

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2018

Dr. Daniel Dörr

<https://gi.cebitec.uni-bielefeld.de/teaching/2018summer/sa>

Übungsblatt 2 vom 17.04.2018

Abgabe am 24.04.2018 vor Beginn der Vorlesung

Aufgabe 1 (Diskrete Metrik)

(3 Punkte)

Auf jeder Menge \mathcal{X} lässt sich die diskrete Metrik $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ definieren:

$$d(x, y) := \begin{cases} 0, & \text{wenn } x = y, \\ 1, & \text{wenn } x \neq y. \end{cases}$$

Weise nach, dass es sich bei d um eine Metrik auf \mathcal{X} handelt, indem Du die Definitionen (3.1) - (3.3) aus dem Skript (S. 14) überprüfst.

Aufgabe 2 (Edit-Operationen und Edit-Sequenz)

(4 Punkte)

Gegeben sei der String $x = \text{ZEICHEN}$.

1. Berechne den String $y = E(x, \mathcal{S}_{KCDSDGCLD})$, gib alle Zwischenschritte an. Wie hoch sind die Kosten der Edit-Sequenz? Gib eine weitere Edit-Sequenz mit maximalen Kosten an, die den gleichen String y ergibt.
2. Gib mit Hilfe der Edit-Funktion die kürzeste Edit-Sequenz an, die x in y überführt.

Aufgabe 3 (Edit-Distanzen)

(4 Punkte)

Die Rekurrenz zur Berechnung der Standard-Edit-Distanz mit Einheitskosten lautet für $1 \leq i \leq |x|, 1 \leq j \leq |y|$:

$$D(i, j) = \min \begin{cases} D(i-1, j-1) + \mathbb{1}_{\{x[i] \neq y[j]\}} \\ D(i-1, j) + 1 \\ D(i, j-1) + 1 \end{cases}$$

Die Rekursionsbasis ist gegeben durch:

$$D(0, j) = j \text{ für } 0 \leq j \leq |y| \text{ und } D(i, 0) = i \text{ für } 0 \leq i \leq |x|$$

1. Wie sehen die Rekurrenz und die Basisfälle für die *Hamming-Distanz* aus?
2. Wie sehen die Rekurrenz und die Basisfälle für die *LCS-Distanz* aus?
3. Ist es möglich, die *Edit+Flip-Distanz* rekursiv zu berechnen? Wenn nein, warum nicht? Wenn ja, gib eine Formel und eine Rekursionsbasis an. Welche Schwierigkeiten könnten zudem hierbei auftreten?

Aufgabe 4 (Berechnung der Standard-Edit-Distanz)

(4 Punkte)

Gegeben seien die Strings $x = \text{ACGAT}$ und $y = \text{CAGTA}$. Benutze zur Berechnung der Standard-Edit-Distanz der beiden Strings eine Edit-Matrix und gib die Distanz an. Benutze außerdem eine Backtracing-Matrix, um alle optimalen Edit-Sequenzen bestimmen zu können. Schreibe eine optimale Edit-Sequenz explizit auf.