CALCUL D'UN ALIGNEMENT LOCAL

Crochemore p257

Considérons et des fonctions de Σ dans $\mathbb{Z}^{-*}$ fixées.
Considérons aussi une fonction de Σ^2 dans $\mathbb{Z}$ ≥ 0 sur la diagonale
< 0 en dehors.

Soient $x = x_1 \ldots x_n$ et $y = y_1 \ldots y_m$ deux mots sur Σ.

On introduit $s_{i,j} = \max \{ Sim(x_k \ldots x_i, y_\ell \ldots y_j) \mid k \in [1..i+1], \ell \in [1..j+1] \}$
= le score d'un alignem^t local de $x_1 \ldots x_i$, $y_1 \ldots y_j$ finissant en i,j et maximal.

NB : On considère des suffixes plutôt que des facteurs quelconques en vue de prolonger ces alignem^t locaux.
Si le score retenu était réalisé par un alignem^t d'un facteur finissant quelques caractères auparavant, cela n'aurait pas de sens de chercher à le prolonger d'une ins, sub ou del ...

Pté

- $\begin{cases} \forall i \in [0..n] & s_{i,0} = 0 \\ \forall j \in [0..m] & s_{0,j} = 0 \end{cases}$

- $$\boxed{\forall (i,j) \in [1..n] \times [1..m] \quad s_{i,j} = \max \begin{cases} 0 \\ s_{i-1,j} + del(x_i) \\ s_{i,j-1} + ins(y_j) \\ s_{i-1,j-1} + sub(x_i, y_j) \end{cases}}$$

Preuve a). Puisque $y_1 \ldots y_0 = \varepsilon$, on cherche en fait ici à aligner un suffixe de x avec le mot vide.
Pour tout facteur non vide on aura un coût de suppression, donc un score < 0.
Il vaut mieux ici ne rien aligner, ce qui fait un score nul.
D'où $s_{i,0} = 0$.
Idem pour $s_{0,j} = 0$.

b) Notons $M_0 = 0$, $M_1 = s_{i-1,j} + del(x_i)$, $M_2 = s_{i,j-1} + ins(y_j)$, $M_3 = s_{i-1,j-1} + sub(x_i, y_j)$.
et $M = \max(M_0, M_1, M_2, M_3)$.

Montrons d'abord que $s_{i,j} \geq M$

- Puisque l'option de ne rien aligner est toujours possible, $s_{i,j} \geq 0 = M_0$ nécessairem^t.

- Considérons $(\tilde{u},\tilde{v})$ un alignem^t justifiant $s_{i-1,j}$, c-à-d que $(\tilde{u},\tilde{v})$ est un alignem^t de $(x_k - x_{i-1},\ y_\ell - y_j)$ pour certains k et ℓ, et que $S(\hat{u},\hat{v}) = s_{i-1,j}$.
Alors $(\tilde{u}x_i,\ \tilde{v}-)$ est un alignem^t de $(x_k - x_i,\ y_\ell - y_j)$ donc
$s_{i,j} \geq S(\tilde{u}x_i,\ \tilde{v}-) = S(\hat{u},\hat{v}) + \underline{del}(x_i) = s_{i-1,j} + \underline{del}(x_i)$. D'où $\underline{s_{i,j} \geq M_1}$

- De m̂ en considérant $(\tilde{u},\tilde{v})$ justifiant $s_{i,j-1}$, $(\tilde{u}-,\tilde{v}y_j)$ est un aligm^t de $(x_k - x_i,\ y_\ell - y_j)$ pour certains k et j donc $s_{i,j} \geq S(\hat{u}-,\tilde{v}y_j) = \underbrace{S(\hat{u},\hat{v})}_{= s_{i,j-1}} + \underline{ins}(y_j)$
D'où $\underline{s_{i,j} \geq M_2}$

- Encore de m̂ on a $s_{i,j} \geq s_{i-1,j-1} + sub(x_i, y_j)$ soit $\underline{s_{i,j} \geq M_3}$ en considérant $(\hat{u},\hat{v})$ justifiant $s_{i-1,j-1}$.

Donc $\underline{s_{i,j} \geq \max(M_0, M_1, M_2, M_3) = M}$.

Réciproquement montrons que $s_{i,j} \leq M$ (ce qui montre finalem^t l'égalité ds chaque cas)
Considérons $(\tilde{u},\tilde{v})$ un alignement de $(x_k - x_i, y_\ell - y_j)$ justifiant $s_{i,j}$, c-à-d tel que $S(\hat{u},\hat{v}) = s_{i,j}$. Notons $r = |\tilde{u}| = |\tilde{v}|$.

