Appendix IV

Dynamic time warping :

For the automatic detection of a pattern in a time series, an approximate or "fuzzy" matching process is required that can capture all the time-series fragments within which the approximate shape of the expected pattern is detected. Specifically, the pattern detection task involves searching a time series $S = s_1, s_2, s_3, \ldots, s_n$ for instances of a given template $T = t_1, t_2, t_3, \ldots, t_m$. In the dynamic time warping (DTW) technique, this pattern detection task is achieved by applying a dynamic programming approach to align the two sequences S and T in a way so that some distance measure is minimized. To achieve a reasonable fit, the time axis may be stretched or compressed.

Shown below in Figure A.1 is a n-by-m grid where each grid-point (i, j) corresponds to an alignment between elements s_i and t_j of S and T respectively. A warping path $W = w_1, w_2, \ldots, w_p$ where each w_k corresponds to some point in the grid, is a sequence of grid points which aligns the elements of S with the elements of T such that the 'distance' between them is minimized.

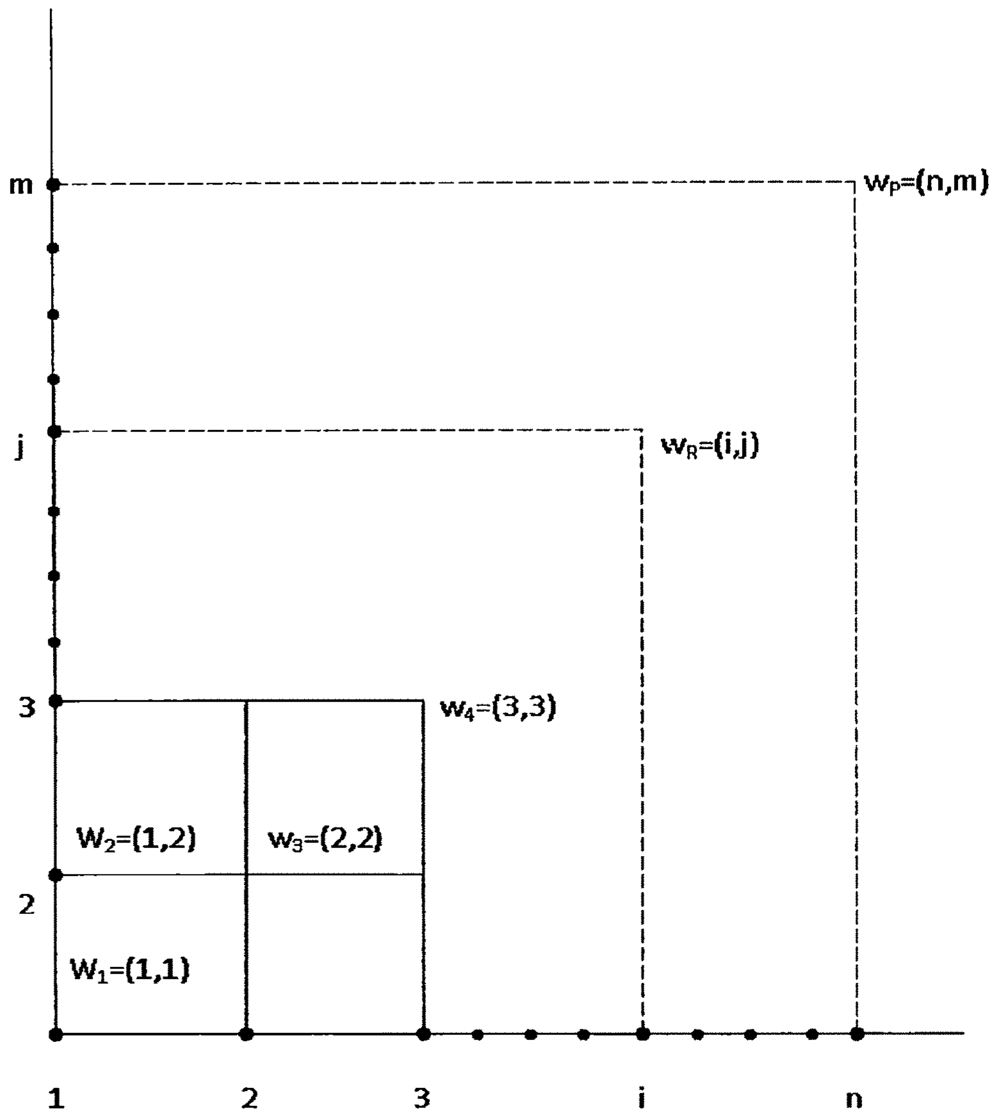

Figure A.1 A warping path W = (w_1, w_2, w_3, ..., w_p)

If δ is the distance measure used to compute the distance between any two elements s_i and t_j of S and T respectively, then the dynamic time-warping problem can be formally defined as a minimization over potential warping paths based on the cumulative distance for each path. i.e.

$$\text{DTW}(S, T) = \min_W \{\sum_{k=1}^{p} \delta(w_k)\}.$$

Searching through all possible warping paths however leads to a combinatorial explosion. Hence, several restrictions are placed on permissible paths between two grid points to reduce the search space – viz. the space of possible warping paths. A few key restrictions are outlined below :

Monotonicity: The points in W are monotonically ordered with respect to time, i.e. for consecutive pairs w_{k-1} and w_k in W, $i_{k-1} \le i_k$ and $j_{k-1} \le j_k$.

Continuity: The allowable steps taken by the path W in the grid are confined to neighboring points. i.e. $i_k - i_{k-1} \le 1$ and $j_k - j_{k-1} \le 1$.

Warping Window: Allowable points are constrained to fall within a given warping window, $| i_k - j_k | \le w$ where w is the size of the warping window, which is a positive integer.

Boundary Conditions: The typical boundary conditions used are $i_1 = j_1 = 1$ and $i_p = n$, $j_p = m$.

The dynamic programming formulation is based on the following recurrence relation which defines the cumulative distance $\gamma(i,j)$ for each grid-point (i, j) :-

$$\gamma(i,j) = \delta(i,j) + \min[\gamma(i-1,j), \gamma(i-1,j-1), \gamma(i,j-1)]$$

i.e. the cumulative distance for a grid-point is the sum of the distance between the elements specified by that grid-point and the minimum of the cumulative distances of the neighboring grid-points.

It may be noted that the cumulative distance associated with the best warping path is simply a raw score. If a template is matched along a segment of time series values that are small in magnitude, the raw cumulative distance will be less than that found when matching a similarly shaped segment with larger values. However, ideally, matches differing only in scale should be comparable – requiring normalization of the raw scores. A simple way to normalize the cumulative distance for this purpose is to divide it by the value of the anchor point i.e. by the value of the initial point of the match in the time series. The normalized cumulative distance can then be used to determine if the degree of fit of the time-series S and the given template T is sufficiently good or not.