Longest Common Subsequence

input: A pair $I = (A[1..m], B[1..n])$ of sequences.

output: The length of the longest common subsequence of A and B. (That is, the longest sequence that is a subsequence of both A and B.)

Here is the algorithm. Fix any input instance $(A[1..m], B[1..n])$.

subproblems. For each pair (i, j) with $i \in \{0, 1, \ldots, m\}$ and $j \in \{0, 1, \ldots, n\}$, define $L[i, j]$ to be the length of the longest common subsequence between $A[1..i]$ and $B[1..j]$. (The first i elements in A, and the first j elements in B.)

recurrence. $L[i, j] = 0$ if $0 \in \{i, j\}$, and, otherwise

$$L[i, j] = \begin{cases} 1 + L[i-1, j-1] & \text{if } A[i] = B[j], \\ \max(L[i-1, j], L[i, j-1]) & \text{if } A[i] \neq B[j]. \end{cases}$$

The intuition is as follows. In the case that $A[i] = B[j]$, there is a longest common subsequence of $A[1..i]$ and $B[1..j]$ that ends in the common final element $A[i]$, and before that final element consists of a longest common subsequence of $A[1..i-1]$ and $B[1..j-1]$ (of length $L[i-1, j-1]$). So in that case, $L[i, j] = 1 + L[i-1, j-1]$.

In the case that $A[i] \neq B[j]$, every common subsequence y of $A[1..i]$ and $B[1..j]$ either (i) doesn't end with $A[i]$, or (ii) doesn't end with $B[j]$. In Case (i), y is a common subsequence of $A[1..i-1]$ and $B[1..j]$. In Case (ii), y is a common subsequence of $A[1..i]$ and $B[1..j-1]$. Thus, the set of common subsequences of $A[1..i]$ and $B[1..j]$ is the union of two subsets:

(i) the set of common subsequences of $A[1..i-1]$ and $B[1..j]$, and

(ii) the set of common subsequences of $A[1..i]$ and $B[1..j-1]$.

So the length of the longest common subsequence of $A[1..i]$ and $B[1..j]$ is the length of the longest sequence in these two subsets. That is, $L[i, j] = \max(L[i-1, j], L[i, j-1])$.

For a careful long-form proof that the recurrence is correct, see Lemma 1 at the end of this note.

As usual, the algorithm solves each subproblem $L[i, j]$ by evaluating the right-hand side of the recurrence. It considers the subproblems in order of increasing size, so each subproblem occurring in the right-hand side has already been computed when $L[i, j]$ is computed. As mentioned above, the algorithm returns $L[m, n]$.

time. There are $(n + 1) \cdot (m + 1) = O(nm)$ subproblems, and for each, the right-hand side of the recurrence takes constant time to evaluate (assuming the smaller subproblems are already solved). So the total time is $O(nm)$.

correctness. A long-form proof that the recurrence relation is correct is below (Lemma 1). Correctness of the algorithm follows by induction (on the size of the subproblem).

This algorithm is equivalent to running the Longest Path algorithm on an a graph with a vertex v_{ij} for each subproblem $L[i, j]$, and appropriately chosen edges of weight 1 or 0.

Lemma 1. *The recurrence relation for $L[i,j]$ (*LONGEST COMMON SUBSEQUENCE*) is correct.*

Proof (long form).

