

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2020

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga · Leonie R. Brockmann · Rebecca K. Pfeil

<https://gi.cebitec.uni-bielefeld.de/teaching/2020summer/sa>

Zusätzliches Übungsblatt 11 vom 02.07.2020

Abgabe am 05.07.2020 bis 24:00 Uhr per Email an jens.stoye@uni-bielefeld.de

Aufgabe 1 (Center-Star-Approximation)

Gegeben sind die Sequenzen $s_1 = \text{ATACT}$, $s_2 = \text{ATCT}$ und $s_3 = \text{GTGT}$. Benutze für deine folgenden Berechnungen Einheitskosten.

1. Berechne die *Center-Sequenz* s_c .
2. Erstelle das multiple Alignment A_c und gib seine Sum-of-Pairs-Kosten an.
3. Was kannst du mithilfe deiner Lösung über die Kosten des optimalen Alignments sagen?
4. Beschreibe in eigenen Worten die Laufzeit- und Speicherplatzkomplexität der Center-Star-Approximation. Unterscheide dabei zuerst die einzelnen Phasen und erkläre dann das Gesamtergebnis.

Aufgabe 2 (Divide-and-Conquer-Alignment)

Gegeben sind die Sequenzen $s_1 = \text{CTC}$, $s_2 = \text{AC}$ und $s_3 = \text{AT}$. Benutze für deine Berechnungen Einheitskosten.

1. Was ist der Unterschied zwischen einem optimalen und einem C-optimalen Schnitt?
2. Erstelle die Zusatzkostenmatrizen für die drei Sequenzen.
3. Gib alle C-optimalen Schnitte an.
4. Nenne alle möglichen optimalen multiplen Alignments, die dir die C-optimalen Schnitte angeben. Gib auch ihre Kosten an.
5. Wieso verkleinert sich der Suchraum von $\mathcal{O}(n^k 2^k)$ beim Sum-of-Pairs-optimalen multiplen Alignment auf $\mathcal{O}(n^{k-1})$ beim Divide-and-Conquer-Alignment?

Aufgabe 3 (Baumalignment)

1. Ein optimales Baumalignment für die Sequenzen s_1, \dots, s_k entspricht im Sequenzraum (V, E) einem minimalen Steinerbaum für die Teilmenge $\{s_1, \dots, s_k\} \subseteq V$.
 - (a) Wie sind dabei V und E definiert?
 - (b) Wie sind die Kantengewichte gewählt?
2. Für die innere Minimierung der Formel für das Baumalignment kann der Fitch-Algorithmus verwendet werden. Finde eine optimale Beschriftung der inneren Knoten des Baums auf der Rückseite unter Einheitskosten. Trage auch Zwischenergebnisse in den Baum ein.

