

Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WS 2011/2012

Dr. Alexander Sczyrba · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011winter/SequenzAnalyse>

Übungsblatt 8 vom 23.12.2011

Abgabe in der Woche nach den Weihnachtsferien

Aufgabe 1 Alignment-Algorithmen

(3 Punkte)

1. Erkläre die Unterschiede zwischen dem *Needleman-Wunsch* und dem *Smith-Waterman* Algorithmus.
2. Für welche biologischen Fragestellungen würdest Du welchen Algorithmus anwenden?

Aufgabe 2 Globales Alignment

(4 Punkte)

Wir definieren die folgenden Scores für die Editoperationen Insertion und Deletion: $\mathcal{I}_c = \mathcal{D}_c = -1$. Für Copy (\mathcal{C}) und Substitution (\mathcal{S}) gilt die Scorematrix $\mathcal{S}_{a,c}$:

	A	B	C
A	+2	-1	-1
B	-1	+2	-1
C	-1	-1	+2

1. Zeichne den globalen Alignmentgraphen für $x = BC$ und $y = AB$ mit *allen* Kanten und Knoten.
2. Füge deiner Zeichnung die Kantenbeschriftungen λ und die Kantengewichte w hinzu.
3. Berechne die Knotenwerte $S(v)$.
4. Gib ein optimales globales Alignment an und markiere dessen Pfad im Graphen.

Aufgabe 3 Lokales Alignment

(5 Punkte)

Gegeben sind die Sequenzen $x = ATCCTATC$ und $y = TTCCTAAAT$. Es sollen die **SCORES** $\mathcal{C} = +2$, $\mathcal{S}_{a,c} = -3$ für $a \neq c$ und $\mathcal{I} = \mathcal{D} = -2$ verwendet werden.

1. Beschreibe (in Worten und mithilfe einer groben Skizze) den Alignmentgraphen von x und y für die Berechnung eines lokalen Alignments mithilfe des Smith-Waterman Algorithmus.
2. Berechne den Score und gib ein optimales lokales Alignment an. Verwende dazu nicht den Graphen, sondern stattdessen eine Matrix, ähnlich der Editmatrix.

Bitte wenden!

Bonusaufgaben (für spannende Weihnachten!)

Die Bearbeitung dieser Aufgaben ist freiwillig. Du kannst dir damit zusätzliche Punkte erarbeiten.

Aufgabe 4 Gapkosten

(3 Bonuspunkte)

1. Erkläre, welche Arten von Gapkostenfunktionen es gibt.
2. Warum sind andere Funktionen sinnvoller als lineare Gapkostenfunktionen? Erkläre dazu den biologischen Hintergrund.

Aufgabe 5 Scorematrizen

(4 Bonuspunkte)

Informiere dich eigenständig über die PAM(t) und BLOSUM(s) Scorematrizen für Proteinsequenzen. Gib die Quellen an, die du benutzt.

1. Fasse in eigenen Worten zusammen, wie diese Scorematrizen aus Proteinsequenzen erstellt werden können.
2. Beschreibe im Detail, wofür die Parameter t und s stehen.
3. Zähle die Unterschiede von PAM und BLOSUM auf.

Aufgabe 6 Globales Alignment mit free end gaps

(5 Bonuspunkte)

Gegeben sind die Sequenzen $x = \text{CCAGGTCG}$ und $y = \text{AGGA}$. Es sollen die **SCORES** $\mathcal{C} = 2$, $\mathcal{S}_{a,c} = 1$ für $a \neq c$ und $\mathcal{I} = \mathcal{D} = 0.5$ verwendet werden.

1. Zeichne den Alignmentgraphen schematisch für das free end gap Alignment.
2. Berechne das globale Alignment mit free end gaps von x und y (in einer Matrix).
3. Gib eine optimale Edit-Sequenz und das dazugehörige optimale Alignment an.

Wir wünschen euch frohe Weihnachten und einen
guten Rutsch ins neue Jahr!