

Design and Analysis of Algorithms

Mohammad GANJTABESH

mgtabesh@ut.ac.ir

School of Mathematics, Statistics and Computer Science, University of Tehran, Tehran, Iran.

Definition

For a given sequence $X=\langle x_1,x_2,\cdots,x_m\rangle$, the sequence $Z=\langle z_1,z_2,\cdots,z_k\rangle$ is a subsequence of X if there exists a strictly increasing sequence $\langle i_1,i_2,\cdots,i_k\rangle$ of indices of X such that for all $j=1,2,\cdots,k$, we have $x_{i_j}=z_j$.

Definition

For a given sequence $X=\langle x_1,x_2,\cdots,x_m\rangle$, the sequence $Z=\langle z_1,z_2,\cdots,z_k\rangle$ is a subsequence of X if there exists a strictly increasing sequence $\langle i_1,i_2,\cdots,i_k\rangle$ of indices of X such that for all $j=1,2,\cdots,k$, we have $x_{i_j}=z_j$.

Definition

Given two sequences X and Y, we say that a sequence Z is a common subsequence of X and Y if Z is a subsequence of both X and Y.

Definition

For a given sequence $X=\langle x_1,x_2,\cdots,x_m\rangle$, the sequence $Z=\langle z_1,z_2,\cdots,z_k\rangle$ is a subsequence of X if there exists a strictly increasing sequence $\langle i_1,i_2,\cdots,i_k\rangle$ of indices of X such that for all $j=1,2,\cdots,k$, we have $x_{i_j}=z_j$.

Definition

Given two sequences X and Y, we say that a sequence Z is a common subsequence of X and Y if Z is a subsequence of both X and Y.

Example

For $X = \langle e, a, b, c, d, e, f, g, h, a, a, d \rangle$ and $Y = \langle b, c, e, a, g, h, b, b, d, e \rangle$, the common subsequence is $Z = \langle b, g, h \rangle$. Another common subsequence is $Z = \langle b, c, e, g, h, d \rangle$.

Definition

Given two sequences X and Y, we say that a sequence Z is a longest common subsequence of X and Y if Z is a subsequence of both X and Y such that:

$$|Z| = max\{|Z'| \ni Z' \text{ is a subsequence of } X \text{ and } Y\}.$$

We denote it by Z = LCS(X, Y).

Definition

Given two sequences X and Y, we say that a sequence Z is a longest common subsequence of X and Y if Z is a subsequence of both X and Y such that:

$$|Z| = max\{|Z'| \ni Z' \text{ is a subsequence of } X \text{ and } Y\}.$$

We denote it by Z = LCS(X, Y).

Example

For $X = \langle e, a, b, c, d, e, f, g, h, a, a, d \rangle$ and

 $Y = \langle b, c, e, a, g, h, b, b, d, e \rangle$, the longest common subsequence is

 $Z = \langle b, c, e, g, h, d \rangle.$

Definition

Given two sequences X and Y, we say that a sequence Z is a longest common subsequence of X and Y if Z is a subsequence of both X and Y such that:

$$|Z| = max\{|Z'| \ni Z' \text{ is a subsequence of } X \text{ and } Y\}.$$

We denote it by Z = LCS(X, Y).

Example

For $X = \langle e, a, b, c, d, e, f, g, h, a, a, d \rangle$ and

 $Y = \langle b, c, e, a, g, h, b, b, d, e \rangle$, the longest common subsequence is

 $Z = \langle b, c, e, g, h, d \rangle.$

Definition

For a given sequence $X = \langle x_1, x_2, \dots, x_m \rangle$, the i^{th} prefix of X, for $i = 0, 1, \dots, m$, is defined as $X_i = \langle x_1, x_2, \dots, x_i \rangle$.

Theorem

Let $X = \langle x_1, x_2, \cdots, x_m \rangle$ and $Y = \langle y_1, y_2, \cdots, y_n \rangle$ be sequences, and let $Z = \langle z_1, z_2, \cdots, z_k \rangle$ be any longest common subsequence of X and Y. We have:

- 1. If $x_m = y_n$, then $z_k = x_m = y_n$ and $Z_{k-1} = LCS(X_{m-1}, Y_{n-1})$.
- 2. If $x_m \neq y_n$, then $z_k \neq x_m$ implies that $Z = LCS(X_{m-1}, Y)$.
- 3. If $x_m \neq y_n$, then $z_k \neq y_n$ implies that $Z = LCS(X, Y_{n-1})$.

