Enhancing Nearest Neighbor Classification Performance through Dynamic Time Warping with Progressive Constraint

Teja Vuppu[a], Teryn Cha[b], Sung-Hyuk Cha[a]
[a] Computer Science Department, Pace University, New York, NY, USA
[b] Computer Science, Essex County College, Newark, NJ, USA
tv20796n@pace.edu, yan@essex.edu, scha@pace.edu

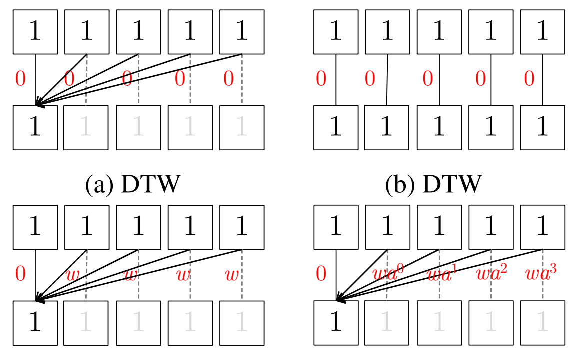
Abstract

Dynamic time warping (DTW) is a widely used metric for comparing time series data, offering elasticity in alignment. While the original DTW allows infinite elasticity without penalty, the wDTW imposes a constant penalty regardless of elastic length. In this study, we propose DTW with a progressive penalty. Experimental evaluations across diverse time series datasets demonstrate the effectiveness of this approach, utilizing nearest neighbor classification. Optimal hyperparameters, including the number of neighbors and progressive weight factor, are jointly identified with the Minkowski p value using Gaussian Process. The proposed methodology shows promise for enhancing performance across various applications leveraging DTW.

Introduction

Dynamic Time Warping (DTW) stands as a widely adopted technique for comparing temporal sequences characterized by variations in speed or timing. Originally formulated for speech recognition by Sakoe and Chiba (Sakoe and Chiba 1978), DTW has found applications spanning diverse domains, including handwriting recognition (Cha and Srihari 2002; Tappert, Suen, and Wakahara 1990) and music retrieval. The core principle of DTW involves determining the optimal alignment between two sequences by dynamically stretching or compressing the time axis of one sequence to minimize the overall distance between corresponding points in both sequences. This adaptive approach enables the identification of similar patterns amidst distortions, delays, or changes in speed.

The original DTW formulation by Sakoe and Chiba (Sakoe and Chiba 1978) imposes no constraint on the stretching of signals. For instance, two signals, $s_1 = (2)$ and $s_2 = (2, 2, 2)$, yield the same DTW values to $s_3 = (1, 1, 1)$; $\text{DTW}((2), (1, 1, 1)) = \text{DTW}((2, 2, 2), (1, 1, 1))$ as illustrated in Figure 1 (a) and (b). In an effort to address this issue, constraints were introduced to DTW in (Herrmann and Webb 2023). However, these constraints remain constant regardless of the elastic length. To mitigate this limitation, this study proposes the introduction of constraints proportional to the length of elasticity. Addressing this concern, the present study advocates for the utilization of progressive constraints as a viable alternative to achieve superior performance in DTW.



(a) DTW (b) DTW (c) constant wDTW (b) progressive wDTW

Figure 1: Dynamic Time Warping Illustration

The remainder of this paper unfolds as follows: firstly, we comprehensively review the original dynamic time warping algorithm and DTW with constant constraint. Next, we define the proposed DTW with progressive constraints. Following this, the subsequent section presents the experimental results obtained through the application of DTW with progressive constraints, shedding light on its comparative performance and efficacy in diverse scenarios.

Dynamic Time Warping

The recursive definition of the original DTW by Sakoe and Chiba (Sakoe and Chiba 1978) is outlined as follows:

$$\text{DTW}(A_{1\sim n}, B_{1\sim m}) = \begin{cases} 0 & \text{if } n = 0 \wedge m = 0 \\ \infty & \text{if } n = 0 \oplus m = 0 \\ \text{subst} & \text{if } n, m > 0 \end{cases} \quad (1)$$

$$\text{subst} = c(a_n, b_m) + \min \begin{pmatrix} \text{DTW}(A_{1\sim n-1}, B_{1\sim m-1}) \\ \text{DTW}(A_{1\sim n}, B_{1\sim m-1}) \\ \text{DTW}(A_{1\sim n-1}, B_{1\sim m}) \end{pmatrix} \quad (2)$$

The full computed table on two sample signals is given Figure 2 (a).

