_1)) dt_1 \quad \le$$

$$\int_{t_1} L \| \widehat{\pi}^c - \pi^c \|_C dt_1 \quad \le \quad T L \epsilon. \quad \square$$

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:*

$$\| \widehat{s}_2 - s_2 \|_C \le \epsilon \quad \Rightarrow$$

$$|Cost(s_1, \widehat{s}_2, \widehat{\pi}^c) - Cost(s_1, s_2, \widehat{\pi}^c)| \le \epsilon T L,$$

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

Proof. Following inequalities demonstrate this statement:

$$|Cost(s_1, \widehat{s}_2, \widehat{\pi}^c) - Cost(s_1, s_2, \widehat{\pi}^c)| =$$

$$| \int_{t_1} (|s_1(t_1) - \widehat{s}_2(\widehat{\pi}^c(t_1))|) dt_1 - \int_{t_1} (|s_1(t_1) - s_2(\widehat{\pi}^c(t_1))|) dt_1 | \le$$

$$\int_{t_1} |\widehat{s}_2(\widehat{\pi}^c(t_1)) - s_2(\widehat{\pi}^c(t_1))| dt_1 \quad \le$$

$$\int_{t_1} \max_{t_1 \in [0,T]} (\widehat{s}_2(\widehat{\pi}^c(t_1)) - s_2(\widehat{\pi}^c(t_1))) dt_1 \quad \le$$

$$\int_{t_1} \| \widehat{s}_2 - s_2 \|_C dt_1 \quad \le \quad T \epsilon. \quad \square$$

These two lemmas demonstrate the robust property for the warping paths cost in the case of small changes in the initial data or in warping path. This paper also puts forward the hypothesis of robust property for the warping path in the case of small changes in the initial data.

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:*

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

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

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.*

V. CALCULATING THE WARPING PATH

Function $\widehat{\pi}^c$: $t_1 \rightarrow t_2$ is a solution of optimization task from the paths definition. The algorithm of building this function in the discrete case uses the dynamic programming. Consider two time series S, C: $S = [s_i]_{i=1}^n$ with length n and $C = [c_i]_{i=1}^m$ with length m. The following constructs the alignment path.

The first stage is building a dissimilarity matrix

$$d \in \mathbb{R}^{n \times m}, \text{where} d_{i,j} = |s_i - c_j|.$$

The second stage is building a transformation matrix $D \in \mathbb{R}^{n \times m}$ using dynamic programming:

$$D_{i,j} = d_{i,j} + \min(D_{i-1,j}, D_{i-1,j-1}, D_{i,j-1}),$$

where $D_{n,m}$ is a DTW distance and warping path is a reverse path from $D_{n,m}$ to $D_{1,1}$ according to the summation direction for each element in path.

This technique can not be applied to the continuous objects since it is impossible to solve the optimization problem in the continuous space. We limit the searching space for $\widehat{\pi}^c$ with the space of parametric functions. In such spaces each set of parameters defines a single object. Define the function $\widehat{\widehat{\pi}}^c$ as an warping path $\widehat{\pi}^c$ approximation. If this approximation is good:

$$\| \widehat{\widehat{\pi}}^c - \widehat{\pi}^c \|_C \leq \epsilon,$$

then cost of this path will not differ greatly from the of real warping path according to the lemma 1. Accept this approximation as a warping path. The searching problem in initial space is reduced to the optimization problem of parameters $\boldsymbol{\theta}$:

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\arg\min} \, Cost(s_1, s_2, \boldsymbol{\theta}) =$$

$$\underset{\boldsymbol{\theta}}{\arg\min} \int_{t_1} |s_1(t_1) - s_2(F(\boldsymbol{\theta})(t_1))| dt_1,$$

where $F(\boldsymbol{\theta})$ maps the parameters into the parametric functions.

This paper suggests the cubic spline approximation for warping path approximation. The nodes number and coordinates (x, y) can vary. Let the number of nodes N be given.

Their coordinates along the axis t_1 are also given. Assume their coordinates along the axis t_2, or $\boldsymbol{\theta} = \{t_{2i}\}_{i=1}^N$, are the optimizing parameters.

