E9:261

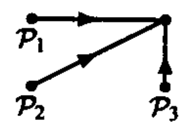
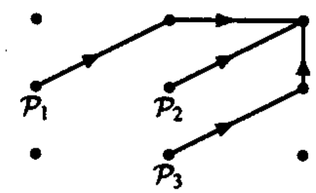
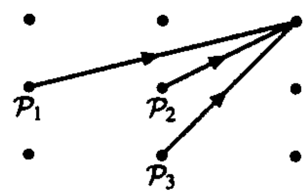
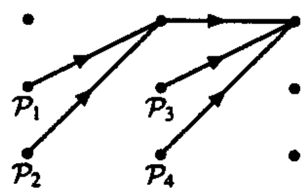
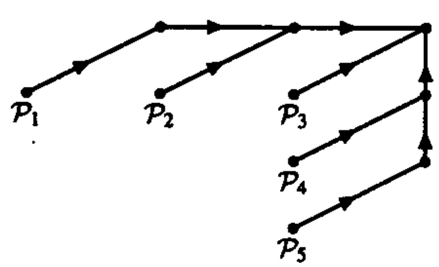
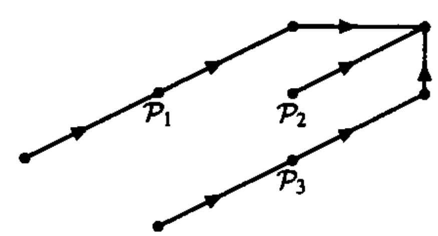
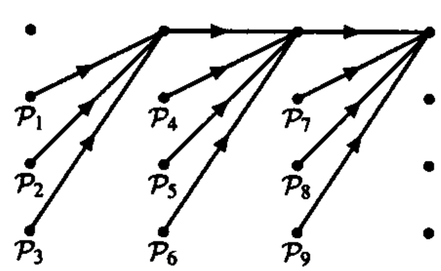
29-02-2016

Recap …

- Dynamic programming
  - solving an optimal search with sub-paths being optimal
- Dynamic Time Warping
  - Constraints on the path
    - end-point constraints
    - monotonicity constraints
    - local constraints
    - global constraints
    - slope weighting
- DTW recursive solution

Local constraints in DTW

TABLE 4.5. Summary of sets of local constraints and the resulting path specifications

Type		Allowable Path Specification
I		$\mathcal{P}_1 \to (1,0)$ $\mathcal{P}_2 \to (1,1)$ $\mathcal{P}_3 \to (0,1)$
II		$\mathcal{P}_1 \to (1,1)(1,0)$ $\mathcal{P}_2 \to (1,1)$ $\mathcal{P}_3 \to (1,1)(0,1)$
III		$\mathcal{P}_1 \to (2,1)$ $\mathcal{P}_2 \to (1,1)$ $\mathcal{P}_3 \to (1,2)$
IV		$\mathcal{P}_1 \to (1,1)(1,0)$ $\mathcal{P}_2 \to (1,2)(1,0)$ $\mathcal{P}_3 \to (1,1)$ $\mathcal{P}_4 \to (1,2)$
V		$\mathcal{P}_1 \to (1,1)(1,0)(1,0)$ $\mathcal{P}_2 \to (1,1)(1,0)$ $\mathcal{P}_3 \to (1,1)$ $\mathcal{P}_4 \to (1,1)(0,1)$ $\mathcal{P}_5 \to (1,1)(0,1)(0,1)$
VI		$\mathcal{P}_1 \to (1,1)(1,1)(1,0)$ $\mathcal{P}_2 \to (1,1)$ $\mathcal{P}_3 \to (1,1)(1,1)(0,1)$
VII		$\mathcal{P}_1 \to (1,1)(1,0)(1,0)$ $\mathcal{P}_2 \to (1,2)(1,0)(1,0)$ $\mathcal{P}_3 \to (1,3)(1,0)(1,0)$ $\mathcal{P}_4 \to (1,1)(1,0)$ $\mathcal{P}_5 \to (1,2)(1,0)$ $\mathcal{P}_6 \to (1,3)(1,0)$ $\mathcal{P}_7 \to (1,1)$ $\mathcal{P}_8 \to (1,2)$ $\mathcal{P}_9 \to (1,3)$

“Fundamentals of Speech Recognition”, Rabiner and Juang

DTW Solution

1. Initialization

$$D_A(1,1) = d(1,1)m(1).$$

2. Recursion

For $1 \leq i_x \leq T_x$, $1 \leq i_y \leq T_y$ such that i_x and i_y stay within the allowable grid, compute

$$D_A(i_x, i_y) = \min_{(i'_x, i'_y)}[D_A(i'_x, i'_y) + \zeta((i'_x, i'_y),\ (i_x, i_y))],$$

where $\zeta((i'_x, i'_y),\ (i_x, i_y))$ is defined by Eq. (4.162).

3. Termination

$$d(\mathcal{X}, \mathcal{Y}) = \frac{D_A(T_x, T_y)}{M_\phi}.$$

$$\zeta((i'_x,\ i'_y), (i_x,\ i_y)) = \sum_{\ell=0}^{L_s} d(\phi_x(T' - \ell), \phi_y(T' - \ell))m(T' - \ell)$$

"Fundamentals of Speech Recognition", Rabiner and Juang

Disadvantages of DTW

- **Based on heuristics**
- **Creating templates from large number of examples can be hard.**
- **With large vocabulary sizes, computationally intractable.**
- **Not a statistical method.**