



Blatt 12

Aufgabe 12.1 Gegeben sei die Levensthein-Distanz als Kostenfunktion. Ein globales Alignment zweier Strings werde mit der Technik des dynamischen Programmierens bestimmt. Zeigen Sie, dass für die Einträge $D_{i,j}$ der dabei generierten Matrix gilt

$$\begin{aligned} |D_{i,j} - D_{i,j-1}| &\leq 1 && \text{für } i \geq 0, j \geq 1 \\ |D_{i,j} - D_{i-1,j}| &\leq 1 && \text{für } i \geq 1, j \geq 0 \\ D_{i,j} - D_{i-1,j-1} &\geq 0 && \text{für } i \geq 1, j \geq 1 \end{aligned}$$

(3 Punkte)

Aufgabe 12.2 Wir betrachten lokale Alignments. Statt der Ähnlichkeit verwenden wir eine Kostenfunktion der folgenden Form:

$$\delta(a, b) = \begin{cases} 0, & a = b \\ > 0, & \text{sonst} \end{cases}$$

Stellen Sie eine Hypothese über die so erhaltenen optimalen lokalen Alignments auf und begründen Sie deren Korrektheit. (2 Punkte)

Aufgabe 12.3 Unter einem approximativen Pair eines Strings S verstehen wir ein Alignment zweier Teilstrings $S_1 = S[i_1..j_1]$ und $S_2 = S[i_2..j_2]$, wobei ausgeschlossen sei, dass ein Zeichen (an der selben Position) mit sich selbst aligniert wird.

Beispiel: Für $S = xaxb$ mit $S_1 = xax$ und $S_2 = xb$ kann man die Alignments

$$\begin{array}{cccc} x & a & x & - \\ - & - & x & b \end{array} \quad \text{und} \quad \begin{array}{ccc} x & a & x \\ x & b & - \end{array}$$

angeben, wobei das erste Alignment kein approximatives Pair darstellt, da $S(3)$ mit sich selbst aligniert wird, wohingegen das zweite Alignments ein approximatives Pair ist.

Ein optimales approximatives Pair ist ein approximatives Pair mit maximaler Ähnlichkeit für das Alignment.

Entwickeln Sie einen Algorithmus, der zu String S ein optimales approximatives Pair bestimmt. Hinweis: Nutzen Sie die DP-Matrix für das lokale Alignment. Die Hauptdiagonale (alle Knoten (i, j) mit $i = j$) spielt dabei eine wichtige Rolle. (3 Punkte)

Aufgabe 12.4 Gegeben sei die Levensthein-Distanz als Kostenfunktion. Im Alignment-Graphen bezeichnen wir die Menge der Knoten (i, j) mit $|i - j| \leq k$ als Schlauch der Breite k um die Hauptdiagonale (alle Knoten (i, j) mit $i = j$).

Zeigen Sie:

- (a) Für Strings S_1, S_2 mit $|S_1| = |S_2|$ und einer Distanz $D(S_1, S_2) \leq 2k$ verbleibt der Pfad des optimalen Alignments im Schlauch der Breite k um die Hauptdiagonale. (2 Punkte)
- (b) Für Strings S_1, S_2 mit einer Distanz $D(S_1, S_2) \leq k$ verbleibt der Pfad des optimalen Alignments im Schlauch der Breite k um die Hauptdiagonale. (2 Punkte)
- (c) Zusatzaufgabe: Formulieren Sie eine schärfere Abschätzung, die die Knoten oberhalb und unterhalb der Hauptdiagonalen getrennt betrachtet. (2 Zusatzpunkte)

Abgabe per E-Mail an grau@informatik.uni-halle.de oder in Raum 4.12.