1. Consider any pair (i,j) with $i \in \{0,\ldots,m\}$ and $j \in \{0,\ldots,n\}$.
2.1. Case 1. Consider the case that $0 \in \{i,j\}$. ($A[1..i]$ or $B[1..j]$ is the empty sequence.)
2.2. The only common subsequence is the empty one, so $L[i,j] = 0$, and the recurrence is correct.
3.1. Case 2. Otherwise $i \geq 1$ and $j \geq 1$.
3.2.1. Case 2.1. First consider the sub-case when $A[i] = B[j]$.
3.2.2. We first show that $L[i,j] \geq L[i-1,j-1]+1$.
3.2.3. Let $x = (x_1, x_2, \ldots, x_\ell)$ be a common subsequence of $A[1..i-1]$ and $B[1..j-1]$, of length $L[i-1,j-1]$ (it exists by definition of L).
3.2.4. Let $x' = (x_1, x_2, \ldots, x_\ell, A[i])$ be obtained by appending $A[i]$ to x.
3.2.5. Since $A[i] = B[j]$, x' is a subsequence of both $A[1..i]$ and $B[1..j]$.
3.2.6. So $L[i,j] \geq |x'| = L[i-1,j-1]+1$.
3.2.7. Next we show $L[i,j] \leq L[i-1,j-1]+1$.
3.2.8. Let $y = (y_1, y_2, \ldots, y_k)$ be a common subsequence of $A[1..i]$ and $B[1..j]$, of length $L[i,j]$.
3.2.9. Let $y' = (y_1, y_2, \ldots, y_{k-1})$ be obtained from y by removing the last element.
3.2.10. Sequence y is a subsequence of $A[1..i]$, so y' is a subsequence of $A[1..i-1]$. (Verify!)
3.2.11. Similarly, y' is a subsequence of $B[1..j-1]$.
3.2.12. So y' is a common subsequence of $A[1..i-1]$ and $B[1..j-1]$.
3.2.13. So $L[i-1,j-1]$ is at least $|y'| = L[i,j]-1$. Rearranging gives $L[i,j] \leq L[i-1,j-1]+1$.
3.2.14. By this and Step 3.2.6, $L[i,j] = L[i-1,j-1]+1$.
3.2.15. So $L[i,j]$ equals the right-hand side of the recurrence in Case 2.1 (when $A[i] = B[j]$).
3.3.1. Case 2.2. In the remaining sub-case $A[i] \neq B[j]$.
3.3.2. First we show $L[i,j] \geq \max(L[i-1,j], L[i,j-1])$.
3.3.3. Let z be a common subsequence of $A[1..i-1]$ and $B[1..j]$ of length $L[i-1,j]$.
3.3.4. Then z is also a common subsequence of $A[i]$ and $B[j]$.
3.3.5. So $L[i,j] \geq |z| = L[i-1,j]$.
3.3.6. By a symmetric argument, $L[i,j] \geq L[i,j-1]$. So $L[i,j] \geq \max(L[i-1,j], L[i,j-1])$.
3.3.7. Next we show that $L[i,j] \leq \max(L[i-1,j], L[i,j-1])$.
3.3.8. Let $y = (y_1, y_2, \ldots, y_\ell)$ be a common subsequence of $A[1..i]$ and $B[1..j]$ of length $L[i,j]$.
3.3.9.1. Case 2.2.1 First consider the case that $y_\ell \neq A[i]$.
3.3.9.2. Then y must be a subsequence of $A[1..i-1]$ (as well as $B[1..j]$).
3.3.9.3. So $L[i-1,j] \geq |y| = L[i,j]$. That is $L[i,j] \leq L[i-1,j]$.
3.3.9.4. So $L[i,j] \leq \max(L[i-1,j], L[i,j-1])$.
3.3.10.1. Case 2.2.2 Otherwise (since $A[i] \neq B[j]$) it must be that $y_\ell \neq B[j]$.
3.3.10.2. By an argument symmetric to the one in Case 2.2.1, $L[i,j] \geq L[i,j-1]$.
3.3.10.3. So $L[i,j] \leq \max(L[i-1,j], L[i,j-1])$.
3.3.11. By Blocks 3.3.9 and 3.3.10, $L[i,j] \leq \max(L[i-1,j], L[i,j-1])$.
3.3.12. By this and Step 3.3.6, $L[i,j] = \max(L[i-1,j], L[i,j-1])$.
3.3.13. So $L[i,j]$ equals the right-hand side of the recurrence in Case 2.2 ($A[i] \neq B[j]$).
3.4. By Blocks 3.2 and 3.3, $L[i,j]$ equals the right-hand side of the recurrence in Case 2 ($0 \notin \{i,j\}$).
4. By Blocks 2 and 3, $L[i,j]$ equals the right-hand side of the recurrence (for any i, j). $\square$

External resources on LONGEST COMMON SUBSEQUENCE

- CLRS Chapter 15
- previous CS 5800 lecture video
  - https://neu.tegrity.com/#/recording/2bed4358-f1ca-43d2-90f4-5980ae9d2a2d?startTime=372331