Longest Common Subsequence (LCS)

張智星 (Roger Jang)
jang@mirlab.org
http://mirlab.org/jang
多媒體資訊檢索實驗室
台灣大學 資訊工程系

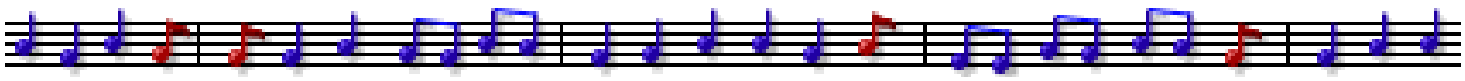
Longest Common Subsequence

- Subsequence
  - Given a string, we can delete some elements to form a subsequence:
    - s1=uvwxyz ➔ s2=uwyz (after deleting v and x)
    - s2 is a subsequence of s1.
- Longest common subsequence (LCS)
  - The similarity of two string can be define as the length of the LCS between them.
  - Example: abcdefg and xzackdfwgh have acdfg as a longest common subsequence

For word suggestion!

Brute-Force Approach to LCS

- A brute-force solution
  - Enumerate all subsequences of X
  - Test which ones are also subsequences of Y
  - Pick the longest one.
- Analysis:
  - If X is of length n, then it has 2^n subsequences
  - This is an exponential-time algorithm!



DP for LCS: 3-step Formula

Quiz!

Three-step DP formula for computing $lcs(\vec{A}, \vec{B})$

1. Optimum-value function

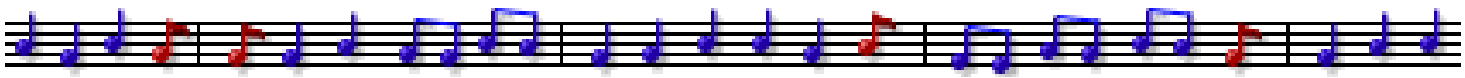
   $lcs(\vec{p}, \vec{q})$ is the length of LCS between string $\vec{p}$ and $\vec{q}$.

2. Recurrent formula

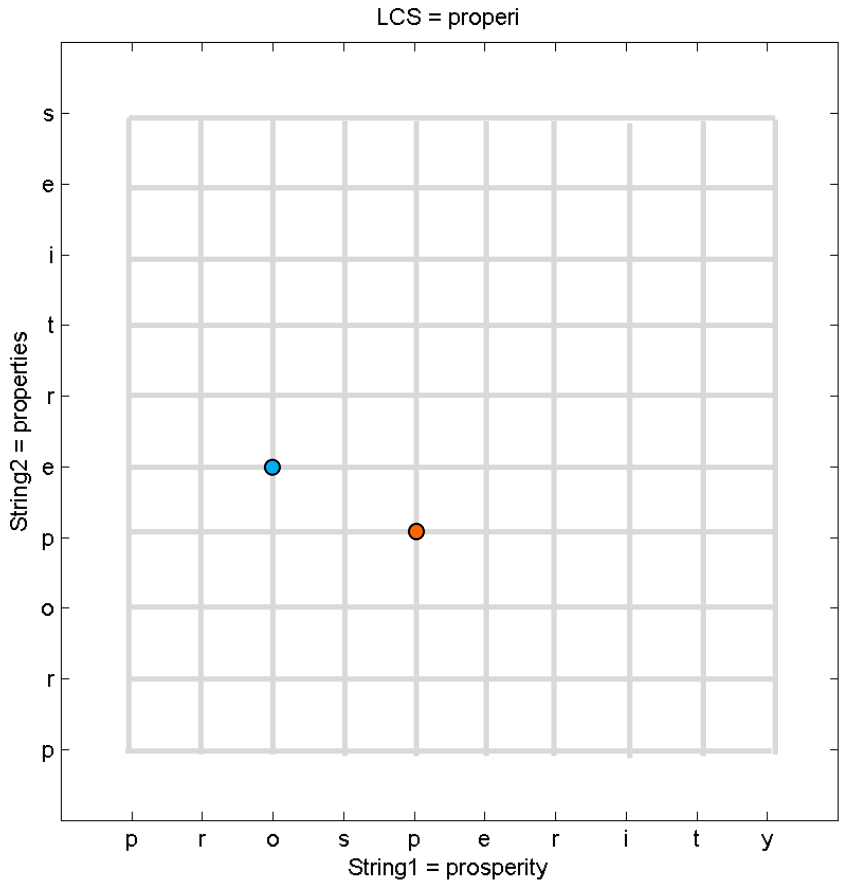
$$lcs(\vec{a}x, \vec{b}y) = \begin{cases} lcs(\vec{a}, \vec{b}) + 1, & \text{if } x = y \\ \max\begin{cases} lcs(\vec{a}x, \vec{b}) \\ lcs(\vec{a}, \vec{b}y) \end{cases}, & \text{if } x \neq y \end{cases}$$

