Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2023
Prof. Dr. Jens Stoye · Tizian Schulz

`https://gi.cebitec.uni-bielefeld.de/teaching/2023summer/sa`

Übungsblatt 5 vom 4.5.2023
Abgabe am 11.5.2023 bis 12:00 Uhr (mittags)

Aufgabe 1 (Affine Gapkosten) (8 Punkte)

1. Warum sollten bei affinen Gapkosten die Gap-open-Kosten d nicht niedriger gewählt werden als die Gap-extension-Kosten e, also $d \geq e$?
2. Zeige, dass affine Gapkosten subadditiv sind, also dass gilt: $g(\ell_1 + \ell_2) \leq g(\ell_1) + g(\ell_2)$
3. Berechne die Matrizen **S**, **V** und **H** der Sequenzen $x =$ `TGAAA` und $y =$ `TTCCGA` für globales Alignment mit affinen Gapkosten mit Hilfe des Gotoh-Algorithmus.

   Verwende dabei $\begin{cases} \text{Score für M\textsc{atch}} & = 3 \\ \text{Score für M\textsc{ismatch}} & = -2 \end{cases}$ $\begin{cases} \text{Kosten für G\textsc{ap-open}}: & d = 3 \\ \text{Kosten für G\textsc{ap-extension}}: & e = 1 \end{cases}$

   Gib alle optimalen Alignments an.

Aufgabe 2 (Suboptimale lokale Alignments) (6 Punkte)

1. Warum ist man bei der Bestimmung von suboptimalen Alignments in erster Linie an *nichtüberlappenden* Alignments interessiert? Welche Probleme möchte man vermeiden?
2. Gegeben seien die Sequenzen $x =$ `TTGATCTC` und $y =$ `AGACGTTA` sowie die folgende Score-Funktion:

$$\text{M\textsc{atch}} = 3\ ; \quad \text{M\textsc{ismatch}} = -2\ ; \quad \text{I\textsc{ndel}} = -2$$

   (a) Berechne die Alignment-Matrix für lokales Alignment mit dem Smith-Waterman-Algorithmus und gib ein optimales Alignment an.
   (b) Berechne die Matrix für nichtüberlappende, suboptimale lokale Alignments nach *Waterman-Eggert* und gib das erste suboptimale Alignment an.
   (c) Aktualisiere die Matrix und gib das nächste suboptimale Alignment an.

Aufgabe 3 (Forward-Backward Technik) (6 Punkte)

Gegeben seien die Sequenzen $x =$ `ATGCAATC` und $y =$ `CTCAGAT`. Verwende **Einheitskosten** für globale Alignments.

1. Berechne die Edit-Matrix D und die Edit-Matrix D^{rev} der reversen Strings.
2. Berechne die Gesamtkostenmatrix T und die Zusatzkostenmatrix C.
3. Gib ein optimales globales Alignment von x und y an und markiere es in D, D^{rev}, T und C.

Aufgabe 4 (Paarweises Alignment mit linearem Speicherbedarf) (8 Punkte)

Zeige anhand der Sequenzen aus Aufgabe 3, wie das optimale Alignment berechnet werden kann, während nur linearer Speicherplatz benötigt wird. Simuliere dazu die Schritte des in der Vorlesung besprochenen Divide-and-Conquer Verfahrens auf Kopien der Edit-Matrix. (Es kann teilweise auf die in Aufgabe 2 berechneten Werte zurückgegriffen werden.)

1. Markiere in einer neuen, leeren Matrix der entsprechenden Größe, für welche Bereiche D und D^{rev} berechnet werden würden bzw. berechne die Werte wenn nötig neu und trage sie ein.
2. Trage die Werte von T für die mittlere Zeile $m' = \lceil m/2 \rceil$ ein. Markiere ein Minimum in T und die entsprechende Spalte, an der die Matrix geteilt wird.
3. Führe rekursiv die oberen Schritte für die beiden Teilmatrizen aus, bis jeweils nur noch eine Zeile übrig ist.

Du kannst diese Seite für deine Lösung verwenden:

Aufgabe 1: $\begin{cases} \text{Score:} & \textsc{Match} = 3\,, \quad \textsc{Mismatch} = -2 \\ \text{Kosten:} & \textsc{Gap-open}: d = 3\,, \quad \textsc{Gap-extension}: e = 1 \end{cases}$

V	ε	T	T	C	C	G	A
ε							
T							
G							
A							
A							
A							

S	ε	T	T	C	C	G	A
ε							
T							
G							
A							
A							
A							

H	ε	T	T	C	C	G	A
ε							
T							
G							
A							
A							
A							

Du kannst diese Seite für deine Lösung verwenden:

Aufgabe 2: Score-Funktion: $\begin{cases} \textsc{Match} = 3 \\ \textsc{Mismatch} = -2 \\ \textsc{Indel} = -2 \end{cases}$

2 (a) Erste Matrix, mit Smith-Waterman berechnet:

	ϵ	T	T	G	A	T	C	T	C
ϵ									
A									
G									
A									
C									
G									
T									
T									
A									

2 (b) Zweite Matrix, nach *Waterman-Eggert* berechnet:

	ϵ	T	T	G	A	T	C	T	C
ϵ									
A									
G									
A									
C									
G									
T									
T									
A									

2 (c) Dritte Matrix, nochmal nach *Waterman-Eggert* berechnet:

	ϵ	T	T	G	A	T	C	T	C
ϵ									
A									
G									
A									
C									
G									
T									
T									
A									

Du kannst diese Seite für deine Lösung verwenden:

Aufgabe 3: Einheitskosten

1.

D	ϵ	C	T	C	A	G	A	T
ϵ								
A								
T								
G								
C								
A								
A								
T								
C								

D^{r}	C	T	C	A	G	A	T	ϵ
A								
T								
G								
C								
A								
A								
T								
C								
ϵ								

2.

T	ϵ	C	T	C	A	G	A	T
ϵ								
A								
T								
G								
C								
A								
A								
T								
C								

C	ϵ	C	T	C	A	G	A	T
ϵ								
A								
T								
G								
C								
A								
A								
T								
C								