Lecture 03

Date 9 Jan 2024

Longest Common Subsequence Problem.

$x_1 = A \quad x_2 = G \; --- \; x_5 = T$

→ A G C A T

→ G A T C

A G C

A C T

C G T ✗

G C T

A T

G A T

Notation :

A sequence $X = \langle x_1, \ldots, x_k \rangle$

$X_i := \langle x_1, \ldots, x_i \rangle$ prefix.

A subsequence of X is a sequence of the form $\langle x_{i_1}, x_{i_2}, \ldots, x_{i_r} \rangle$

$1 \le i_1 < i_2 < \cdots i_r \le k$

Def^n:-

Let X and Y be two sequences. We say the Z is a common subsequence of X & Y

if Z is subseq of X

& Z " " " Y.

Def^n:- We say that Z is a longest common subsequence of X & Y if for all common subsequence W, $|W| \leq |Z|$

$X = \langle A, T, G, C, A \rangle$ 5

$Y = \langle T, G, C, A, T, G, C \rangle$ 7

A, T, G, C 4

T, G, C A

Thm :- Let $X = \langle x_1, \ldots, x_m \rangle$ and $Y = \langle y_1, \ldots, y_n \rangle$ be two sequences. Let $Z = \langle z_1, \ldots, z_k \rangle$ be an LCS of X and Y.

① Let $x_m = y_n$, then $z_k = x_m = y_n$ and $Z_{k-1} = \langle z_1, \ldots, z_{k-1} \rangle$ is an lcs of X_{m-1} & Y_{n-1}.

② If $x_m \neq y_m$ and $z_k \neq x_m$ then Z is an LCS of X_{m-1} & Y.

③ If $x_m \neq y_m$ and $z_k \neq y_n$ then Z is an LCS of X & Y_{n-1}

Proof:- ① If $z_k \neq x_m$ then append $x_m (= y_n)$ at the end of Z to get a longer common ss. Contra!

Note Z_{k-1} is a css of X_{m-1} & Y_{n-1}

Suppose Z_{k-1} is not an LCS of " "

Let W be an LCS of X_{m-1} & Y_{n-1}

$$|Z_{k-1}| < |W|$$

$$\underbrace{\underline{Z_{k-1} + \langle x_m \rangle}}_{\overset{\displaystyle Z}{\shortparallel}} \quad \underbrace{W + \langle x_m \rangle}$$

$W + \langle x_m \rangle$ will be a common ss of X & Y of length longer than the length of Z

contra!

(2) $x_m \neq y_n$ & $z_k \neq x_m$. then

Z is css of X_{m-1} & Y

Suppose Z is not a LCS of X_{m-1} & Y

then <u>HW:</u> contra!

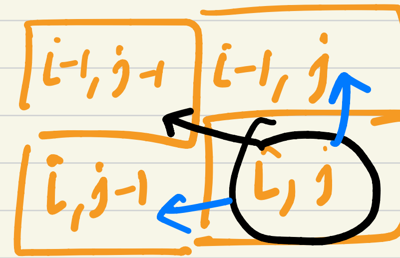
Let $c[i,j]$ = length of an LCS of X_i & Y_j

$$c[i,j] = \begin{cases} 0 & i=0 \text{ or } j=0 \\ c[i-1,j-1]+1 & i,j>0,\ x_m = y_n \\ \max\{c[i-1,j],\ c[i,j-1]\} & i,j>0,\ x_m \neq y_n \end{cases}$$

ε → string of length 0.

$$Z_{k-1} = LCS(X_{i-1}, Y_{j-1}) \rightarrow c[i-1,j-1]$$

i	$X \backslash Y$	ε	A	T	G	C	A
	j	0	1	2			
0	ε	0	0	0	0	0	0
1	T	0	0	1	1	1	1
2	G	0	0	1	2	2	2
3	C	0	0	1	2	3	3
4	A	0	1	1	2	3	4
	T	0					
	G	0					
	C	0					



$x_i = y_j$

$x_i \neq y_j$

HW:- Pseudocode [CLRS]

Optimal Binary Search Tree

Keys 1, 2, 3

How many BST's are there with keys 1, 2, 3 ?