Tracking Recurring Patterns in Time Series Using Dynamic Time Warping

Rik van der Vlist
Quby B.V.
Delft University of Technology
The Netherlands
rikvandervlist@gmail.com

Cees Taal
Quby B.V.
The Netherlands
chtaal@gmail.com

Richard Heusdens
Delft University of technology
The Netherlands
r.heusdens@tudelft.nl

***Abstract*—Dynamic time warping (DTW) is a distance measure to compare time series that exhibit similar patterns. In this paper, we will show how the *warping path* of the DTW algorithm can be interpreted, and a framework is proposed to extend the DTW algorithm. Using this framework, we will show how the dynamic programming structure of the DTW algorithm can be used to track repeating patterns in time series.**

***Index Terms*—dynamic programming, dynamic time warping, time series analysis**

I. Introduction

Many recorded signals exhibit some kind of cyclic behavior where the signal values are repeated according to a certain pattern, but where the frequency of the pattern is time-varying. For example in [1], household power consumption is estimated by optically sensing a rotating disc with a photoreflective sensor, where the rotational speed of the disc is proportional to the power consumption. As a consequence, the reflective pattern of one full rotation is frequency-modulated by this consumption. Another example can be found in the field of remaining life estimation in rotating equipment where vibration measurements are used of rolling element bearings [2]. A defect in one of the bearing raceways will induce a clear pattern in the vibration signal, where the rotational speed of the bearing will modulate its frequency. Also, times series with time-varying frequency include the repeating pulse signal in ECG heart-rate data where its underlying frequency can be altered by different causes [3].

Typically, the fluctuation of the frequency of these type of signals may vary fast over time. As a consequence, conventional frequency estimation techniques based on (short-time) Fourier analysis may not be applicable, or require advanced post-processing [4], since they assume a stationary signal within their analysis window. In this paper we will demonstrate a technique to estimate the phase of these recurrent patterns from a noisy sensor reading based on dynamic time warping (DTW).

DTW is a technique where two signals can be aligned in time or *warped*, as to optimally fit the other signal within a certain bound. First proposed by Sakoe and Chiba [5], DTW offers an alternative to the conventional Euclidean distance measure [6] for time series that exhibit a similar pattern but that are not synchronized in time. A simple example of this would be two signals that both have a clear peaks but where the peak occurs at slightly different times. The DTW algorithm would first align the two peaks, and then compare the samples in between the peaks. Originally, the application of DTW was focused on matching speech with a library of recorded vowel sounds [7], [8], but nowadays it is a widely used technique with various kinds of time series data [6], [9], [10]. The DTW algorithm matches each of the samples of one signal with a sample of the other signal in such a way that the cost between the matched samples is minimized. The similarity of the signals is then defined by the total cost of all matched samples, and this cost can be used as a distance measure. In addition to the computed cost, the algorithm also returns an alignment between the two signals called the *warping path*. In this paper it will be shown how this alignment can be used to track repeating patterns in time series.

The DTW algorithm relies on a technique called dynamic programming (DP). The principle of DP was developed by Bellman in the 1950s, and relies on Bellman's optimality principle, which states that an optimal solution has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal solution with regard to the state resulting from the first decision [11]. As a consequence, DP breaks a dynamic optimization problem into a sequence of simpler subproblems, and allows for an efficient implementation of the combinatorial problem of finding the alignment between the two samples. It is this property of the dynamic time warping algorithm that will allow us to also find an efficient solution for the tracking of repeating patterns in time series.

In Section II, a more detailed description of the DTW algorithm is provided which will be used to show the equivalence of the distance between the aligned signals and the cost of the warping path. Subsequently, we will show in Section III how these results can be used to track repeating signal patterns.

II. Dynamic time warping

DTW aligns the samples of two signals $\boldsymbol{x} = (x[0], \dots, x[i], \dots, x[N-1])$, and $\boldsymbol{y} = (y[0], \dots, y[j], \dots, , y[M-1])$ by searching for a *warping path* [12, p. 482]. The goal of the algorithm is to find a mapping between the samples of $\boldsymbol{x}$ and $\boldsymbol{y}$, so that the

