

Übung zu Algorithmen auf Sequenzen

Blatt 12

Ausgabe: 26.01.2017 Besprechung: 02.02.2017

Aufgabe 12.1

Überprüfen Sie, ob und wo sich die Sequenzen $s = \text{tactgaccgca}$ und $t = \text{cccattag}$ mit folgenden Scoringwerten überlappen (free end gap Alignment):

- Match: +2
- Mismatch: -1
- InDel (Gap): -2

Erstellen Sie hierfür auch das entsprechende Alignment, indem Sie die Traceback-Matrix mitberechnen.

Aufgabe 12.2

Aus dem universellen Alignment-Algorithmus für zwei Strings s und t und den vier Graphtopologien für verschiedene Alignment-Varianten lassen sich vier konkrete Alignment-Algorithmen ableiten.

Stets sei $S[i, j]$ der Score des optimalen Pfades von v_o nach (i, j) , $\gamma > 0$ seien die Gapkosten (Score $-\gamma$) und $score(a, b)$ sei der Ähnlichkeitsscore der Zeichen a und b aus dem Alphabet.

Jeder konkrete Algorithmus besteht aus den Punkten:

- Bedeutung von $S[i, j]$ im konkreten Fall
- Berechnungsvorschrift für $S[i, j]$, dabei in der Regel Fallunterscheidung zwischen $i = j = 0$, $i = 0 < j$, $j = 0 < i$ und $i, j > 0$.
- Berechnung des Ergebnisses

Im Fall des globalen Alignments gilt:

- $S[i, j]$ ist der optimale Alignment-Score des Präfixes der Länge i von s mit dem Präfix der Länge j von t .
- $S[0, 0] = 0$,
 $S[i, 0] = -\gamma i$ für $i > 0$,
 $S[0, j] = -\gamma j$ für $j > 0$,
 $S[i, j] = \max\{S[i-1, j-1] + score(s_{i-1}, t_{j-1}), S[i-1, j] - \gamma, S[i, j-1] - \gamma\}$.
- Der optimale globale Alignment-Score ist $S[m, n]$.

Stellen Sie entsprechende konkrete Algorithmen für die anderen drei Alignment-Varianten auf (semiglobal = Mustersuche, Überlappsuche = free end gaps, lokal).