

Advanced Algorithms in Bioinformatics (P4)

Sequence and Structure Analysis

Freie Universität Berlin, Institut für Informatik
Knut Reinert, Sandro Andreotti
Sommersemester 2012

7. Exercise sheet, 27. June 2012

Discussion: 4. July 2012

Exercise 1.

Nussinov SCFG

- Formulate the inside and outside algorithm for the Nussinov SCFG.
- Show how to use your inside and outside variables to calculate the probability that positions i and j are base-paired, summed over all structures.

Exercise 2.

Covariance models

```
Ara.ven. GGUCGC..GUCAACAGUGUUUGAU.CGAAC
Onc.ner. UUACCUGCUUCAACAGUGCUUGAACGGCAA
Gal.gal. UUUAUC..AGUGACAGCGUUCACU.AUAAA
Del.leu. UCGUUC..GUCCUCAGUGCAGGGC.AACAG
SS_cons <<<<<...<<<<<.....>>>>>.>>>>>
```

- Build a covariance model for the alignment of RNAs. Visualize it as a tree, like presented in the lecture.
- Analyze the runtime and memory consumption of the CYK for CM database search.

Exercise 3.

CNF

Convert the production rule $W \rightarrow aWbWWc$ (a, b, c terminal symbols) into Chomsky normal form. If the probability of the original production is p , show the probabilities for the productions in normal form.

Exercise 4.

Minimal Resolved Match Refinement

Prove the following Lemma:

Lemma 1. *There exists a unique resolved refinement \bar{S} of S of minimal cardinality.*

Proof: Sketch: Consider two different resolved refinements \mathcal{S}_1 and \mathcal{S}_2 of \mathcal{S} , both of minimal cardinality. Divide proof into two cases. 1) $(supp_A(\mathcal{S}_1) \neq supp_A(\mathcal{S}_2))$ 2) $(supp_A(\mathcal{S}_1) = supp_A(\mathcal{S}_2), supp_B(\mathcal{S}_1) = supp_B(\mathcal{S}_2))$