Approximation Algorithms for Dynamic Time Warping on Run-Length Encoded Strings

Zoe Xi and William Kuszmaul*

Abstract

Dynamic Time Warping (DTW) is a well-known similarity measure for comparing strings that encode time series data. First introduced in 1968, DTW has found applications to a large variety of areas including bioinformatics, signature verification, and speech recognition. The standard dynamic-programming algorithm for computing the DTW distance between two strings is quadratic-time in terms of the lengths of the two strings. It is also known that there are no strongly subquadratic-time algorithms for computing DTW distance under the Strong Exponential Time Hypothesis (SETH).

The standard quadratic-time algorithm for computing DTW distance is often too slow in practice, and many attempts have been made to design and analyze algorithms for computing and approximating DTW distance on special strings. In particular, there is a recently published algorithm for computing the DTW distance between two run-length encoded (RLE) strings that is cubic-time in terms of the numbers of runs in the two strings (rather than in terms of the numbers of characters). It remains an open question whether a (near) quadratic-time algorithm is possible for this problem.

Given two strings x and y, where x is a RLE string with k runs and y is a RLE string with l runs, we show that it is possible to obtain a provably good approximation for $DTW(x,y)$ in near-quadratic time. In particular, we present three algorithmic results:

- The first algorithm computes a 2-approximation for $DTW(x,y)$ in $O(kl)$-time, that is, it returns a number w satisfying $DTW(x,y) \leq w \leq 2 \cdot DTW(x,y)$.
- Given $\varepsilon > 0$, the second algorithm computes a $(1+\varepsilon)$-approximation for $DTW(x,y)$ in $\tilde{O}(kl/\varepsilon^2)$-time while assuming that the underlying characters are taken from some Hamming space (that is, the distance between any two letters is either 0 or 1).
- Given $\varepsilon > 0$, the third algorithm computes a $(1+\varepsilon)$-approximation for $DTW(x,y)$ in $\tilde{O}(kl/\varepsilon^3)$-time, while allowing for characters in the strings to be taken from an arbitrary metric space in which distances are polynomially-bounded integers.

For the time being, our presented algorithms are the state-of-art in terms of their formally proven time complexities.

*William Kuszmaul is the mentor of Zoe Xi for the research reported in this paper.