

Algorithmische Bioinformatik Übungsblatt 4

Ausgabe: 05. November 2019 · Besprechung: 12. November

Aufgabe 4.1 Textmodelle und HMMs sind äquivalent.

Beweise die eine Richtung dieser Aussage, indem Du zu einem gegebenen HMM (Q, q_0, A, E, e) ein Textmodell (C, c_0, Σ, ϕ) angibst, so dass die gleichen Beobachtungs-Sequenzen mit gleichen Wahrscheinlichkeiten generiert werden.

Beweise die zweite Richtung dieser Aussage, indem Du zu einem gegebenen Textmodell (C, c_0, Σ, ϕ) ein HMM (Q, q_0, A, E, e) angibst, so dass die gleichen Beobachtungs-Sequenzen mit gleichen Wahrscheinlichkeiten generiert werden.

Szenario Wir betrachten ein (kriminelles!) Casino, das ein Würfelspiel veranstaltet. Ein Gewinn wird nur beim Wurf einer Sechs gezahlt. Der Spieler glaubt, dass ein fairer Würfel verwendet wird.

Tatsächlich gibt es aber drei verschiedene Würfel: einen fairen (F), einen unfairen (U), wobei jede Augenzahl zwischen 1 und 5 mit Wahrscheinlichkeit $1/5$ auftritt, sowie einen Würfel zum Anlocken naiver Kunden (A), bei dem die Sechs mit Wahrscheinlichkeit $2/7$ und die restlichen Augenzahlen mit Wahrscheinlichkeit $1/7$ auftreten.

Begonnen wird mit dem fairen Würfel ($q_0 = F$). (Vor dem ersten Wurf findet aber bereits ein Übergang statt!) Die Übergänge sind wie folgt. Zu 80% wird der aktuelle Würfel beibehalten. Von F wird mit je 10% Wahrscheinlichkeit zu U bzw. A gewechselt. Von U oder A wird mit 20% Wahrscheinlichkeit zurück zu F gewechselt.

Aufgabe 4.2 Erstelle 100 Simulationsreihen (in verschiedenen Dateien) zu je 1000 Würfeln. Speichere dabei sowohl die Beobachtungen als auch die Zustände.

Aufgabe 4.3 Berechne die Logarithmen der Wahrscheinlichkeiten der 100 Beobachtungsreihen (Forward-Algorithmus). Stelle die 100 Ergebnisse in einem Histogramm dar.

Aufgabe 4.4 Berechne für jede der 100 Beobachtungsreihen den Viterbi-Pfad. Welchen Anteil hat der Viterbi-Pfad jeweils an der Gesamtwahrscheinlichkeit der Beobachtung?

Aufgabe 4.5 Vergleiche Viterbi-Pfad und wahren Zustandspfad. Wie oft sagt der Viterbi-Pfad den korrekten Zustand vorher?

Die vorstehenden Aufgaben erfordern, dass man die HMM-Algorithmen programmiert. Wer das nicht möchte, kann sich im Kopf Folgendes überlegen.

Aufgabe 4.6 Berechne im Casino jeweils die (Gesamt-)Wahrscheinlichkeit, Viterbi-Wahrscheinlichkeit und Viterbi-Pfad für die Beobachtungsreihen 6, 66, 666, 6666, 66666, ... (bis es langweilig wird).