# cuTimeWarp: Accelerating Soft Dynamic Time Warping on GPU

Alex Kyllo Afrooz Rahmati

March 4, 2021

**Abstract**

This report explores techniques for optimizing the computation of Soft Dynamic Time Warping, a differentiable sequence dissimilarity measure, on graphics processing units (GPU), for the purpose of enabling high-performance machine learning on time series datasets.

## Introduction

Time series machine learning is a research area with countless useful applications such as recognizing sounds and gestures. Clustering or classifying large time series datasets is challenging partly because of the need to define a measure of dissimilarity between two time series. Practical applications require finding common structure despite different speeds or phases; a word means the same whether spoken quickly or slowly. Another requirement is that the measure must be differentiable so that its gradient can be used as a loss function to minimize in model fitting. Finally, the measure must be efficient to calculate. To this end we will explore GPU acceleration of Soft Dynamic Time Warping (Soft-DTW) [1], a differentiable dissimilarity measure, to enable high performance time series machine learning.

Dynamic Time Warping (DTW) is an algorithm to compute the dissimilarity between two time series, which may vary in speed and phase. The basic algorithm for DTW is to use Bellman's recursion, a dynamic programming technique, to find the lowest-cost path diagonally across a pairwise distance matrix. The computation cost for this approach is quadratic(mn) for time series vectors of length m and n [1]. The formula for the DTW between time series x and y is:

$$DTW(x,y) = min_{\pi} \sqrt{\sum_{(i,j)\in\pi} d(x_i, y_j)^2}$$

Where $d(x_i, y_j)^2$ is the cost function, typically pairwise squared Euclidean distance. The loss function for DTW is not differentiable due to the min operation within the formula; a small change in the input time series may result in zero change in the path cost. However, we can create a differentiable version called Soft-DTW by replacing the min with a soft-min function [1]:

$$\text{soft-min}_{\gamma}(a_1, ..., a_n) = -\gamma log \sum_i e^{-a_i/\gamma}$$

Hence, Soft-DTW is parameterized by the smoothing constant gamma, which becomes a tunable hyperparameter in machine learning model training applications.

A common technique in machine learning with Soft-DTW is the computation of barycenters, which are centroids within the space of a set of time series. The differentiability of Soft-DTW allows for barycenter finding via gradient descent, and then new observations can be clustered or classified by finding the closest barycenter.

## Methods

## Results

## Discussion

## References

[1] M. Cuturi and M. Blondel, "Soft-DTW: A differentiable loss function for time-series," *arXiv:1703.01541 [stat]*, Feb. 20, 2018. arXiv: `1703.01541`. [Online]. Available: `http://arxiv.org/abs/1703.01541` (visited on 01/16/2021).