Dynamic Programming

Algorithms: Design and Analysis, Part II

Sequence Alignment
Optimal Substructure

Problem Definition

Recall: Sequence alignment. [Needleman-Wunsch score = Similarity measure between strings]

Example:

A	G	G	G	C	T
A	G	G	-	C	A

Total penalty $= \alpha_{\text{gap}} + \alpha_{\text{AT}}$

Input: Strings $X = x_1 \ldots x_m$, $Y = y_1 \ldots y_n$ over some alphabet Σ (like $\{$A,C,G,T$\}$)

- Penalty α_{gap} for inserting a gap, α_{ab} for matching a & b [presumably $\alpha_{ab} = 0$ of $a = b$]

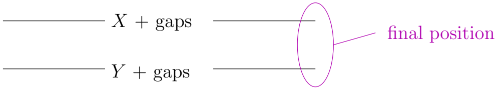
Feasible solutions: Alignments - i.e., insert gaps to equalize lengths of the string

Goal: Alignment with minimum possible total penalty

A Dynamic Programming Approach

Key step: Identify subproblems. As usual, will look at structure of an optimal solution for clues.
[i.e., develop a recurrence + then reverse engineer the subproblems]

Structure of optimal solution: Consider an optimal alignment of X, Y and its final position:



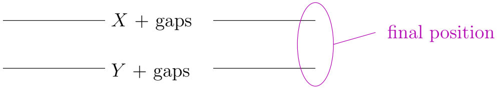
Question: How many <u>relevant</u> possibilities are there for the contents of the final position?

A) 2 C) 4
B) 3 D) mn

Case 1: x_m, y_n matched, case 2: x_m matched with a gap, case 3: y_n matched with a gap [Pointless to have 2 gaps]

Optimal Substructure

(1) x_m & y_n, (2) x_m & gap, (3) y_n & gap



Point: Narrow optimal solution down to 3 candidates.

Optimal substructure: Let $X' = X - x_m$, $Y' = Y - y_n$.

If case (1) holds, then induced alignment of X' & Y' is optimal.
If case (2) holds, then induced alignment of X' & Y is optimal.
If case (3) holds, then induced alignment of X & Y' is optimal.

Optimal Substructure (Proof)

Proof: [of Case 1, other cases are similar]

By contradiction. Suppose induced alignment of X', Y' has penalty P while some other one has penalty $P^* < P$.

$\Rightarrow$ Appending $\begin{matrix} x_m \\ y_n \end{matrix}$ to the latter, get an alignment of X and Y

with penalty $P^* + \alpha_{x_m y_n} < \boxed{P + \alpha_{x_m y_n}}$

Contents of final position

Penalty of original alignment

$\Rightarrow$ Contradicts optimality of original alignment of X & Y. QED!