

MAIOR SUBSEQUÊNCIA COMUM

Programação dinâmica

Carlos Eduardo
Julio Cesar Rocha

Pablo Felipe
Victor Le Roy

O PROBLEMA

O problema da maior subsequência comum também conhecido como “*longest common subsequence (LCS)*”, é encontrar a maior subsequência comum a todas as sequências em um conjunto de sequências (geralmente apenas duas sequências).

O LCS possui grande aplicação na área de bioinformática (análise de uma sequência de DNA) e na área de processamento de linguagem natural, usado também como um sistema de revisão para gerenciadores de versão (como o Git).

A SOLUÇÃO

- Subestrutura ótima
 - Pode ser subdivido em subproblemas menores e mais simples
- Subproblemas sobrepostos
 - Solução de problemas mais difíceis que surgem a partir de um mais fáceis

Problemas com essas duas propriedades são passíveis de abordagens de programação dinâmica, nas quais as soluções dos subproblemas são memorizadas, ou seja, as soluções dos subproblemas são salvas para reutilização.

A SOLUÇÃO - PARTE 2

Para a solução, começamos criando uma matriz (inicializada com 0) para armazenar os valores que serão usados como referência para a solução. O tamanho da maior subsequência comum será o último valor da matriz criada. A atualização dos valores da matriz é feita analisando se o item X da 1ª sequência é igual ao item Y da 2ª sequência, se forem iguais, então o valor do atual elemento será o valor do elemento à esquerda do elemento acima do atual elemento somado 1, caso sejam diferentes, o valor do elemento atual será o maior valor entre o elemento a sua esquerda e o elemento acima.

O PSEUDOCÓDIGO (SOLUCIONANDO O LCS)

X , Y = Strings escolhidas para solução do problema

T1 , T2 = Tamanho das strings

LCS = Matriz contendo as strings e solução do problema

```
T1 ← tamanho X
```

```
T2 ← tamanho Y
```

```
LCS [T2+1][T1+1] ← 0    # Inserir 0 em todas as posições
```

```
ENQUANTO i < T1 + 1 :
```

```
    ENQUANTO j < T2 + 1 :
```

```
        SE i = 0 OU j = 0 :
```

```
            LCS[i][j] ← 0
```

```
        OU SE X[i-1] = Y[j-1] :
```

```
            LCS[i][j] ← LCS[i-1][j-1] + 1
```

```
        SE NÃO :
```

```
            LCS[i][j] ← max( LCS[i-1][j] e LCS[i][j-1] )
```

```
RETORNE LCS [T1][T2]    # Indica a maior sequência na matriz
```

O PSEUDOCÓDIGO - PARTE 2 (DEMONSTRAR A SEQUÊNCIA)

X , Y = Strings escolhidas para solução do problema

Index = Maior “tamanho” da matriz LCS

T1 , T2 = Tamanho das strings

LCS_S = Lista contendo a strings solução

```
T1 ← tamanho X
T2 ← tamanho Y
Index ← LCS [T1][T2]
LCS_S [Index + 1] ← “ ”    # Inserir espaços em todas posições
i ← T1, j ← T2
```

```
ENQUANTO i > 0 E j > 0:
    SE X[i-1] = Y[j-1]:
        LCS_S[i][j] ← X[i-1]
        i, j, Index -= 1
    OU SE LCS [i-1][j] > LCS [i][j-1] :
        i -= 1
    SE NÃO :
        j -= 1
```

```
RETORNE LCS_S # Retorna a lista
```

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	0	0
B	0	0	0	0	0	0
C	0	0	0	0	0	0
B	0	0	0	0	0	0

$T1 \leftarrow \text{tamanho } X$

$T2 \leftarrow \text{tamanho } Y$

$LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Criação da matriz e seta todas as posições com valor de 0;

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	0	0
B	0	0	0	0	0	0
C	0	0	0	0	0	0
B	0	0	0	0	0	0

$T1 \leftarrow \text{tamanho } X$

$T2 \leftarrow \text{tamanho } Y$

$LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$ 

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Para todas as execuções em vermelho:

Execução do laço de repetição ($i = 1, j = [1 \dots 3]$):

“SE NÃO :” é verdadeiro;

 Verifica-se qual valor máximo entre os blocos adjacentes (esquerda, cima)

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	0
B	0	0	0	0	0	0
C	0	0	0	0	0	0
B	0	0	0	0	0	0

$T1 \leftarrow \text{tamanho } X$

$T2 \leftarrow \text{tamanho } Y$

$LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$ 

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Execução do laço de repetição ($i = 1, j = 4$):

“A” comum entre as duas sequências ;

“**OU SE** $X[i-1] = Y[j-1]$:” é verdadeiro;

$LCS[i][j] = LCS[i-1][j-1] + 1$; # Posição de cima a esquerda de “A”

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	1
B	0	0	0	0	0	0
C	0	0	0	0	0	0
B	0	0	0	0	0	0

$T1 \leftarrow \text{tamanho } X$
 $T2 \leftarrow \text{tamanho } Y$
 $LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$ ←

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Execução do laço de repetição ($i = 1, j = 5$):

 “**SE NÃO** :” é verdadeiro;

 Verifica-se qual valor máximo entre os blocos adjacentes (esquerda, cima)

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	1
B	0	1	0	0	0	0
C	0	0	0	0	0	0
B	0	0	0	0	0	0

$T1 \leftarrow \text{tamanho } X$

$T2 \leftarrow \text{tamanho } Y$

$LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$ 

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Execução do laço de repetição ($i = 2, j = 1$):

“B” comum entre as duas sequências ;

“**OU SE** $X[i-1] = Y[j-1]$:” é verdadeiro;