The continuity property is satisfied for the path approximation. Introduce boundaries for the optimizing parameters for keeping the monotony and the boundary conditions. Define

$$t_{21} = t_{11} \quad t_{2N} = t_{1N},$$

with

$$t_{2(i)} \leq t_{2(i+1)}, \quad i \in \{1, ..., N-1\}.$$

Formulate the following lemma in the case of this type of approximation.

Lemma 3 *The cost of warping path does not vary greatly in the case of small changes in the parameter vector that defines the path approximation:*

$$\| \widehat{\theta} - \theta \|_2 \leq \epsilon \Rightarrow |Cost(s_1, \widehat{s}_2, F(\widehat{\theta})) - Cost(s_1, s_2, F(\theta))| \leq \delta.$$

Proof: if the coordinate y_i of the spline's node i does not vary greatly, the warping path between two time series will not change a lot according to the norm $\| . \|_C$. Further proof follows from Lemma 1. $\square$

VI. EXPERIMENTAL PART

This part investigates properties of the new distance function. The data is collected the set of time series describing human activity [12].

The computation experiment runs as follows. Split the whole set into six sets according to the human activity type and build a centroid for each set. Demonstrate continuous object creation from discrete one for all objects and centroids. Choose the best number of nodes for warping path approximation and reveal the dependencies between number of nodes, DTW distance and calculation time.

The initial data is a set of discrete time series, which is the ordered list of measurements $\{s_i\}_{i=1}^n$. It is necessary to present them in the $\{(x_i, y_i)\}_{i=1}^n$ form for their future interpolation. Let $x_i = i, \quad y_i = s_i, \quad i = 1, \ldots, n$.

Cubic splines interpolate time series. The example of this interpolation is on the Fig. 1. The point line is a real time series and the continuous line is its smooth cubic interpolation. One can apply any interpolation or approximation type for getting the continuous object if more accurate method exists.

This type of interpolation was chosen because of it's applicability to theoretical calculations of the integral from the Def. 3.

The same type of parametric functions approximates the warping path. Nodes number N is a structure parameter and their coordinates along the first time axis are fixed. Their coordinates along the second time axis are model parameters. If N is small then approximation can't fit the real path and we will overstate the cost of this path. The computation complexity will rise with increasing N. Fig. 2 demonstrates these statements with the experimental evaluations. Distance and computation complexity were measured for one data class
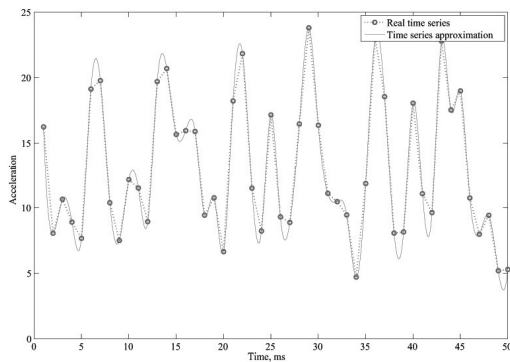
Fig. 1. The example of the cubic spline interpolation

and results were averaged. Fig. 2 includes mean and standard deviation for these measurements. The nodes number was chosen according these results.
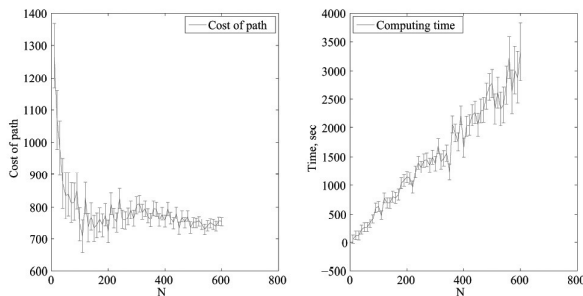

Fig. 2. The dependencies between number of nodes N, path cost and calculation time

We build and average the distance matrix between all time series and centroids for each separate class for investigating distance properties using leave-one-out technique: for each TS we left it out from the sample for building centroids. Mean distance values for each class are shown in the Table I. Each row demonstrates the average distances between one class and all centroids. The smallest value in the row defines the closest centroid's class. The closest centroid for each class has the same label.