	1	4	1	2	1	2	1
3	2	3	5	6	8	9	7
1	2	5	3	4	4	5	5
4	5	2	5	5	7	6	8
2	6	4	3	3	4	4	5
1	6	7	3	4	3	4	4

(a) DTW

	1	4	1	2	1	2	1
3	2	4	7	9	12	14	17
1	3	5	4	6	7	9	10
4	7	3	7	6	9	9	12
2	9	6	4	5	7	8	10
1	10	10	5	5	5	7	8

(b) constant ($w = 1$)-DTW

Figure 2: Dynamic Time Warping Illustration

The DTW with penalty proposed in (Herrmann and Webb 2023) modifies the subst part as follows:

$$\text{subst} = c(a_n, b_m) + \min\begin{pmatrix} \text{DTW}(A_{1\sim n-1}, B_{1\sim m-1}) \\ \text{DTW}(A_{1\sim n}, B_{1\sim m-1}) + w \\ \text{DTW}(A_{1\sim n-1}, B_{1\sim m}) + w \end{pmatrix} \quad (3)$$

As shown in Figure 1 (c), constraints remain constant regardless of the elastic length. The full computed table on two sample signals is given Figure 2 (b).

The proposed DTW with progressive constraint modifies the subst part as follows:

$$\text{subst} = c(a_n, b_m) + \min\begin{pmatrix} \text{DTW}(A_{1\sim n-1}, B_{1\sim m-1}) \\ \text{DTW}(A_{1\sim n}, B_{1\sim m-1}) + w\alpha^{H(n,m)-1} \\ \text{DTW}(A_{1\sim n-1}, B_{1\sim m}) + w\alpha^{V(n,m)-1} \end{pmatrix} \quad (4)$$

$$\text{where } \alpha \geq 1$$

If $\alpha = 1$, it yields the same constraint as the constant constraint. If $\alpha > 1$, the constraint becomes progressive. It is illustrated in Figure 1 (d).

The horizontal and vertical elastic length information need to be computed by the following equations:

$$H(n,m) = \begin{cases} H(n, m-1) + 1 & \text{if } c_h = \min(c_d, c_h, c_v) \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$V(n,m) = \begin{cases} V(n-1, m) + 1 & \text{if } c_v = \min(c_d, c_h, c_v) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$\text{where } \begin{cases} c_d = \text{DTW}(A_{1\sim n-1}, B_{1\sim m-1}) \\ c_h = \text{DTW}(A_{1\sim n}, B_{1\sim m-1}) + w\alpha^{H(n,m-1)} \\ c_v = \text{DTW}(A_{1\sim n-1}, B_{1\sim m}) + w\alpha^{V(n-1,m)} \end{cases}$$

Experiments

We consider the Bird/Chicken dataset from the UCR Time Series Archive, aimed at distinguishing between the outline of a bird and a chicken. The dataset comprises 20 instances of each class.

Both the original DTW and wDTW algorithms yield an accuracy of 75%. However, employing Gaussian Process to optimize hyperparameters, particularly the progressive constraint, enhances the accuracy to 85% at $w = 0.01$ and $\alpha = 1.2$.

Conclusion

In conclusion, we have introduced a novel dynamic time warping with progressive constraint. Our experimental findings demonstrate its superiority over conventional DTW and DTW with constant constraint. We anticipate that our proposed method holds promise for enhancing the performance of various applications employing DTW. Future research directions include exploring the efficacy of our approach on larger and more diverse datasets.

References

Cha, S.-H., and Srihari, S. N. 2002. *Approximate String Matching for Character Recognition and Analysis*, volume 13 of *COMBINATORIAL OPTIMIZATION series*. Springer.

Herrmann, M., and Webb, G. I. 2023. Amercing: An intuitive and effective constraint for dynamic time warping. *Pattern Recognition* 137:109333.

Sakoe, H., and Chiba, S. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26(1):43–49.

Tappert, C. C.; Suen, C. Y.; and Wakahara, T. 1990. The state of the art in online handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(8):787–808.