$LCS[i][j] = LCS[i-1][j-1] + 1$; # Posição de cima a esquerda de “B”

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	1
B	0	1	1	1	1	0
C	0	0	0	0	0	0
B	0	0	0	0	0	0

$T1 \leftarrow \text{tamanho } X$
 $T2 \leftarrow \text{tamanho } Y$
 $LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$ 

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Para todas as execuções em vermelho:

Execução do laço de repetição ($i = 2, j = [2...4]$):

“**SE NÃO** :” é verdadeiro;

Verifica-se qual valor máximo entre os blocos adjacentes (esquerda, cima)

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	1
B	0	1	1	1	1	2
C	0	0	0	0	0	0
B	0	0	0	0	0	0

$T1 \leftarrow \text{tamanho } X$

$T2 \leftarrow \text{tamanho } Y$

$LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$ 

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Execução do laço de repetição ($i = 2, j = 5$):

“B” comum entre as duas sequências ;

“**OU SE** $X[i-1] = Y[j-1]$:” é verdadeiro;

$LCS[i][j] = LCS[i-1][j-1] + 1$; # Posição de cima a esquerda de “B”

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	1
B	0	1	1	1	1	2
C	0	1	1	0	0	0
B	0	0	0	0	0	0

$T1 \leftarrow \text{tamanho } X$

$T2 \leftarrow \text{tamanho } Y$

$LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$ 

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Para todas as execuções em vermelho:

Execução do laço de repetição ($i = 3, j = [1...2]$):

“SE NÃO :” é verdadeiro;

 Verifica-se qual valor máximo entre os blocos adjacentes (esquerda, cima)

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	1
B	0	1	1	1	1	2
C	0	1	1	2	0	0
B	0	0	0	0	0	0

$T1 \leftarrow \text{tamanho } X$
 $T2 \leftarrow \text{tamanho } Y$
 $LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições


ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$ 

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Execução do laço de repetição ($i = 3, j = 3$):

“C” comum entre as duas sequências ;

“**OU SE** $X[i-1] = Y[j-1]$:” é verdadeiro;

$LCS[i][j] = LCS[i-1][j-1] + 1$; # Posição de cima a esquerda de “C”

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	1
B	0	1	1	1	1	2
C	0	1	1	2	2	2
B	0	0	0	0	0	0

$T1 \leftarrow \text{tamanho } X$
 $T2 \leftarrow \text{tamanho } Y$
 $LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$ ←

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Para todas as execuções em vermelho:

Execução do laço de repetição ($i = 3, j = [1...2]$):

 “**SE NÃO** :” é verdadeiro;

 Verifica-se qual valor máximo entre os blocos adjacentes (esquerda, cima)

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	1
B	0	1	1	1	1	2
C	0	1	1	2	2	2
B	0	1	0	0	0	0

$T1 \leftarrow \text{tamanho } X$
 $T2 \leftarrow \text{tamanho } Y$
 $LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições


ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$ 

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Execução do laço de repetição ($i = 4, j = 1$):

 “B” comum entre as duas sequências ;

 “**OU SE** $X[i-1] = Y[j-1]$:” é verdadeiro;

$LCS[i][j] = LCS[i-1][j-1] + 1$; # Posição de cima a esquerda de “B”

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	1
B	0	1	1	1	1	2
C	0	1	1	2	2	2
B	0	1	1	2	2	0

$T1 \leftarrow \text{tamanho } X$
 $T2 \leftarrow \text{tamanho } Y$
 $LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$ 

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Para todas as execuções em vermelho:

Execução do laço de repetição ($i = , j = [2...4]$):

“SE NÃO :” é verdadeiro;

 Verifica-se qual valor máximo entre os blocos adjacentes (esquerda, cima)

EXEMPLO DE EXECUÇÃO

		B	D	C	A	B
	0	0	0	0	0	0
A	0	0	0	0	1	1
B	0	1	1	1	1	2
C	0	1	1	2	2	2
B	0	1	1	2	2	3

$T1 \leftarrow \text{tamanho } X$
 $T2 \leftarrow \text{tamanho } Y$
 $LCS[T2+1][T1+1] \leftarrow 0$ # Inserir 0 em todas as posições

ENQUANTO $i < T1 + 1$:

ENQUANTO $j < T2 + 1$:

SE $i = 0$ **OU** $j = 0$:

$LCS[i][j] \leftarrow 0$

OU SE $X[i-1] = Y[j-1]$:

$LCS[i][j] \leftarrow LCS[i-1][j-1] + 1$ 

SE NÃO :

$LCS[i][j] \leftarrow \max(LCS[i-1][j] \text{ e } LCS[i][j-1])$

RETORNE $LCS[T1][T2]$ # Indica a maior sequência na matriz

Execução do laço de repetição ($i = 4, j = 5$):

 “B” comum entre as duas sequências ;

 “**OU SE** $X[i-1] = Y[j-1]$:” é verdadeiro;

$LCS[i][j] = LCS[i-1][j-1] + 1$; # Posição de cima a esquerda de “B”

REFERÊNCIAS

Simulador: <https://www.cs.usfca.edu/~galles/visualization/DPLCS.html>

Programiz: <https://www.programiz.com/dsa/longest-common-subsequence>

Wikipedia: [https://en.wikipedia.org/wiki/Longest_common_subsequence_problem#Computing the length of the LCS](https://en.wikipedia.org/wiki/Longest_common_subsequence_problem#Computing_the_length_of_the_LCS)

Geeksforgeeks: <https://www.geeksforgeeks.org/longest-common-subsequence-dp-4/>

Marcos Castro: <https://www.youtube.com/watch?v=2brK-02vB6M>