Applied Soft Computing 48 (2016) 650–659

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

# An area-based shape distance measure of time series

Xiao Wang [a], Fusheng Yu [b,*], Witold Pedrycz [c]

[a] School of Information Engineering, Shandong Youth University of Political Science, Jinan 250103, China
[b] School of Mathematical Sciences, Beijing Normal University, Laboratory of Mathematics and Complex Systems, Ministry of Education, Beijing 100875, China
[c] Department of Electrical and Computer Engineering, University of Alberta, Edmonton T6R 2V4 AB, Canada

ARTICLE INFO

*Article history:*
Received 15 February 2016
Received in revised form 22 June 2016
Accepted 22 June 2016
Available online 30 July 2016

*Keywords:*
One-dimensional time series
Area-based shape distance
$l_1$ trend filtering
Dynamic time warping

ABSTRACT

For any two one-dimensional time series of equal or non-equal length, we propose a new method to determine their shape distance. Each of the original time series is represented by a sequence of linear segments which are produced by $l_1$ trend filtering. As the dimensionality of this representation ranges between time series, dynamic time warping (DTW) method is used to calculate the distance between time series. In contrast to the standard dynamic time warping method, here the element of the new distance matrix concerns the distance between two linear segments instead of two elements of the original time series. More specifically, the distance between the two linear segments is calculated as the area of a triangle which is formed by the two linear segments after their translation and connection. In brief, the new measure can be regarded as the dynamic time warping distance computed in a piecewise linear space. Furthermore, we show that new distance measure quantitatively reflects the shape's difference between two one-dimensional time series. The simulation experiments presented in this paper illustrate the performance of the proposed method.



## 1. Introduction

A time series is a sequence of real numbers representing the values of a certain variable over time, and arises in numerous domains, including finance, medicine, business and industry. Typical problems concerning time series include classification [13,17], clustering [6,7,31], similarity search [12,28], forecasting [1,22] and outlier detection [23]. These investigations rely heavily on the ability of a formulation of a suitable distance between time series. This makes studies on distance measures fundamental and timely. As a result, numerous time series distances have been proposed including such well-known examples as, Euclidean distance [5], dynamic time warping distance [2,15], edit distance [3], longest common subsequence distance [27], symbolic aggregate approximation distance [25,26], as well as numerous extensions of these [8,9,13,15,20]. Other works related to the distance measure between two time series also include the feature-based method [11,19,21,30].

Although there are many different distance measures, by far the most common distance measure used in case of two time series of equal length is the Euclidean distance. However, in some practical problems, such as stock time series, people may pay more attention to the similarity concerning the shape of the time series. Specifically, in one-dimensional stock data, the shape reflects trend information which includes not only the time-axis information but also the value-axis information. Since the Euclidean distance cannot completely reflect the shape difference, it may fail when we want to measure the shape difference between two time series.

Additionally, we often encounter pairs of time series of non-equal lengths, which limits the usefulness of the Euclidean distance. In order to handle this problem, dynamic time warping (DTW) algorithm [2] was proposed by Berndt and Clifford to eliminate misalignment and to express the distance between time series by ignoring both global and local shifts in time. As a result, DTW algorithm gives rise to a DTW distance and helps realize the matching process. Dynamic time warping exhibits two weaknesses. First, DTW may ignore capturing an overall shape of the time series. In other words, this pure value-oriented comparison makes the value-based DTW ignore the context of points, such as their positions in local features and their relations in terms of possible overall trends. Furthermore, DTW has a quadratic time and space complexity that limits its use to only small-size (short) time series. In order to reduce computing overhead and space complexity, some constraints are imposed on the warping route. In the existing literature [31], Yu et al. proposed granular dynamic time warping (GDTW) on

* Corresponding author.
*E-mail address:* yufusheng@bnu.edu.cn (F. Yu).

http://dx.doi.org/10.1016/j.asoc.2016.06.033
1568-4946/

granular time series which can greatly improve computing speed. But, in the GDTW method the original time series is segmented within equal-length granular windows and the obtained granules cannot reflect the trend information of any granular window. Hence, GDTW is also a pure value-oriented measure by ignoring corresponding timeline. The other shape-based distance measures, such as [18,29], are based on discrete symbolic representation, which also require the pair of time series to be of equal length and segment the pair of time series alignment.

