CMPSC 465
Data Structures and Algorithms
Spring 2022

Instructor: Chunhao Wang

Dynamic Programming

Dynamic Programming

Prelude

Key steps to design DP algorithms

1. Identify subproblems
2. Recurrence

   e.g. $L(j) = 1 + \max\{L(i) : a_i < a_j\}$
3. Base case

Dynamic Programming

Edit Distance (Textbook Section 6.3)

Edit distance

Motivation: consider DNA sequences $x = ACGTA$, $y = ATCTG$.

Note $|x| \neq |y|$ in general

Question: how far away are x and y?

Definition

The **edit distance** between x and y, denoted by $d(x, y)$, is the minimum number of insertions, deletions, and substitutions needed to transform x to y

Consider the following **alignments**:

$$\begin{array}{lcccccc} \text{x:} & \text{A} & \text{-} & \text{C} & \text{G} & \text{T} & \text{A} \\ \text{y:} & \text{A} & \text{T} & \text{C} & \text{-} & \text{T} & \text{G} \end{array}$$

$$\text{cost}: 3 \ (\text{optimal})$$

$$\begin{array}{lccccccc} \text{x:} & \text{A} & \text{C} & \text{-} & \text{-} & \text{G} & \text{T} & \text{A} \\ \text{y:} & \text{A} & \text{T} & \text{C} & \text{T} & \text{G} & \text{-} & \text{-} \end{array}$$

$$\text{cost}: 5$$

So $d(x, y) = 3$

Edit distance — subproblem

Consider two strings

$$x = x_1 x_2 \cdots x_m \quad \text{and} \quad y = y_1 y_2 \cdots y_n$$

Subproblem: consider prefix $x_1 \cdots x_i$ and $y_1 \cdots y_j$ ($i \leq m$, $j \leq n$)

Define

$$E(i, j) = d(x_1 \cdots x_i, y_1 \cdots y_j)$$

Optimal solution: $E(m, n)$

How to use the solution to the subproblems to solve $E(i, j)$?

Recurrence (I)

Look at the rightmost column:

Case 1 $\begin{matrix} x_1 & \cdots & x_{i-1} & x_i \\ y_1 & \cdots & y_j & - \end{matrix}$

Contributes 1 to the cost plus the cost of alignment $\begin{matrix} x_1 & \cdots & x_{i-1} \\ y_1 & \cdots & y_j \end{matrix}$

$$E(i,j) = 1 + E(i-1,j)$$

Case 2 $\begin{matrix} x_1 & \cdots & x_i & - \\ y_1 & \cdots & y_{j-1} & y_j \end{matrix}$

Contributes 1 to the cost plus the cost of alignment $\begin{matrix} x_1 & \cdots & x_i \\ y_1 & \cdots & y_{j-1} \end{matrix}$

$$E(i,j) = 1 + E(i,j-1)$$

Case 3 $\begin{matrix} x_1 & \cdots & x_{i-1} & x_i \\ y_1 & \cdots & y_{j-1} & y_j \end{matrix}$

$$E(i,j) = \begin{cases} E(i-1,j-1) & \text{if } x_i = y_j \\ 1 + E(i-1,j-1) & \text{otherwise} \end{cases}$$

Recurrence (II)

The recurrence:

$$E(i,j) = \min\{1 + E(i-1,j), 1 + E(i,j-1), \mathrm{diff}(i,j) + E(i-1,j-1)\},$$

where

$$\mathrm{diff}(i,j) = \begin{cases} 1 & \text{if } x_i \neq y_j \\ 0 & \text{otherwise} \end{cases}$$

Optimal solution: $E(m,n)$

Base case: $E(0,0) = 0$, $E(i,0) = i$, $E(0,j) = j$

Filling the table

$$E(i,j) = \min\{1 + E(i-1,j), 1 + E(i,j-1), \mathrm{diff}(i,j) + E(i-1,j-1)\},$$

$$\begin{array}{ccccccccc}
E(0,0) & & E(0,1) & \cdots & & E(0,n-1) & & E(0,n) \\
& \searrow & \downarrow & & & & & \\
E(1,0) & \rightarrow & E(1,1) & \cdots & & \cdots & & \cdots \\
\vdots & & & & & & & \\
& & & & \searrow & \downarrow & \searrow & \downarrow \\
E(m-1,0) & & & & \rightarrow & E(m-1,n-1) & \rightarrow & E(m-1,n) \\
& & & & \searrow & \downarrow & \searrow & \downarrow \\
E(m,0) & & & & \rightarrow & E(m,n-1) & \rightarrow & E(m,n)
\end{array}$$

Running example

$x = ACGTA$ and $y = ATCTG$

$$
\begin{array}{cccccccccccc}
 & & & A & & T & & C & & T & & G \\
 & 0 & & 1 & & 2 & & 3 & & 4 & & 5 \\
 & & \searrow & & & & & & & & & \\
\mathrm{A} & 1 & & 0 & \rightarrow & 1 & & 2 & & 3 & & 4 \\
 & & & & \searrow & & \searrow & & & & & \\
\mathrm{C} & 2 & & 1 & & 1 & & 1 & & 2 & & 3 \\
 & & & & & & \searrow & \downarrow & & & & \\
\mathrm{G} & 3 & & 2 & & 2 & & 2 & & 2 & & 2 \\
 & & & & & & & & \searrow & & & \\
\mathrm{T} & 4 & & 3 & & 2 & & 3 & & 2 & & 3 \\
 & & & & & & & & & & \searrow & \\
\mathrm{A} & 5 & & 4 & & 3 & & 3 & & 3 & & 3
\end{array}
$$