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

alignment between $\boldsymbol{x}$ and $\boldsymbol{y}$ and the alignment between $\boldsymbol{y}$ and $\boldsymbol{x}$ are jointly optimized in some sense. The warping path is defined as a set of, say K, index pairs

$$\mathcal{W} = \{(i_0, j_0), \dots (i_{K-1}, j_{K-1})\}.$$

Let $\Delta i_k = i_k - i_{k-1}$ denote the difference between two consecutive indices with $\Delta i_0 = 1$, and let Δj_k be similarly defined. The total cost of a warping path can then be defined as

$$D(\mathcal{W}) := \sum_{k=0}^{K-1} \alpha(\Delta i_k, \Delta j_k)\, d(i_k, j_k)\,,$$

where $\alpha(\Delta i_k, \Delta j_k)$ and $d(i_k, j_k)$ represent the transitions cost between two successive index pairs and alignment cost for a given index pair, respectively. The transition cost determines the weights of different steps in the warping path. Examples are $\alpha(\Delta i_k, \Delta j_k) = \Delta i_k + \Delta j_k$ for the symmetric DTW algorithm, or $\alpha(\Delta i_k, \Delta j_k) = \Delta i_k$ for asymmetric DTW [5]. A common choice for the alignment cost is the squared error $d(i_k, j_k) = (x[i_k] - y[j_k])^2$. The objective of DTW is to find the optimal path, say $\mathcal{W}^*$, that minimizes the cost $D(\mathcal{W})$. In many applications, this cost is used as a distance measure to define the similarity of two signals.

We will now define the *phase function* that maps the signal $\boldsymbol{x}$ onto $\boldsymbol{y}$, or vice versa. The phase function that aligns $\boldsymbol{x}$ to $\boldsymbol{y}$ will be referred to as $\boldsymbol{\theta_x} = (\theta_x[0], \dots, \theta_x[N-1])$, while the function that aligns $\boldsymbol{y}$ to $\boldsymbol{x}$ is referred to as $\boldsymbol{\theta_y} = (\theta_y[0], \dots, \theta_y[M-1])$. The phase function $\boldsymbol{\theta_x}$ is defined as the mapping between an index i and the first index j that is encountered in the warping path. That is

$$\theta_x[i] := \min_j \{j \,|\, (i, j) \in \mathcal{W}\}, \tag{1}$$

and similarly for $\boldsymbol{\theta_y}$

$$\theta_y[j] := \min_i \{i \,|\, (i, j) \in \mathcal{W}\}.$$

With this, the aligned signals can be expressed as

$$\begin{aligned} \boldsymbol{y}[\boldsymbol{\theta_x}] &= (y[\theta_x[0]], \dots, y[\theta_x[N-1]]), \\ \boldsymbol{x}[\boldsymbol{\theta_y}] &= (x[\theta_y[0]], \dots, x[\theta_y[M-1]]). \end{aligned}$$

Note that in the DTW algorithm we have $\Delta i_k, \Delta j_k \in \{0, 1\}$, which means that $i_k \geq i_{k-1}$ and $j_k \geq j_{k-1}$. As a consequence, we have

$$\begin{aligned} \theta_x[i_k] &= \min_j \{j \,|\, (i_k, j) \in \mathcal{W}\} \\ &\geq \min_{j'} \{j' \,|\, (i_{k-1}, j') \in \mathcal{W}\} \\ &= \theta_x[i_{k-1}]. \end{aligned}$$

That is, θ_x is monotone, and similarly for θ_y. Figure 1 shows an example of two warped signals and their corresponding warping path. The original signals $\boldsymbol{x}$ and $\boldsymbol{y}$ are indicated by the solid lines in the side and bottom subplot, respectively, whereas the warped signals are depicted as dashed lines. The warping path which maps the signals into each other is shown in the center plot (+). The colors in the center plot (cost matrix) represent the costs $d(i, j)$; the darker the color, the lower the cost. As indicated in the figure, the optimal path $\mathcal{W}^*$ is the path through the cost matrix having minimal cost.

