Speech Recognition and Processing

Lecture 4

Yossi Keshet
Agenda

- DFT Example
- Spectrogram
- Dynamic Time Warping
DFT Example
Spectrogram
Waveform to Spectrogram
Rotate it by 90 degrees
MAP spectral amplitude to a grey level (0-255) value. 0 represents black and 255 represents white. Higher the amplitude, darker the corresponding region.
Spectrogram

• Phonemes and their properties are better observed in spectrograms

Vowel (a)  Consonants (p)  Fricatives (sh)
Dynamic Time Warping (DTW)
Consider the task of Automatic Speech Recognition, A simple approach would be to build a 1-NN Classifier.

“Bird”?

“Cat”

“Dog”

“Bird”
Dynamic Time Warping

How can we compare 2 sequences of different length?
Linear alignment
Linear alignment
Linear alignment
Linear alignment
Linear alignment

This is not a desired alignment since we cannot compute a frame-by-frame distance.
DTW Intuition
DTW: Alignment
DTW: Alignment
DTW: Alignment
DTW: Distance

\[(d_{i,j}^2) = (\frac{\sum_{t=0}^{T} ((x_{i,t} - y_{j,t})^2)}{T})\]
DTW: Alignment
In practice
Consider the following time series:

| Time series A | -0.87 | -0.84 | -0.85 | -0.82 | -0.23 | 1.95 | 1.36 | 0.60 | 0.0 | -0.29 |
|               | -0.88 | -0.91 | -0.84 | -0.82 | -0.24 | 1.92 | 1.41 | 0.51 | 0.03| -0.18 |

| Time series B | -0.60 | -0.65 | -0.71 | -0.58 | -0.17 | 0.77 | 1.94 |
|               | -0.46 | -0.62 | -0.68 | -0.63 | -0.32 | 0.74 | 1.97 |
DTW

- Consider the following matrix

- Each cell \((i, j)\) holds \(d(TS_A[i], TS_B[j])\)

- A path from \((0,0)\) to \((m, n)\) represents an alignment between \(TS_A, TS_B\)

- Can only move forward in time (why?)

- We wish to find the alignment that carries a minimal distance penalty.

\[
d\left(\begin{bmatrix} 0.0 \\ 0.03 \end{bmatrix}, \begin{bmatrix} -0.58 \\ -0.63 \end{bmatrix}\right)
\]
DTW

<table>
<thead>
<tr>
<th>Time Series A</th>
<th>-0.87</th>
<th>-0.84</th>
<th>-0.85</th>
<th>-0.82</th>
<th>-0.23</th>
<th>1.95</th>
<th>1.36</th>
<th>0.60</th>
<th>0.0</th>
<th>-0.29</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.88</td>
<td>-0.91</td>
<td>-0.84</td>
<td>-0.82</td>
<td>-0.24</td>
<td>1.92</td>
<td>1.41</td>
<td>0.51</td>
<td>0.03</td>
<td>-0.18</td>
</tr>
</tbody>
</table>

Time Series B

Diagram showing the Dynamic Time Warping (DTW) algorithm with arrows indicating the warping path between two time series.
DTW

Time Series A

-0.87  -0.84  -0.85  -0.82  -0.23  1.95  1.36  0.60  0.0  -0.29
-0.88  -0.91  -0.84  -0.82  -0.24  1.92  1.41  0.51  0.03  -0.18

Time Series B

-0.60  -0.65  -0.58  -0.17  0.77  0.74  1.97  
-0.65  -0.66  -0.32  0.74  
-0.46  -0.62  -0.32  

Grid with arrows connecting corresponding values in the two time series.
Can we compute the cost of the minimal path efficiently?
DTW: Dynamic Programming

\[
DTW[0,0] = d(0,0)
\]

\[
DTW[i, j] = d(i, j) + \min\left(\begin{array}{c}
DTW[i - 1, j - 1] \\
DTW[i, j - 1] \\
DTW[i - 1, j]
\end{array}\right)
\]
**DTW: Example**

$X_1 = \{1, 2, 3, 5, 5, 5, 6\}$

$X_2 = \{1, 1, 2, 2, 3, 5\}$

$$d(x, y) = |x - y|$$

$$DTW[i, j] = d(i, j) + \min \begin{pmatrix} DTW[i - 1, j - 1] \\ DTW[i, j - 1] \\ DTW[i - 1, j] \end{pmatrix}$$

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>11</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>20</td>
<td>14</td>
<td>14</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>
Recall our 1-NN classifier,

“?”

\[ DTW_{dist} = 10.3 \]

\[ DTW_{dist} = 15.1 \]

\[ DTW_{dist} = 9.4 \]

“Cat”

“Dog”

“Bird”
Time-normalization constraints

• Endpoint constraints: $i = 1, \ldots, T_{X_1}, j = 1, \ldots, T_{X_2}$

• Monotonically conditions: no negative slope. The constraints eliminates the possibility of reverse warping along the time axis.

• Local continuity constraints

• Global path constraints

• Slope weighting
Local continuity constraints

<table>
<thead>
<tr>
<th>Type</th>
<th>Allowable Path Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$p_1 \rightarrow (1, 0)$</td>
</tr>
<tr>
<td></td>
<td>$p_2 \rightarrow (1, 1)$</td>
</tr>
<tr>
<td></td>
<td>$p_3 \rightarrow (0, 1)$</td>
</tr>
<tr>
<td>II</td>
<td>$p_1 \rightarrow (1, 1)(1, 0)$</td>
</tr>
<tr>
<td></td>
<td>$p_2 \rightarrow (1, 1)$</td>
</tr>
<tr>
<td></td>
<td>$p_3 \rightarrow (1, 1)(0, 1)$</td>
</tr>
<tr>
<td>III</td>
<td>$p_1 \rightarrow (2, 1)$</td>
</tr>
<tr>
<td></td>
<td>$p_2 \rightarrow (1, 1)$</td>
</tr>
<tr>
<td></td>
<td>$p_3 \rightarrow (1, 2)$</td>
</tr>
<tr>
<td>IV</td>
<td>$p_1 \rightarrow (1, 1)(1, 0)$</td>
</tr>
<tr>
<td></td>
<td>$p_2 \rightarrow (1, 2)(1, 0)$</td>
</tr>
<tr>
<td></td>
<td>$p_3 \rightarrow (1, 1)$</td>
</tr>
<tr>
<td></td>
<td>$p_4 \rightarrow (1, 2)$</td>
</tr>
<tr>
<td>V</td>
<td>$p_1 \rightarrow (1, 1)(1, 0)(1, 0)$</td>
</tr>
<tr>
<td></td>
<td>$p_2 \rightarrow (1, 1)(1, 0)$</td>
</tr>
<tr>
<td></td>
<td>$p_3 \rightarrow (1, 1)$</td>
</tr>
<tr>
<td></td>
<td>$p_4 \rightarrow (1, 1)(0, 1)$</td>
</tr>
<tr>
<td></td>
<td>$p_5 \rightarrow (1, 1)(0, 1)(0, 1)$</td>
</tr>
</tbody>
</table>
Global path constraints

The effect of global path constraints and range limiting on the allowable regions of time-wrapping functions.
Slope weighting - local
Slope weighting - global
Questions?