

Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SoSe 2008

Prof. Dr. Jens Stoye · Dipl.-Inform. Peter Husemann

<http://gi.cebitec.uni-bielefeld.de/teaching/2008summer/sequenzanalyse>

Blatt 8 vom 13.06.2008

Abgabe am 27.06. in der Vorlesung

Aufgabe 1 Fitch-Algorithmus

Der Fitch-Algorithmus berechnet für eine gegebene Baumtopologie mit beschrifteten Blättern eine Beschriftung der inneren Knoten, die minimale Kosten verursacht (*small parsimony problem*). Die in Kapitel 14 auf Seite 146 des Skripts beschriebene Variante des Fitch-Algorithmus verwendet zur Berechnung eine Minimierung der (Einheits-)Kosten. Der originale Fitch-Algorithmus (siehe nächste Seite) nutzt bei der Berechnung hingegen Mengenoperationen. Zeige, dass die beiden Varianten äquivalent sind.

Aufgabe 2 Sankoff-Algorithmus

Beschreibe die Grundidee des Sankoff-Algorithmus für das Baumalignment-Problem.

Aufgabe 3 Steinerbaum

Berechne den minimalen Steinerbaum für die folgenden Mengen von Sequenzen. Benutze, sofern möglich, sowohl die Hamming-, als auch die Einheitskostendistanz.

1. Menge: $s_1 = \text{TGCA}$, $s_2 = \text{GGAA}$, $s_3 = \text{TACA}$, $s_4 = \text{GGCT}$.
2. Menge: $s_1 = \text{TGCA}$, $s_2 = \text{TGAC}$, $s_3 = \text{TGCC}$, $s_4 = \text{GTGAA}$.

The Original Fitch Algorithm

The original description of the algorithm by Fitch [1] is as follows:

For each node v of the tree we construct its *state set* S_v in the following recursive way.

1. For each leaf v , set $S_v = \{x\}$ if x is the character at that leaf.
2. The state sets of the internal nodes are computed in a bottom-up pass of the tree. Assume an internal node v with children u and w .
 - a. If the intersection of the state sets S_u and S_w is non-empty, then S_v becomes this intersection;
 - b. otherwise S_v becomes the union of its children's state sets.

More formally:

$$S_v = \begin{cases} S_u \cap S_w & \text{if } S_u \cap S_w \neq \emptyset \\ S_u \cup S_w & \text{otherwise.} \end{cases}$$

The most parsimonious reconstruction of states at the internal nodes is then obtained in a top-down pass according to the following rules:

1. If the state set of the root contains only one element, assign it to the root, otherwise arbitrarily assign one of the elements in the state set to the root.
2. Let u be a child of node v , and let x_v denote the state assigned to v .
 - a. If x_v is contained in S_u , assign it to node u as well.
 - b. Otherwise, arbitrarily assign any state from S_u to node u .

The length of the most parsimonious tree can then easily be read from the state assignments in the tree.

Literatur

- [1] Walter M. Fitch. Toward defining the course of evolution: Minimum change for a specific tree topology. *Systematic Zoology*, 20(4):406–416, December 1971.