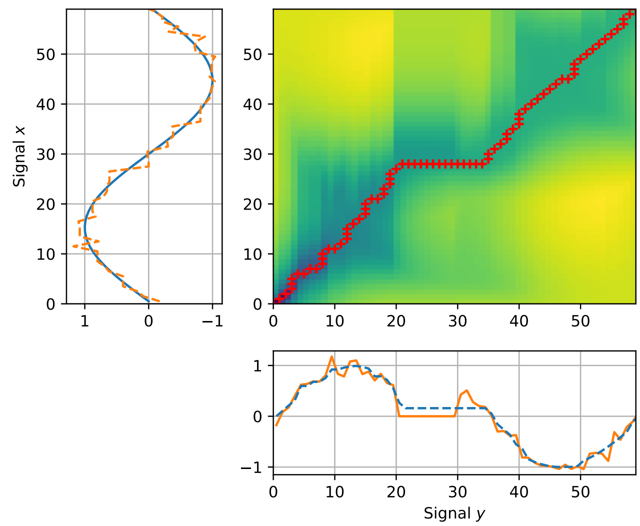

Fig. 1. Example of two warped signals and their corresponding warping path. The original signals $\boldsymbol{x}$ and $\boldsymbol{y}$ are shown by the solid lines in the side and bottom subplot, respectively, whereas the warped signals $\boldsymbol{y}[\boldsymbol{\theta_x}]$ and $\boldsymbol{x}[\boldsymbol{\theta_y}]$ are depicted as dashed lines. The warping path which maps the signals into each other is shown in the center plot (+).

A. Equivalence of the warping path cost and distance between aligned signals

In this section we will show that the cost $D(\mathcal{W})$ associated with the warping path $\mathcal{W}$ is given by the sum of the distances between the original and aligned signals. This equivalence relation will be used in section Section III to modify the DTW algorithm to track recurring patterns in time series. The Euclidean distance between the signals can be expressed as

$$d(\boldsymbol{x}, \boldsymbol{y}[\boldsymbol{\theta}_x]) = \sum_{i=0}^{N-1} (x[i] - y[\theta_x[i]])^2\,, \tag{2}$$

and

$$d(\boldsymbol{x}[\boldsymbol{\theta}_y], \boldsymbol{y}) = \sum_{j=0}^{M-1} (x[\theta_y[j]] - y[j])^2\,.$$

We have the following result.

Proposition II.1. *Let* $\alpha(\Delta i_k, \Delta j_k) := \Delta i_k + \Delta j_k$, *where* $\Delta i_k, \Delta j_k \in \{0, 1\}$. *Then*

$$D(\mathcal{W}) = d(\boldsymbol{x}, \boldsymbol{y}[\boldsymbol{\theta}_x]) + d(\boldsymbol{x}[\boldsymbol{\theta}_y], \boldsymbol{y})\,. \tag{3}$$

Proof. We have.

$$\begin{aligned} D(\mathcal{W}) &= \sum_{k=0}^{K-1} d(i_k, j_k)(\Delta i_k + \Delta j_k), \\ &= \sum_{k=0}^{K-1} d(i_k, j_k)\Delta i_k + \sum_{k=0}^{K-1} d(i_k, j_k)\Delta j_k. \end{aligned} \tag{4}$$

Note that if $i_k = i_{k-1}$, then $\Delta i_k = 0$ and similarly for Δj_k. Hence, (4) reduces to

$$D(\mathcal{W}) = \sum_{n=0}^{N-1} d(n, \theta_x[n]) + \sum_{m=0}^{M-1} d(\theta_y[m], m)\,,$$
$$= d(\boldsymbol{x}, \boldsymbol{y}[\boldsymbol{\theta}_x]) + d(\boldsymbol{x}[\boldsymbol{\theta}_y], \boldsymbol{y}).$$

This completes the proof. □

The choice $\alpha(\Delta i_k, \Delta j_k) = \Delta i_k + \Delta j_k$ results in a warping path that jointly optimizes the Euclidean distances between the warped signals. This algorithm is referred to as the symmetric DTW algorithm. Alternatively, we could define $\alpha(\Delta i_k, \Delta j_k) = \Delta i_k$. By inspection of the proof of Proposition II.1, it follows straightforwardly that in that case we have $D(\mathcal{W}) = d(\boldsymbol{x}, \boldsymbol{y}[\boldsymbol{\theta}_x])$, which results in the so-called asymmetric DTW algorithm [5]. Note that in some recent applications of the DTW algorithm, $\alpha(\Delta i_k, \Delta j_k) = 1$ is used implicitly (e.g. [13], [14]). In that case, there is no direct relationship between the distance between the aligned signals and the path cost $D(\mathcal{W})$.

III. Tracking recurring patterns using DTW

Having shown the relationship between the cost of the warping path and the distance between the aligned and unaligned signals, we will now extend the algorithm to track recurring patterns in time series. Figure 2 shows an example of such a situation, where a noisy observation is shown in the top subplot (dashed line) together with the warped signal (solid line). Clearly, a repeating pattern is visible in the graph, but the frequency (change of the phase) of the pattern changes over time. In the bottom plot, both the true phase (dashed line) and the estimated one (solid line) of the pattern is shown. For the first half of the signal, the phase is increasing and the frequency of the pattern is positive, but for the second half of the signal the phase decreases and the frequency becomes negative. When the frequency is negative, the pattern occurs in reversed order.

