Time Series Similarity Measures

Samuel Caekaert
Gianforte School of Computing
Montana State University
Bozeman, MT 59715, USA

1. Introduction

Similarity measures play an important role in various data mining tasks, with clustering being especially important. This work will review similarity measures and discuss their application with respect to time series data.

2. Similarity Measures

Numerous similarity measures have been applied to time series data. The most straightforward measure is to simply calculate any L_n distance between two time series. When this is done between two time series, it is sometimes referred to as a *lock step* measure. Another common approach is to extract features or coefficients from the underlying series and then use the extracted information to calculate similarity. Some common approaches for extracting features include Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), and Auto-regression (AR) models (Serra and Arcos, 2014). Other types of measures that are commonly applied to time series are referred to as *elastic* similarity measures. These measures are appropriate for series that may not be of equal speed or are not aligned. Some elastic measures include Dynamic Time Warping (DTW), Edit distance for real sequences (EDR), Time-warped edit distance (TWED), and Minimum jump costs (MJC) (Serra and Arcos, 2014).

Serra and Arcos (2014) compare the above measures by using them as the underlying similarity measure for the K-Nearest-Neighbor (KNN) algorithm. They compare the performance on out of sample test data by calculating the rank of each measure on 45 time series datasets. They find that the measures result in the following average rank: { TWED: 2.29, DTW: 3.00, MJC: 3.04, EDR: 3.42, L2: 4.40, FC: 5.07, AR: 6.80 }. They also give the number of times each metric performs best as { TWED: 20 (6), EDR: 7 (2), DTW: 6 (2), MJC: 5 (0), FC: 4 (2), L2: 2 (0), AR: 1 (1) }.

2.1 Dynamic Time Warping

Dynamic Time Warping (DTW) is an elastic similarity measure that optimally aligns time series using dynamic programming. A recursive equation is defined by Serra and Arcos (2014) as,

$$D_{i,j} = f\left(x_i, y_j\right) + \min\left\{D_{i,j-1}, D_{i-1,j}, D_{i-1,j-1}\right\}. \tag{1}$$

This equation is calculated for all $i \in 1, \ldots, M$ and $j \in 1, \ldots, N$, where M and N are the number of values in each time series. The function $f()$ can be any L_n distance measure.

In the most straightforward implementation of DTW, an N by M matrix is calculated and populated using two loops. In this way, the time and space complexity are $O(N * M)$. Gold and Sharir (2018) propose an algorithm that runs in $O\left(n^2 \log\log\log n / \log\log n\right)$ time and space complexity. Recently, Tralie and Dempsey (2020) proposed a divide and conquer algorithm that runs with a space complexity of $O(N + M)$. The space complexity benefits come at the cost of average time complexity, with the authors stating it requires twice as much computation.

Below is Python3 code that calculates DTW distance between the two series from the presentation. The output from this code is shown in Figure 1.

```
import numpy as np
import matplotlib.pyplot as plt
from dtaidistance import dtw # https://pypi.org/project/dtaidistance/
from dtaidistance import dtw_visualisation as dtwvis

c1 = np.array([0,0,1,1,0])
c2 = np.array([0,1,1,0,0])

path = dtw.warping_path(c1, c2)
offset = 2 # Offset for curve2
plt.plot(c1)
plt.plot([c2[i] + offset for i in range(len(c2))])
for i,line in enumerate(path):
    x_coord = [line[0], line[1]]
    y_coord = [c1[line[0]], c2[line[1]]+offset]
    plt.plot(x_coord,y_coord,'g')

plt.title("The DTW distance between C1 and C2 is %s" % dtw.distance(c1,c2))
plt.show()
```

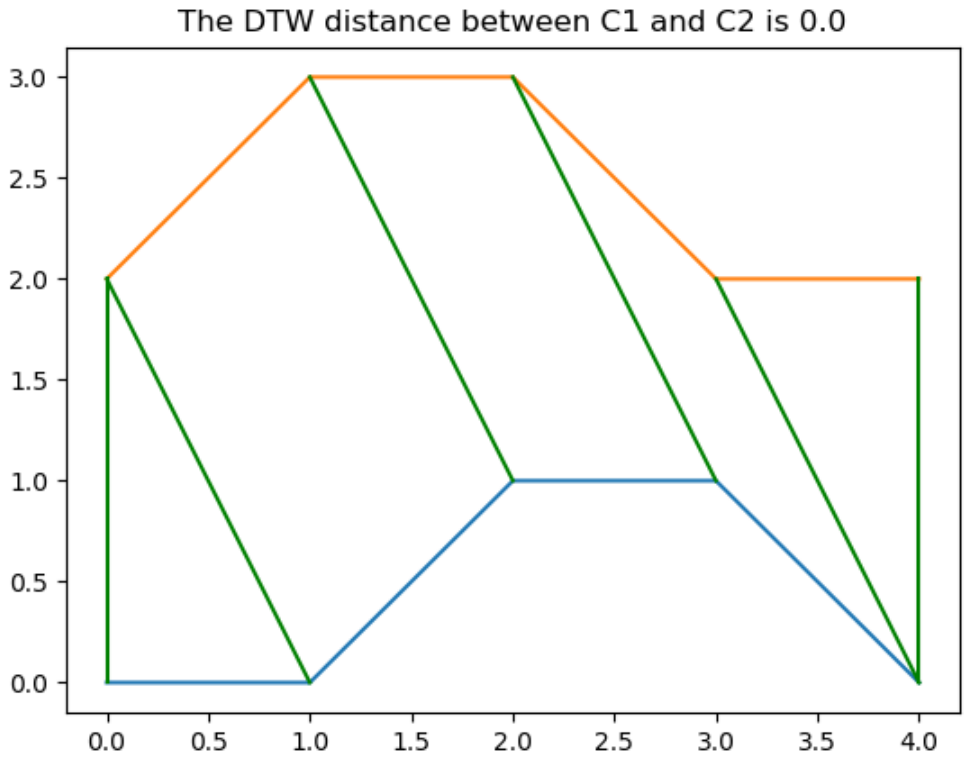

Figure 1: DTW Warping Path between two curves

References

Omer Gold and Micha Sharir. Dynamic time warping and geometric edit distance: Breaking the quadratic barrier. *ACM Transactions on Algorithms (TALG)*, 14(4):1–17, 2018.

Joan Serra and Josep Ll Arcos. An empirical evaluation of similarity measures for time series classification. *Knowledge-Based Systems*, 67:305–314, 2014.

Christopher Tralie and Elizabeth Dempsey. Exact, parallelizable dynamic time warping alignment with linear memory. *arXiv preprint arXiv:2008.02734*, 2020.