TABLE I. THE MEAN INTRACLASS VALUES

	Walk	Run	Up	Down	Sit	Lie
Walk	**693**	803	811	733	1165	1143
Run	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**

Paper uses the nearest centroid method for object classification. The accuracy measuring was simple: one object was dropen from the sample, and centroids for all classes were built without using this object. Then the distance matrix between this dropen time series and all centroids were built. After that this object was given the class of the nearest centroid.

Discrete DTW and continuous DTW gave following results: 85% and 83% accuracy relatively. These results don't

vary greatly. The number of keeping data were halved but the time complexity increased in compare to the effecient DTW realisations. Some of standard for this task techniques beat proposed approach (90 % for neural networks and 89% for using approximation parameters for future classification). But the number of results obtained in the experiment is not enough for its statistical significance.

VII. CONCLUSION

A novel approach in continuous objects analysis was introduced in this paper. It is based on the DTW distance measure between time series and has same properties. It has few additional robust properties. It is universal for applying different approximation types. Bezier splines will be used in future work. Many optimization methods can be used to accelerate the computation time and the quality of path searching. Now this research is in progress. The experiment will be hold on different datasets. This approach will be improved and compared with other techniques for some statistical results.

REFERENCES

[1] J. Hippisley-Cox, Y. Vinogradova and C. Coupland, "Time Series Analysis for selected clinical indicators from the Quality and Outcomes Framework 2001-2006", QRESEARCH and The Information Centre for health and social care, 2007.

[2] R. Palaniappan, "Biological Signal Analysis", *Ventus Publishing*, 2010.

[3] S. Sanei and J.A. Chambers, "EEG Signal Processing", *JohnWiley&Son*, New York, 2007.

[4] R.M. Narayanan, S. Smith and K.A. Gallagher, "A Multifrequency Radar System for Detecting Humans and Characterizing Human Activities for Short-Range Through-Wall and Long-Range Foliage Penetration Applications", *International Journal of Microwave Science and Technology.*, vol. 2014, 2014, p. 21.

[5] D.K. Kilic and O. Ugur, "Multiresolution analysis of S&P500 time series", *Annals of Operations Research.*, 2016, pp. 1-20.

[6] P. Cortez, M. Rio, M. Rocha and P. Sousa, "Multi-scale Internet traffic forecasting using neural networks and time series methods", *ExpertSystems.*, Vol. 29, No. 2., 2012, pp. 143-155.

[7] P. Scheuermann, X. Wang, H. Ding, G. Trajcevski and E. Keogh, "Querying and mining of time series data: experimental comparison of representations and distance measures", *Very Large Data Bases (VLDB)*, 2008, pp.1542-1552.

[8] S. Salvador and P. Chan, "Fastdtw: Toward accurate dynamic time warping in linear time and space", *Workshop on Mining Temporal and Sequential Data*, 2004, p. 11.

[9] E.J. Keogh, M.J. Pazzani, "Scaling up Dynamic Time Warping to Massive Datasets", *Principles of Data Mining and Knowledge Discovery*, Prague: springer Berlin Heidelberg, 1999, pp. 1-11.

[10] P. McKinley and M. Levine, "Cubic Spline Interpolation", 1998, Web: http://msemac.redwoods.edu/ darnold/math45/laproj/Fall98/SkyMeg/Proj.PDF

[11] Carl de Boor, "A Practical Guide to Splines", *Springer-Verlag.*, 1978, pp. 113-114.

[12] Data from accelerometer. Web: http://sourceforge.net/p/mlalgorithms/ TSLearning/data/preprocessed_large.csv. (accessed November 15, 2015).

[13] 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)*, 2014, Chicago, pp. 470-479.

[14] B.B. Alia, Y. Masmoudib and S. Dhouibc, "DTW-Global Constraint Learning Using Tabu Search Algorithm", *Procedia Computer Science.*, Vol. 82, 2016, pp. 12?19.

[15] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria and E. Keogh, "Searching and mining trillions of time series subsequences under dynamic time warping", *KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining.*, 2012, pp. 262-270.