To model this kind of signal, let $\boldsymbol{y}$ denote a template pattern of length M, and let $\boldsymbol{y}[\boldsymbol{\theta_x}]$ denote a quasi-periodic repetition of the template pattern of length N, where $\boldsymbol{\theta_x}$ is a *latent* phase function (of length N), which we want to estimate. We observe a noisy observation $\boldsymbol{x}$ of the (quasi) periodic signal $\boldsymbol{y}[\boldsymbol{\theta_x}]$ given by

$$\boldsymbol{x} = \boldsymbol{y}[\boldsymbol{\theta_x}] + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon}$ is an additive white Gaussian noise signal. Note that when $\boldsymbol{y}[\boldsymbol{\theta_x}]$ is quasi-periodic, $\boldsymbol{\theta_x}$ is quasi-periodic as well. As an example, for a repeating pattern of constant frequency, $\boldsymbol{\theta_x}$ has the shape of a sawtooth function, where each period of the sawtooth corresponds to one repetition of the pattern. A more general example of a phase function is shown in the bottom plot of Figure 2 (dashed line). To estimate the phase function from a noisy observation $\boldsymbol{x}$, the template signal $\boldsymbol{y}$ must either be known beforehand, or must be estimated from the signal during periods where the frequency is constant. Estimating the latent phase function can be considered as the problem

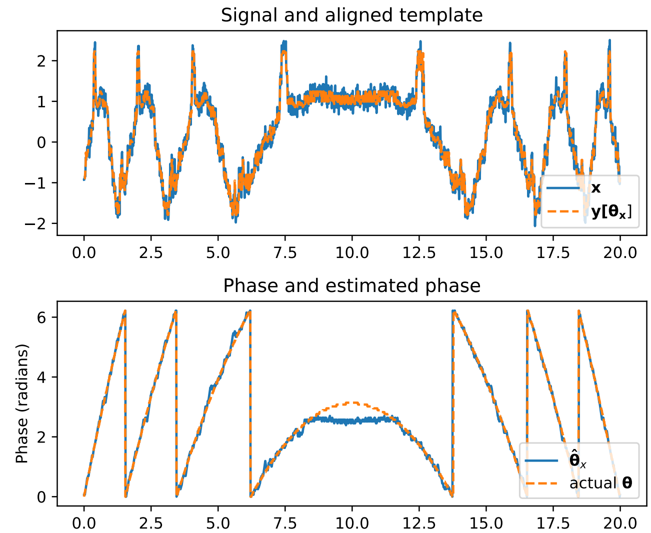

Fig. 2. This figure shows a signal that exhibits a repeating pattern with time-varying frequency. In the second half of the graph, the frequency of the pattern becomes negative and the pattern occurs in reverse order. On the bottom plot, the output of the dynamic programming based-tracking algorithm is shown, indicating the actual and estimate phase of the pattern.

of tracking the template pattern as a function of time. Given a squared-error distortion criterion, the maximum likelihood estimator (MLE) for $\boldsymbol{\theta}_x$ is given by

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{x}} = \min_{\boldsymbol{\theta}} d(\boldsymbol{x}, \boldsymbol{y}[\boldsymbol{\theta}])\,. \tag{5}$$

Without any additional contraints, the MLE will result in unrealistic phase functions since the estimates of the samples $\theta_x[i]$ are obtained independently. This can be overcome by requiring the phase function to be monotonic *within* each periodic repetition of the template signal. Note that with the original DTW algorithm, as presented in Section II, two patterns are aligned which can be viewed as the alignment of a template to a one-period signal. In that case, requiring monotonicity alone is sufficient to guarantee a realistic phase function. In the case of mapping a template to a (quasi) periodic signal, we need an additional requirement that samples of the template are mapped within the same period, and not mapped to samples of succeeding periods. In order to do so, we put an additional constraint that limits the change (slope) of the phase function. That is, we define the problem of finding the phase function as

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{x}} = \min_{\boldsymbol{\theta}} \quad d(\boldsymbol{x}, \boldsymbol{y}[\boldsymbol{\theta}_x])$$
$$\text{s.t.} \quad \theta_x[i] - \theta_x[i{-}1] \in \{0, \cdots, s_{\max}\}\,,$$

where $s_{\max}$ denotes the maximum slope of the phase function. This optimization problem can be solved using DTW by introducing specific restrictions on Δi_k and Δi_k as $\Delta i_k = 1$ and $\Delta j_k \in \{0, \cdots, s_{\max}\}$.

