

Übungen zur Vorlesung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2014/15

Martin Vingron · Juliane Perner · Annkatrin Bressin

Blatt 4 · Ausgabe am 03.11.2014

Abgabe am 10.11.2014 vor Beginn der Vorlesung

Name:

Matrikelnummer:

Übungsgruppe:

Aufgabe 1 (10 Punkte; Praxis). Auf der Vorlesungsseite finden Sie zwei Guide-trees^{1 2} und die dazugehörigen DNA-Sequenzen³. Installieren Sie ClustalW⁴ und rekonstruieren Sie multiple Alignments dieser Sequenzen unter Verwendung von jeweils einem der Guide-trees als Input. Sind die Ergebnisse identisch?

Aufgabe 2 (20 Punkte; Theorie). Wir haben aus einem multiplen Alignment einen phylogenetischen Baum rekonstruiert. In der Vorlesung haben Sie gelernt, dass man nun Bootstrapping anwenden kann um zu testen wie 'sicher' man sich der Kanten in diesem Baum sein kann. Dazu merkt sich in jeder Bootstrapping-Iteration ob eine bestimmte Bi-partition der Knoten durch eine Kante im entstanden Baum möglich ist.

1. Konstruieren Sie aus den folgenden Bi-partitionen einen Baum. In welchem Fall ist das ohne Widerspruch möglich? Woher kommt der Widerspruch?

(a)	A B C D E F	(b)	A D B C E F
	B E F A C D		B E F A C D
	A C D B E F		A C D B E F
	E B A F C D		E F A B C D

2. Beweisen Sie die folgende Aussage: Bi-partitionen, die während des Bootstrappings in mindestens 50% der Iterationen vorgekommen sind, ergeben zusammen wieder einen (nicht unbedingt binären) Baum.

Aufgabe 3 (10 Punkte; Theorie). Eine Person wurde auf HIV getestet. Mit einer Wahrscheinlichkeit von 99.5% liefert der Test korrekterweise bei einer infizierten Person ein positives Ergebnis (Sensitivität). Die Wahrscheinlichkeit, mit der der Test ebenfalls korrekterweise kein positives Ergebnis bei einer nicht infizierten Person liefert (Spezifität), beträgt 99.3%. Berechnen Sie wie groß die Wahrscheinlichkeit ist, dass eine Person tatsächlich infiziert ist, wenn der Test ein positives Ergebnis ausgibt und diese Person einer niedrigen/hohen Risikogruppe (mit einer HIV Prävalenz von 0.01%/ 0.18%) angehört. Nutzen Sie dazu das Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

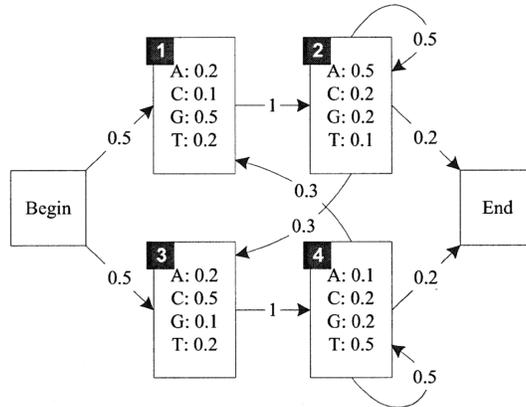
¹Material 1: <http://www.molgen.mpg.de/Algorithmische-Bioinformatik-WS1415/u4/guidetree1>

²Material 2: <http://www.molgen.mpg.de/Algorithmische-Bioinformatik-WS1415/u4/guidetree2>

³Material 3: <http://www.molgen.mpg.de/Algorithmische-Bioinformatik-WS1415/u4/randomDNA>

⁴<http://www.clustal.org/clustal2/>

Aufgabe 4 (35 Punkte; Theorie). Das folgende HMM generiert DNA Sequenzen:



1. Geben Sie zu diesem HMM das Alphabet Σ , die Zustandsmenge S und die Matrix der Übergangswahrscheinlichkeiten A an.
2. Bestimmen Sie per Hand den Viterbi-Pfad für die Beobachtung *TATA* mittels Viterbi-Algorithmus.
3. Berechnen Sie per Hand die Forward- und Backward-Matrix für *TATA*. Welche Wahrscheinlichkeiten lassen sich damit berechnen?

Aufgabe 5 (25 Punkte; Praxis). In der Vorlesung wurde beispielhaft das HMM des *occasionally dishonest casino's* besprochen. Im Kasino werden zwei Würfel eingesetzt: Ein fairer Würfel (F), bei dem alle Ergebnisse gleichverteilt sind, und ein "loaded" Würfel (L), bei dem alle Ergebnisse außer der 6, die mit der Wahrscheinlichkeit von 0.5 fällt, gleichverteilt sind. Die Übergangswahrscheinlichkeiten sind $FF = 0.95$ und $LL = 0.9$.

Sie sollen nun folgende Experimente mit den Beobachtungen, die in der Textdatei⁵ auf der Vorlesungsseite abgelegt wurden, ausführen. Diese Datei enthält die Augenzahlen von 300 Würfeln und den dabei verwendeten Würfel. Nutzen Sie für ihre Experimente den ausführbaren Python-Code für den Viterbi-Algorithmus von Wikipedia⁶.

1. Geben Sie die Stationäre Verteilung für die Zustände (L und F) an und vergleichen Sie diese mit der beobachteten Häufigkeiten der Zustände. Warum sind diese ungefähr identisch?
2. Bestimmen Sie für die beobachtete Sequenz des Würfel-experiments den Viterbi-Pfad.
3. Drehen Sie die beobachtete Sequenz um, und bestimmen Sie dann erneut den Viterbi-Pfad. Vergleichen Sie nun den Viterbi-Pfad der ursprünglichen Sequenz mit dem umgedrehten resultierenden Viterbi-Pfad. Begründen Sie ihre Beobachtung.

⁵Material 4: <http://www.molgen.mpg.de/Algorithmische-Bioinformatik-WS1415/u4/casino>

⁶http://en.wikipedia.org/wiki/Viterbi_algorithm