Theorem

Let $X = \langle x_1, x_2, \cdots, x_m \rangle$ and $Y = \langle y_1, y_2, \cdots, y_n \rangle$ be sequences, and let $Z = \langle z_1, z_2, \cdots, z_k \rangle$ be any longest common subsequence of X and Y. We have:

- 1. If $x_m = y_n$, then $z_k = x_m = y_n$ and $Z_{k-1} = LCS(X_{m-1}, Y_{n-1})$.
- 2. If $x_m \neq y_n$, then $z_k \neq x_m$ implies that $Z = LCS(X_{m-1}, Y)$.
- 3. If $x_m \neq y_n$, then $z_k \neq y_n$ implies that $Z = LCS(X, Y_{n-1})$.

Proof.

The proof is based on contradiction...

Proof.

- If z_k ≠ x_m, then we could append x_m = y_n to Z to obtain a common subsequence of X and Y of length k + 1 (Contradiction!). Thus, we must have z_k = x_m = y_n.
 Now, we show that the prefix Z_{k-1} (of length k 1) is LCS(X_{m-1}, Y_{n-1}). Suppose that there is a common subsequence W of X_{m-1} and Y_{n-1} with length greater than k 1. Then, appending x_m = y_n to W produces a common subsequence of X and Y whose length is greater than k, which is a contradiction.
- 2. If $z_k \neq x_m$, then Z is a common subsequence of X_{m-1} and Y. If there were a common subsequence W of X_{m-1} and Y with length greater than k, then W would also be a common subsequence of X_m and Y, contradicting the assumption that Z is an LCS(X,Y).
- 3. Similar to case 2.

Let $X = \langle x_1, x_2, \dots, x_m \rangle$ and $Y = \langle y_1, y_2, \dots, y_n \rangle$ be sequences. Also let $C[i,j] = |LCS(X_i, Y_i)|$. Now we can write C[i,j] as follows:

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

Let $X = \langle x_1, x_2, \dots, x_m \rangle$ and $Y = \langle y_1, y_2, \dots, y_n \rangle$ be sequences. Also let $C[i,j] = |LCS(X_i, Y_j)|$. Now we can write C[i,j] as follows:

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

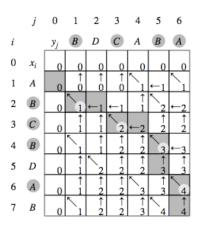
The solution can be constructed correspondingly by the following matrix:

$$B[i,j] = \begin{cases} \leftarrow \text{ or } \uparrow & \text{if } i = 0 \text{ or } j = 0, \\ \nwarrow & \text{if } i,j > 0 \text{ and } x_i = y_j, \\ \leftarrow \text{ or } \uparrow & \text{if } i,j > 0 \text{ and } x_i \neq y_i. \end{cases}$$

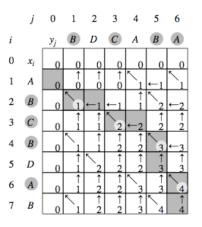
Longest Common Subsequence: Algorithm

```
LCS-LENGTH(X, Y)
    m \leftarrow length[X]
 2 n \leftarrow length[Y]
 3 for i \leftarrow 1 to m
            do c[i,0] \leftarrow 0
     for j \leftarrow 0 to n
 6
            do c[0, i] \leftarrow 0
      for i \leftarrow 1 to m
 8
            do for j \leftarrow 1 to n
 9
                      do if x_i = y_i
10
                             then c[i, j] \leftarrow c[i - 1, j - 1] + 1
                                   b[i, i] \leftarrow "\\"
11
12
                             else if c[i-1, j] \ge c[i, j-1]
13
                                      then c[i, j] \leftarrow c[i-1, j]
14
                                             b[i, j] \leftarrow "\uparrow"
15
                                      else c[i, j] \leftarrow c[i, j-1]
                                             b[i, j] \leftarrow "\leftarrow"
16
17
      return c and b
```

Longest Common Subsequence: Algorithm



Longest Common Subsequence: Algorithm



```
PRINT-LCS(b, X, i, j)

1 if i = 0 or j = 0

2 then return

3 if b[i, j] = ``````

4 then PRINT-LCS(b, X, i - 1, j - 1)

5 print x_i

6 elseif b[i, j] = ``\uparrow``

7 then PRINT-LCS(b, X, i, j - 1)

8 else PRINT-LCS(b, X, i, j - 1)
```

Exercises

- 1. Determine an LCS of $\langle 1,0,0,1,0,1,0,1\rangle$ and $\langle 0,1,0,1,1,0,1,1,0\rangle.$
- 2. Give an $O(n^2)$ -time algorithm to find the longest monotonically increasing subsequence of a sequence of n numbers.
- Give an O(nlog n)-time algorithm to find the longest monotonically increasing subsequence of a sequence of n numbers.