   Boundary condition : $lcs(\vec{a}, [\,]) = lcs([\,], \vec{b}) = 0.$

3. Answer: $lcs(\vec{A}, \vec{B})$

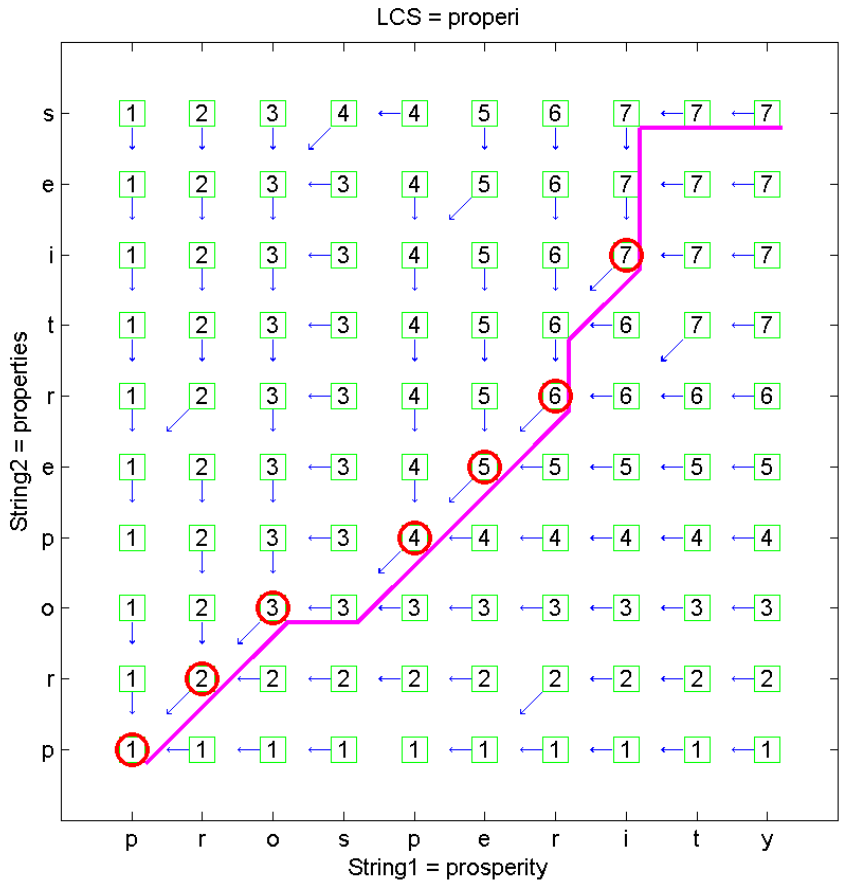


DP for LCS: Table Filling (1/2)



- $lcs(prosp, prop) = lcs(pros, prop) + 1$
- $lcs(pro, prope) = \max\begin{cases} lcs(pro, prop) \\ lcs(pr, prope) \end{cases}$

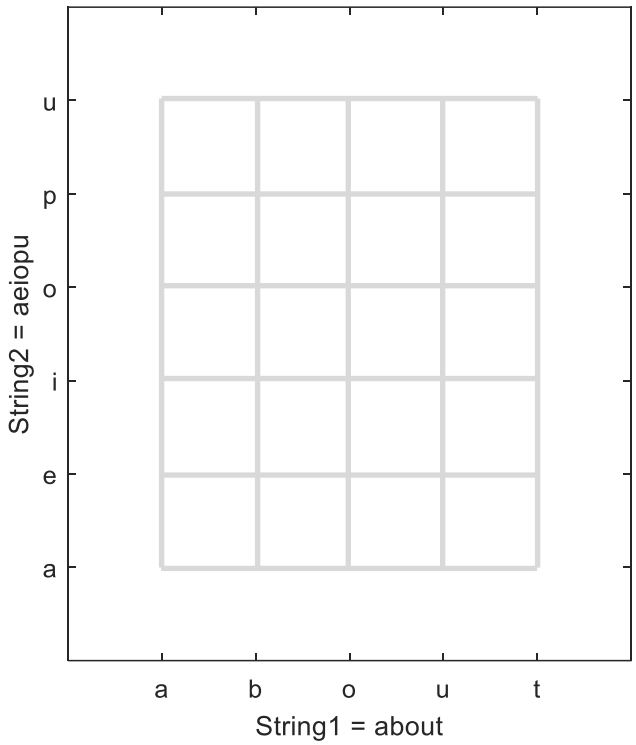
DP for LCS: Table Filling (2/2)



