Design and Analysis of Algorithms

Mohammad GANJTABESH

mgtabesh@ut.ac.ir

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

Longest Common Subsequence

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 $x_{i_j} = z_j$.

Longest Common Subsequence

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.

Longest Common Subsequence

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

Longest Common Subsequence

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 $Z = LCS(X, Y)$.

Longest Common Subsequence

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

Longest Common Subsequence

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

Longest Common Subsequence

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* $Z = LCS(X, Y_{n-1})$.

Longest Common Subsequence

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... $\square$

Longest Common Subsequence

Proof.

1. If $z_k \neq x_m$, 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.

$\square$

Longest Common Subsequence

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

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

Longest Common Subsequence

Let $X = \langle x_1, x_2, \cdots, x_m \rangle$ and $Y = \langle y_1, y_2, \cdots, y_n \rangle$ 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:

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

Longest Common Subsequence: Algorithm

```
LCS-LENGTH(X, Y)
m ← length[X]
n ← length[Y]
for i ← 1 to m
      do c[i, 0] ← 0
for j ← 0 to n
      do c[0, j] ← 0
for i ← 1 to m
      do for j ← 1 to n
              do if x_i = y_j
                    then c[i, j] ← c[i − 1, j − 1] + 1
                         b[i, j] ← "↖"
                    else if c[i − 1, j] ≥ c[i, j − 1]
                            then c[i, j] ← c[i − 1, j]
                                 b[i, j] ← "↑"
                            else c[i, j] ← c[i, j − 1]
                                 b[i, j] ← "←"
return c and b
```

Longest Common Subsequence: Algorithm

	j	0	1	2	3	4	5	6
i		y_j	B	D	C	A	B	A
0	x_i	0	0	0	0	0	0	0
1	A	0	↑ 0	↑ 0	↑ 0	↖ 1	← 1	↖ 1
2	B	0	↖ 1	← 1	← 1	↑ 1	↖ 2	← 2
3	C	0	↑ 1	↑ 1	↖ 2	← 2	↑ 2	↑ 2
4	B	0	↖ 1	↑ 1	↑ 2	↑ 2	↖ 3	← 3
5	D	0	↑ 1	↖ 2	↑ 2	↑ 2	↑ 3	↑ 3
6	A	0	↑ 1	↑ 2	↑ 2	↖ 3	↑ 3	↖ 4
7	B	0	↖ 1	↑ 2	↑ 2	↑ 3	↖ 4	↑ 4

Longest Common Subsequence: Algorithm

i	j	0	1	2	3	4	5	6
		y_j	B	D	C	A	B	A
0	x_i	0	0	0	0	0	0	0
1	A	0	↑ 0	↑ 0	↑ 0	↖ 1	← 1	↖ 1
2	B	0	↖ 1	← 1	← 1	↑ 1	↖ 2	← 2
3	C	0	↑ 1	↑ 1	↖ 2	← 2	↑ 2	↑ 2
4	B	0	↖ 1	↑ 1	↑ 2	↑ 2	↖ 3	← 3
5	D	0	↑ 1	↖ 2	↑ 2	↑ 2	↑ 3	↑ 3
6	A	0	↑ 1	↑ 2	↑ 2	↖ 3	↑ 3	↖ 4
7	B	0	↖ 1	↑ 2	↑ 2	↑ 3	↖ 4	↑ 4

```
PRINT-LCS(b, X, i, j)
    if i = 0 or j = 0
        then return
    if b[i, j] = "↖"
        then PRINT-LCS(b, X, i − 1, j − 1)
             print x_i
    elseif b[i, j] = "↑"
        then PRINT-LCS(b, X, i − 1, j)
    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.
3. Give an $O(n \log n)$-time algorithm to find the longest monotonically increasing subsequence of a sequence of n numbers.