Taking into account the above drawbacks, we present an accurate shape distance measure between two one-dimensional time series. Before calculating their distance, we do some preprocessing by mapping the original time series into the space of their piecewise linear approximation. Firstly, we use $l_1$ trend filtering [16] to determine trends presented in a given time series. Then, we connect some key points such as kinks, knots, or changes in slope induced by $l_1$ trend filtering forming a piecewise linear representation (PLR) of the original time series. The PLR method can extract main features of time series and at the same time effectively reduce data dimensionality [10]. After $l_1$ trend filtering and PLR, the original time series is represented by a sequence of linear segments of non-equal lengths, and two linear segments coming from the different transformed time series are usually out of alignment on the time-axis. Obviously, the trend information of the original time series is preserved in all the linear segments. Before calculating the distance between any two time series, one has to decide on how to accurately measure the distance between any two linear segments originating from different time series. In order to speed up the calculation and fully take into account the time-axis information of these linear segments, we construct a triangle through translation and connection and consider the area to measure their distance. Let us note that an area can reflect the time-axis and value-axis information. In the sequel, we employ DTW algorithm in the piecewise linear space to form the final DTW distance of two given time series.

To investigate the aforementioned problem carefully, this paper is organized as follows: we devote Section 2 to a discussion of all necessary prerequisites such as $l_1$ trend filtering and dynamic time warping algorithm. The original area-based shape distance measure is presented in details, and the area-based shape distance algorithm is proposed in Section 3. Following that, a series of experiments are performed in Section 4 to illustrate the performance of the area-based shape distance algorithm. Finally, some conclusions are made in Section 5.

For convenience, some notations used in the later sections are introduced as follows:

$X=\{(t_1,v_1),(t_2,v_2),\ldots,(t_m,v_m)\}$: one-dimensional time series of length $m$, where $t_i$ and $v_i$ denote time-axis and value-axis coordinates of the $i$th element of $X$, respectively, $t_i<t_{i+1}$, $v_i\in R^1$ and $m$ is a positive integer $(i=1,b,\ldots,m)$;
$\boldsymbol{x}=(x_1,x_2,\ldots,x_m)$: one-dimensional time series of length $m$, where $x_t$ denotes the value-axis information at time $t$ $(t=1,2,\ldots,m)$, $\boldsymbol{x}$ can be briefly described as $\boldsymbol{x}=\{x_t\}_{t=1}^m$;
$[1:N]:[1:N]=\{1,2,\ldots,N\}$;
$a\wedge b: a\wedge b=\min\{a,b\}$;
$|D|$: the determinant of matrix $D$;
$|X|:|X|=\sum_{i=1}^m t_i/m$, if $X=\{(t_1,v_1),(t_2,v_2),\ldots,(t_m,v_m)\}$;
$\hat{X}:\hat{X}=\sum_{i=1}^m v_i/m$, if $X=\{(t_1,v_1),(t_2,v_2),\ldots,(t_m,v_m)\}$;
$\|\boldsymbol{u}\|_1$: the $L_1$ norm of vector $\boldsymbol{u}=(u_1,u_2,\ldots,u_n)$, i.e., $\|\boldsymbol{u}\|_1=\sum_i|u_i|$;
$\|\boldsymbol{u}\|_2$: the $L_2$ norm of vector $\boldsymbol{u}=(u_1,u_2,\ldots,u_n)$, i.e., $\|\boldsymbol{u}\|_2=\sqrt{\sum_i|u_i^2|}$;
$\bar{L}=\langle(t_L,v_L),(t_R,v_R)\rangle$: the linear segment $\bar{L}$, where $(t_L,v_L)$ and $(t_R,v_R)$ represent the left and right endpoint of $\bar{L}$ in the two-dimensional plane, respectively.

## 2. Preliminaries

In this section, we briefly recall $l_1$ trend filtering [16], and introduce dynamic time warping algorithm [2] employed in our method.

