Consider n streams of time series describing the utilization traces of n resources, each at some (fixed) granularity. Denote the space of a time series as S

Define a function d:

$$\begin{aligned} d_{a,b} \colon S^2 &\to [0,1] \\ (x[a:b], y[a:b]) &\mapsto [0,1] \end{aligned} \tag{1}$$

where d maps a portion of two time series to a real number in $[0,1]$ that represents the distance between the partial time series.

Additionally, we require d to satisfy the triangle inequality:

$$d(x,y) \leq d(x,z) + d(z,y) \tag{2}$$

Out of the box Dynamic Time Warping (DTW) does not satisfy the triangle inequality, so we leverage an approximate lower-bound DTW [1].

Suppose $p_1, p_2 \in [0,1]$ with $p_1 \geq p_2$. Then, for a hash function $h \in \mathcal{H}, \Pr[h(x) = h(y)] \geq p_1 \forall\, x, y \in S$ with $d(x,y) \leq r$ for some fixed r. Additionally, for $x, y \in S$ with $d(x,y) > r, \Pr[h(x) = h(y)] \leq p_2$ for the same, fixed r.

Bibliography

[1] Daniel Lemire. Faster retrieval with a two-pass dynamic-time-warping lower bound. *Pattern Recogn.*, 42(9):2169–2180, September 2009.