- Observations
  - LCS='properi' or 'propert' (which is obtained by keeping multiple back-tracking paths)
  - A match occurs when the node has a 45-degree back-tracking path

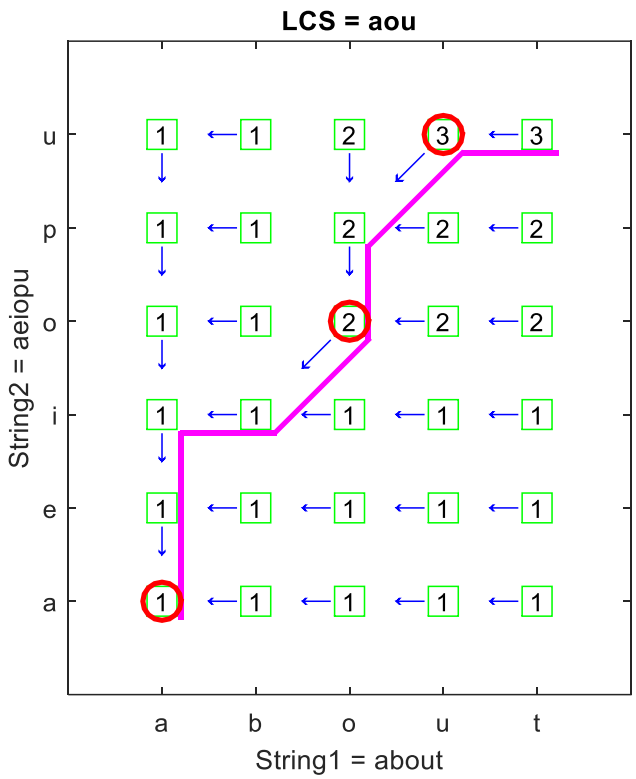
DP for LCS: Quiz for Table Filling

Quiz!



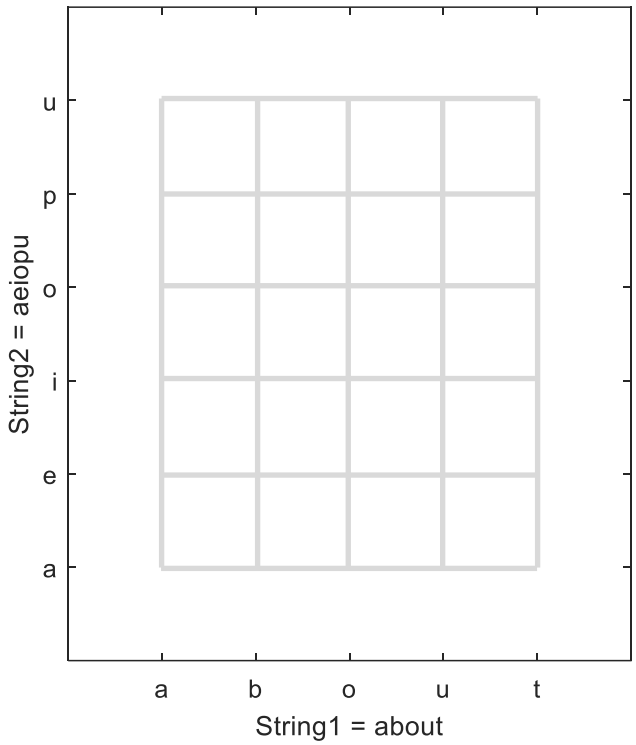
- Hints
  - Create a (m+1)*(n+1) matrix for table filling
  - Fill row 0 and column 0 with 0 first to establish the base cases of boundary conditions
  - Fill all the other elements in a layer-by-layer manner.

Quiz Solution



- To create this plot
  - Download Machine Learning Toolbox
  - Run lcs('about', 'aeiopu', 1) under MATLAB

LCS In Terms of Path Finding



- Note that all DP problem can be visualized as path finding…