

Vorlesung: Spezielle Algorithmen auf Sequenzen
Sommersemester 2005

Übungen

Übung 2, Diskussion: 05.05.2005.

1. Edit-Matrix.

Gegeben seien zwei Strings $u = \text{GTCCA}$ und $v = \text{GCCAA}$. Berechnen Sie die Edit-Matrix des Algorithmus zur Berechnung der Editdistanz gemäß folgender Kostenfunktionen und visualisieren Sie die minimierenden Pfade.

(a) Einheitskosten:

$$\delta(\alpha \rightarrow \beta) = \begin{cases} 0 & : \alpha, \beta \in \mathcal{A} \wedge \alpha = \beta \\ 1 & : \text{sonst} \end{cases}$$

(b) Hamming-Kosten:

$$\delta(\alpha \rightarrow \beta) = \begin{cases} 0 & : \alpha, \beta \in \mathcal{A} \wedge \alpha = \beta \\ 1 & : \alpha, \beta \in \mathcal{A} \wedge \alpha \neq \beta \\ \infty & : \text{sonst} \end{cases}$$

2. Anzahl optimaler Alignments.

Die minimierenden Pfade in einer $(n \times m)$ Edit-Matrix entsprechen den optimalen Alignments. Demnach ist es möglich die Anzahl der verschiedenen optimalen Alignments zu erhalten, indem die Anzahl dieser Pfade berechnet wird.

Entwickeln Sie einen Algorithmus, der diese Anzahl in $O(nm)$ Zeit berechnet.

Hinweis: Benutzen Sie dynamische Programmierung.

3. Edit-Distanz.

Sei $\mathbf{A} = \begin{pmatrix} T & G & A & G & A & G & - & - & - & A \\ T & - & A & C & G & G & T & C & C & - \end{pmatrix}$ ein Alignment.

(a) Berechnen Sie die *gewichteten Alignmentkosten* $D(\mathbf{A})$ mit affinen Gapkosten. Die affinen Gapkosten sind gegeben durch $gapinit = 3$ und $gapextend = 1$ und die Kostenfunktion durch die Matrix:

δ	A	C	G	T
A	0	4	1	4
C	4	0	4	2
G	1	4	0	5
T	4	2	5	0

(b) Berechnen Sie für das gleiche Alignment den gewichteten Ähnlichkeitsscore $S(\mathbf{A})$ mit affinen Gapkosten. Die affinen Gapkosten sind gegeben durch $gapinit = -3$ und $gapextend = -1$ und der Ähnlichkeitsscore durch die Matrix:

σ	A	C	G	T
A	3	-1	1	-1
C	-1	5	-1	2
G	1	-1	3	-1
T	-1	2	-1	5

- (c) Berechnen Sie für das gleiche Alignment den gewichteten Ähnlichkeitsscore $S(\mathbf{A})$ mit allgemeinen Gapkosten. Die Gapkosten $g(l)$ sind gegeben durch

1	1	2	3	4	5	6
	-10	-10	-3	-11	-11	-4

und der Ähnlichkeitsscore ist gegeben durch die Matrix

σ	A	C	G	T
A	3	-1	1	-1
C	-1	5	-1	2
G	1	-1	3	-1
T	-1	2	-1	5

4. Implementieren Sie Gotohs Algorithmus zur Berechnung von paarweisen Alignments mit affinen Gapkosten.