A. Recurring patterns with backwards motion

Recall our example of the rotating disc and the photo reflective sensor as discussed in Section I. In this example,

TABLE I
THIS TABLE SHOWS THE RELATIONSHIPS BETWEEN DIFFERENT VARIANTS OF THE DTW ALGORITHM AND THE EQUIVALENT OPTIMIZATION PROBLEMS.

DTW constraints	Equivalent optimization problem	Comments
$\alpha(\Delta i_k, \Delta j_k) = \Delta i_k + \Delta j_k$ $\Delta i_k \in \{0, 1\}$ $\Delta j_k \in \{0, 1\}$	$\min_{\boldsymbol{\theta}} \; d(\boldsymbol{x}[\boldsymbol{\theta}_y], \boldsymbol{y}) + d(\boldsymbol{x}, \boldsymbol{y}[\boldsymbol{\theta}_x])$ s.t. $\theta_x[i] \geq \theta_x[i{-}1]$ $\theta_y[j] \geq \theta_y[j{-}1]$	Symmetric DTW [5]
$\alpha(\Delta i_k, \Delta j_k) = \Delta i_k$ $\Delta i_k \in \{0, 1\}$ $\Delta j_k \in \{0, 1\}$	$\min_{\boldsymbol{\theta}} \; d(\boldsymbol{x}, \boldsymbol{y}[\boldsymbol{\theta}_x])$ s.t. $\theta_x[i] \geq \theta_x[i{-}1]$	Asymmetric DTW [5]
$\alpha(\Delta i_k, \Delta j_k) = \Delta i_k$ $\Delta i_k = 1$ $\Delta j_k \in \{0, \cdots, s_{\max}\}$	$\min_{\boldsymbol{\theta}} \; d(\boldsymbol{x}, \boldsymbol{y}[\boldsymbol{\theta}_x])$ s.t. $\lvert\theta_x[i] - \theta_x[i{-}1]\rvert \in \{0, \cdots, s_{\max}\}$	Equation (6)

it can happen that the disc rotates backwards, for example in the case where a household has solar panels installed and has a negative energy consumption during daytime. In terms of the phase function, this means that the function is not monotonically increasing anymore, but can decrease as well. Still, due to physical limitations, the slope of this phase function will be limited. To allow for this "backward rotation" phenomena, we adapt our optimization problem to

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\boldsymbol{x}} = \; & \min_{\boldsymbol{\theta}} \quad d(\boldsymbol{x}, \boldsymbol{y}[\boldsymbol{\theta}_x]) \\ & \text{s.t.} \quad \lvert\theta_x[i] - \theta_x[i{-}1]\rvert \in \{0, \cdots, s_{\max}\}. \end{aligned} \tag{6}$$

An example of this model is shown in Figure 2, where the signal is obtained from a typical photo reflective sensor used to monitor analog electricity meters [1]. As the meter spins backwards and forwards, the phase (position) of the meter is tracked using the adapted DTW algorithm where we set $s_{\max} = 1$.

Table I summarizes the relationship between minimizing Euclidean distances between aligned signals (first column) and the corresponding DTW algorithms (second column).

B. Numerical analysis

In this section, a numerical assessment of the performance of the algorithm is provided. Since no other algorithms are known to the author that implement this kind of phase tracking of recurring patterns in noise, the algorithm was compared to a threshold-based peak detection method. This algorithm assumes there is one clear peak per revolution in the signal, and increments the phase estimate by one revolution every time this peak is detected. Figure 3 shows a comparison of the algorithm of this paper with the heuristic algorithm based on detecting peaks. As can be seen from Figure 3, the algorithm outperforms the peak detection algorithm by providing a higher granularity for the phase estimate. The signal was generated with a linearly incrementing frequency from the same pattern as Figure 2, and with a SNR of roughly 6 dB. The experiment was run 1000 times for different realisations of the noise, and the RMSE between the estimated phase and the true phase was measured. The RMSE of the proposed method was roughly twenty times lower than the phase estimate based on peak detection. In a different experiment with templates with multiple peaks, the peak detection algorithm overcounted the number of peaks. For signals with decrementing phase (backwards pattern), the algorithm based on peak detections could not distinguish between forwards and backwards frequency of the signal.