- Si $r = 0$, alors $s_{i,j} = S(\varepsilon,\varepsilon) = 0$. Comme on a exhibé plus haut des alignm^t de suffixe de $(x_\cdot - x_i,\ y_\cdot - y_j)$ de score M_1, M_2 ou M_3, le fait que $s_{i,j} = 0$ implique néc que $M_1 \leq 0$, $M_2 \leq 0$ et $M_3 \leq 0$. Donc $\max(M_0, M_1, M_2, M_3) = 0$ et alors on a bien $\underline{s_{i,j} = 0 = M}$.

- Si $\underline{\tilde{u}_r = -}$, alors néc $\hat{v}_r = y_j$ donc $(\hat{u},\hat{v}) := (\tilde{u}_1 - \tilde{u}_{r-1},\ \tilde{v}_1 \text{———} \tilde{v}_{r-1})$ est un alignement de $(x_k - x_i,\ y_\ell - y_{j-1})$ donc $S(\hat{u},\hat{v}) \leq s_{i,j-1}$.
Or $S(\hat{u},\tilde{v}) = S(\hat{u},\hat{v}) + \underline{ins}(y_j)$ donc $\underline{s_{i,j} \leq s_{i,j-1} + \underline{ins}(y_j)} = M_2$
or $s_{i,j} \geq M_2$ donc $s_{i,j} = M_2$ et néc $\begin{matrix} M_2 \geq M_0 \\ M_2 \geq M_1 \\ M_2 \geq M_3 \end{matrix}$ aut dit $M_2 = M$.
donc $\underline{s_{i,j} = M}$

- Si $\underline{\tilde{v}_r = -}$ alors $\tilde{u}_r = x_i$ et $(\hat{u},\hat{v}) := (\tilde{u}_1 - \tilde{u}_{r-1},\ \tilde{v}_1 - \tilde{v}_{r-1})$ est un aligm^t de $(x_k - x_{i-1}, y_\ell - y_j)$
donc $\underline{s_{i,j} = S(\tilde{u},\tilde{v}) = S(\hat{u},\hat{v}) + \underline{del}(x_i) \leq s_{i-1,j} + \underline{del}(x_i) = M_1}$
or $s_{i,j} \geq M_1$, donc $s_{i,j} = M_1$ et alors néc $M_1 = M$ d'où $\underline{s_{i,j} = M}$

- Sinon $\tilde{u}_r = x_i$, $\tilde{v}_r = y_j$ et $(\hat{u},\hat{v}) := (\tilde{u}_1 - \tilde{u}_{r-1},\ \tilde{v}_1 - \tilde{v}_{r-1})$ est un
donc $\underline{s_{i,j}} = S(\hat{u},\tilde{v}) = S(\hat{u},\hat{v}) + \underline{sub}(x_i, y_j) \leq \underbrace{s_{i-1,j-1} + \underline{sub}(x_i, y_j)}_{= M_3}$
or $s_{i,j} \geq M_3$ donc $s_{i,j} = M_3$ et alors néc.
$M = M_3$, d'où $\underline{s_{i,j} = M}$

$\square$

Algo

LOC(x, y)

$n \leftarrow$ longueur (x)
$m \leftarrow$ longueur (y)

$T \leftarrow$ Tableau indexé par $[0..n] \times [0..m]$
initialisé à 0

Pour j allant de 1 à m
  Pour i allant de 1 à n
    $T[i,j] \leftarrow \max \{ T[i-1,j-1] + sub(x_i, y_j) \; ; \; T[i-1,j] + del(x_i), \; T[i,j-1] + ins(y_j), \; 0 \}$

max $\leftarrow 0$
Pour j allant de 1 à m
  Pour i allant de 1 à n
    Si $T[i,j]$ > max
    alors max $\leftarrow T[i,j]$

Retourner max

Rq

* Ici les cas de base sont compris dans la mise à 0 du tableau T puisque $s_{i,0} = s_{0,j} = 0$

⚠ Ici à ne pas renvoyer $T[n,m]$.
En effet le meilleur alignement local n'a aucune raison de se trouver à la fin, de se terminer en x_n et y_m.

On a stocké dans $T[i,j]$ la valeur $s_{i,j}$ du meilleur alignement en terme de score, d'un suffixe de $x_1 - x_i$ et d'un suffixe de $y_1 - y_j$, parce que c'est ce dont on a besoin pour prolonger.

Il faut donc chercher le max en parcourant toutes les de positions de fin possible.

Complexité La complexité temporelle est en $O(n \times m)$
——— spatiale ———

Algobis

Pour renvoyer l'alignem^t local et non plus seulem^t son score.

LOC_BIS (x, y)

$n \leftarrow$ longueur (x)
$m \leftarrow$ ______ (y).

