

Longest Common Subsequence (LCS)

Chapter 15



4/29/11 CS380 Algorithm Design and Analysis 1

Longest Common Subsequence

- Problem: Let $x_1x_2\dots x_m$ and $y_1y_2\dots y_n$ be two sequences over some alphabet.
 - We assume they are strings of characters
- Find a longest common subsequence (LCS) of $x_1x_2\dots x_m$ and $y_1y_2\dots y_n$

4/29/11 CS380 Algorithm Design and Analysis 2

Example

- $x_1x_2x_3x_4x_5x_6x_7x_8 = b a c b f f c b$
- $y_1y_2y_3y_4y_5y_6y_7y_8y_9 = d a b e a b f b c$
- Longest Common Subsequence is:

A subsequence is a set of characters that appear in left- to-right order, but not necessarily consecutively.

4/29/11 CS380 Algorithm Design and Analysis 3

Dynamic Programming

- LCS can be solved using dynamic programming
 1. Characterize the structure of an optimal solution
 2. Recursively define the value of an optimal solution
 3. Compute the value of an optimal solution bottom-up
 4. Construct an optimal solution from the computed information

Step 1

- Characterizing a longest subsequence
- Optimal substructure: If $z = z_1z_2\dots z_p$ is a LCS of $x_1x_2\dots x_m$ and $y_1y_2\dots y_n$, then at least one of these must hold
 - $x_m = y_n$, and $z_1z_2\dots z_{p-1}$ is an LCS of $x_1x_2\dots x_{m-1}$ and $y_1y_2\dots y_{n-1}$,
 - $x_m \neq y_n$, and $z_1z_2\dots z_p$ is an LCS of $x_1x_2\dots x_{m-1}$ and $y_1y_2\dots y_n$,
 - $x_m \neq y_n$, and $z_1z_2\dots z_p$ is an LCS of $x_1x_2\dots x_m$ and $y_1y_2\dots y_{n-1}$.

Step 2: Recursive Solution

Let c_{ij} = length of LCS of $x_1x_2\dots x_i$ and $y = y_1y_2\dots y_j$.

$$c[i,j] = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0, \\ 1 + c[i-1,j-1] & \text{if } x_i = y_j, \\ \max(c[i-1,j], c[i,j-1]) & \text{if } x_i \neq y_j. \end{cases}$$

$$b[i,j] = \begin{cases} "_ \rightarrow" & \text{if } x_i = y_j, \\ "\uparrow" & \text{if } x_i \neq y_j \text{ and } c[i-1,j] \geq c[i,j-1], \\ "\leftarrow" & \text{if } x_i \neq y_j \text{ and } c[i-1,j] < c[i,j-1]. \end{cases}$$

We compute the $c[i,j]$ and $b[i,j]$ in order of increasing $i+j$, or alternatively in order of increasing i , and for a fixed i , in order of increasing j .

Step 3 & 4: Example

- $x_1x_2x_3x_4x_5x_6x_7x_8 = b a c b f f c b$
- $y_1y_2y_3y_4y_5y_6y_7y_8y_9 = d a b e a b f b c$

4/28/11

CS380 Algorithm Design and Analysis

7

Example

	0	1 d	2 a	3 b	4 e	5 a	6 b	7 f	8 b	9 c
0										
1 b										
2 a										
3 c										
4 b										
5 f										
6 f										
7 c										
8 b										

4/28/11

CS380 Algorithm Design and Analysis

8

Another Example

- What is the LCS in:
 - epidemiologist
 - refrigeration

4/29/11

CS380 Algorithm Design and Analysis

9

- Congratulations William and Kate!



Source: <http://www.bbc.co.uk>