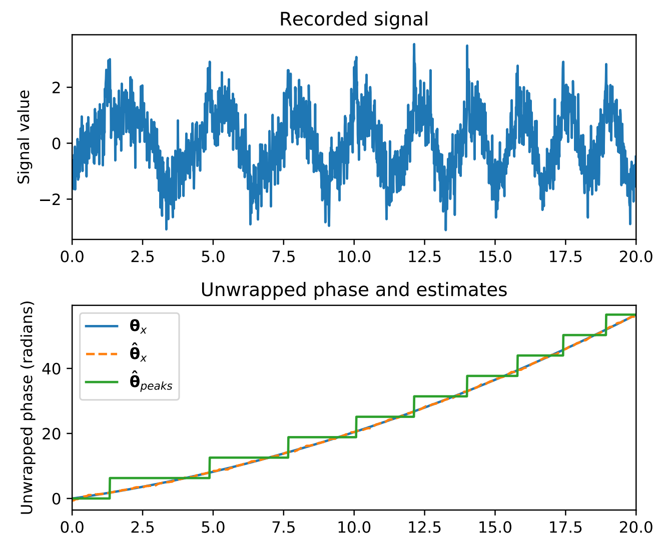

Fig. 3. This figure shows an example of the unwrapped phase estimated for a noisy signal with a repeating pattern. In the top image, the true phase is compared to the estimate from the algorithm presented in this paper and a heuristic estimate based on counting the peak in the signal.

IV. CONCLUSION

An interpretation for the cost of the DTW algorithm is provided in this paper, and it was shown that this cost is equal to the cost of the aligned signals that can be obtained from the output of the algorithm. Then, using this equivalence, a method was presented to adapt the DTW algorithm to efficiently find an estimator for the phase of a recurring pattern in time series. An implementation of this algorithm can be found under [15].

REFERENCES

[1] M. B. and, “Nonintrusive appliance load monitoring based on an optical sensor,” in *2003 IEEE Bologna Power Tech Conference Proceedings,*, vol. 4, June 2003, pp. 8 pp. Vol.4–.

[2] N. Tandon and A. Choudhury, “A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings,” *Tribology international*, vol. 32, no. 8, pp. 469–480, 1999.

[3] M. Schmidt, M. Baumert, A. Porta, H. Malberg, and S. Zaunseder, “Two-dimensional warping for one-dimensional signalsâĂŤconceptual framework and application to ecg processing,” *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5577–5588, 2014.

[4] H.-T. Wu, G. F. Lewis, M. I. Davila, I. Daubechies, S. W. Porges *et al.*, “Optimizing estimates of instantaneous heart rate from pulse wave signals with the synchrosqueezing transform,” *Methods of Information in Medicine*, vol. 55, no. 5, pp. 463–472, 2016.

[5] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition,” *IEEE transactions on acoustics, speech, and signal processing*, vol. 26, no. 1, pp. 43–49, 1978.

[6] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, “Time-series clustering–a decade review,” *Information Systems*, vol. 53, pp. 16–38, 2015.

[7] L. Rabiner, A. Rosenberg, and S. Levinson, “Considerations in dynamic time warping algorithms for discrete word recognition,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 6, pp. 575–582, 1978.

[8] H. F. Silverman and D. P. Morgan, “The application of dynamic programming to connected speech recognition,” *IEEE ASSP Magazine*, vol. 7, no. 3, pp. 6–25, 1990.

[9] T.-c. Fu, “A review on time series data mining,” *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181, 2011.

[10] P. Esling and C. Agon, “Time-series data mining,” *ACM Computing Surveys (CSUR)*, vol. 45, no. 1, p. 12, 2012.

[11] R. Bellman, *Dynamic Programming*. Princeton University Press, Princeton, NJ., 1957.

[12] S. Theodoridis and K. Koutroumbas, “Pattern recognition,” 2009.

[13] G. Li, Y. Wang, M. Li, and Z. Wu, “Similarity match in time series streams under dynamic time warping distance,” in *Computer Science and Software Engineering, 2008 International Conference on*, vol. 4. IEEE, 2008, pp. 399–402.

[14] A. Zifan, S. Saberi, M. H. Moradi, and F. Towhidkhah, “Automated ecg segmentation using piecewise derivative dynamic time warping,” *International Journal of Biological and Medical Sciences*, vol. 1, no. 3, 2006.

[15] R. Van der Vlist, “DTW tracking algorithm for repeating patterns,” https://cas.tudelft.nl/download/recurring_patterns_dtw.zip, 2019.