$T \leftarrow$ tableau indicé par $[0..n] \times [0..m]$ initialisé à 0
$M \leftarrow$ ______

Pour j allant de 1 à m
  Pour i allant de 1 à n
    $RAZ \leftarrow 0$
    $INS \leftarrow T[i, j-1] + ins(y_j)$
    $DEL \leftarrow T[i-1, j] + del(x_i)$
    $SUB \leftarrow T[i-1, j-1] + sub(x_i, y_j)$
    Si $\max(RAZ, INS, DEL, SUB) = RAZ$
    alors $T[i,j] \leftarrow 0$
      $M[i,j] \leftarrow (-1,-1)$
    Si ______ $= INS$
    alors $T[i,j] \leftarrow INS$
      $M[i,j] \leftarrow (i, j-1)$
    Si ______ $= DEL$
    alors $T[i,j] \leftarrow DEL$
      $M[i,j] \leftarrow (i-1, j)$
    Si ______ $= SUB$
    alors $T[i,j] \leftarrow SUB$
      $M[i,j] \leftarrow (i-1, j-1)$

$i_m \leftarrow 0$; $j_m \leftarrow 0$, res. $\leftarrow$ mot vide
Pour j allant de 1 à m
  Pour i ______ 1 ______ n
    Si $T[i,j] > T[i_m, j_m]$
    alors $i_m \leftarrow i$; $j_m \leftarrow j$

Pour i allant de 0 à n
  $M[i, 0] \leftarrow (-1,-1)$

Pour j allant de 1 à m
  $M[0, j] \leftarrow (-1,-1)$

$\tilde{x} \leftarrow \varepsilon$; $i \leftarrow i_m$
$\tilde{y} \leftarrow \varepsilon$, $j \leftarrow j_m$

Tant que $M[i,j] \neq (-1,-1)$
  match $M[i,j]$ with
  | $(i, j-1) \rightarrow \begin{cases} \tilde{x} \leftarrow -\tilde{x} \\ \tilde{y} \leftarrow y_j \tilde{y} \end{cases}$
  | $(i-1, j) \rightarrow \begin{cases} \tilde{x} \leftarrow x_i \tilde{x} \\ \tilde{y} \leftarrow -\tilde{y} \end{cases}$
  | $(i-1, j-1) \rightarrow \begin{cases} \tilde{x} \leftarrow x_i \tilde{x} \\ \tilde{y} \leftarrow y_j \tilde{y} \end{cases}$
  $(i,j) \leftarrow M[i,j]$
retourner $(\tilde{x}, \tilde{y}, T[i_m, j_m])$

Cplx
complexité spatiale $O(n \times m)$
______ temporelle $O(n \times m)$

Rq Pour favoriser le plus long alignement on veille à mettre le test du cas d'une remise à zéro (RAZ) en premier.

En effet si la possibilité d'un score nulle est offerte par un align^t qu'on allonge d'un pas INS, DEL ou SUB, c'est cette hypothèse qu'on retiendra dans le tableau mémoire M, puisque dans l'un de ces cas on écrase la valeur précédem^t inscrite par le cas RAZ.

En fait les tests sont présentés par ordre inverse de priorité, l'idée étant que le dernier est avantagé puisqu'il l'emporte.

Ainsi on privilégie 1 la SUB, 2 LA DEL, 3 L'INS, 4 RAZ.

ex

del $\equiv -1$ ins (voyelle) $= -1$, (consonne) $= -2$,

sub (voyelle, voyelle') $= -1$
(consonne, consonne') $= -1$
(voyelle, consonne) $= -2$
(voyelle, m̂ voyelle) $= 3$
(consonne, m̂ consonne) $= 1$.

x = GRAVITER $n = 8$

y = AGRANDIR $m = 8$

	/	A	G	R	A	N	D	I	R
/	0	0	0	0	0	0	0	0	0
G	0	0	1	0	0	0	0	0	0
R	0	0	0	2	1	0	0	0	1
A	0	3	1	1	5	3	1	0	0
V	0	2	2	0	4	4	2	1	0
I	0	1	0	0	3	2	2	$\boxed{5}$	3
T	0	0	0	0	2	2	1	4	4
E	0	0	0	0	1	0	0	3	3
R	0	0	0	1	0	0	0	2	4

meilleur al. loc. $\begin{pmatrix} G & R & A & \_ & V & I \\ G & R & A & N & D & I \end{pmatrix}$, de score 5.

1 1 3 -2 -1 3

autre meilleur al. local, de score aussi 5 $\begin{pmatrix} G & R & A \\ G & R & A \end{pmatrix}$

1 1 3