### 2.1. $l_1$ trend filtering

Given is a one-dimensional time series $\boldsymbol{y}=\{y_t\}_{t=1}^n$ of length $n$ which consists of an underlying trend component $\boldsymbol{x}=\{x_t\}_{t=1}^n$ and a random component $\boldsymbol{z}=\{z_t\}_{t=1}^n$. Our goal is to obtain the trend estimate $\boldsymbol{x}$ by minimizing the objective function in the form

$$\min \quad Q(\boldsymbol{x})=\frac{1}{2}\sum_{t=1}^{n}(y_t-x_t)^2+\lambda\sum_{t=2}^{n-1}|x_{t-1}-2x_t+x_{t+1}| \tag{1}$$

where $\lambda$ is a nonnegative parameter used to control a trade-off between the smoothness of $\boldsymbol{x}$ and the size of the residual $\boldsymbol{z}$ $(\boldsymbol{z}=\boldsymbol{y}-\boldsymbol{x})$. The first term in the objective function (1) expresses the residual $\boldsymbol{y}-\boldsymbol{x}$; while the second term quantifies the smoothness of the estimated trend $\boldsymbol{x}$. In the second term, $x_{t-1}-2x_t+x_{t+1}$ is the second order difference of the time series $\boldsymbol{x}$ computed at time $t$; it is zero if and only if the three points $x_{t-1}, x_t, x_{t+1}$ are linear.

The objective function (1) can be rewritten in the matrix form as:

$$\min \quad Q(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2+\lambda\|D\boldsymbol{x}\|_1 \tag{2}$$

where $D$ is the second-order difference matrix:

$$D=\begin{bmatrix}1 & -2 & 1 & 0 & \cdots & 0 & 0\\ 0 & 1 & -2 & 1 & \cdots & 0 & 0\\ & & \ddots & \ddots & \ddots & & \\ 0 & 0 & \cdots & 1 & -2 & 1 & 0\\ 0 & 0 & \cdots & 0 & 1 & -2 & 1\end{bmatrix}.$$

The objective function (2) is strictly convex and coercive in $\boldsymbol{x}$, so it has a unique minimizer $\boldsymbol{x}^{lt}$. (The superscript ‘lt’ stands for ‘$l_1$ trend’.)

Let $\boldsymbol{x}=A\boldsymbol{\theta}$; then the objective function (2) is transformed into a $l_1$-regularized least squares problem:

$$\min \quad Q(\boldsymbol{\theta})=\|A\boldsymbol{\theta}-\boldsymbol{y}\|_2^2+\lambda\sum_{i=3}^{n}|\theta_i| \tag{3}$$

where $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_n)\in R^n$ is a vector of some variables and $A$ is a lower triangular matrix:

$$A=\begin{bmatrix}1 & 0 & 0 & 0 & \cdots & 0\\ 1 & 1 & 0 & 0 & \cdots & 0\\ 1 & 2 & 1 & 0 & \cdots & 0\\ 1 & 3 & 2 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \ddots & \\ 1 & n-1 & n-2 & \cdots & 2 & 1\end{bmatrix}.$$

The solution $\boldsymbol{\theta}^{lt}$ to the optimization problem (3) and the $l_1$ trend estimate $\boldsymbol{x}^{lt}$ are related by

$$\boldsymbol{x}^{lt}=A\boldsymbol{\theta}^{lt}. \tag{4}$$

From the standard result in $l_1$-regularized least squares [4,24], the solution $\boldsymbol{\theta}^{lt}$ to the optimization problem (3) is a piecewise linear function of the regularization parameter $\lambda$. Combining (4), the regularization path of $l_1$ trend filtering is also piecewise linear. More specifically, there are $p$ integer time points $t_i$ satisfying the condition $1=t_1<t_2<\ldots<t_p=n$, whereas $x_t^{lt}$ is an affine function in the $i$th time window $[t_i, t_{i+1}]$, i.e. $x_t^{lt}$ $(t\in[t_i,t_{i+1}])$ can be expressed as a linear segment $\bar{X}_i=\langle(t_i,x_{t_i}),(t_{i+1},x_{t_{i+1}})\rangle$ $(i=1,2,\ldots,p-1)$.