Technische Universität
Kaiserslautern

Fachbereich Informatik
AG Algorithmen und Komplexität

Issue Date: 12.11.2012
Version: 2012-11-12 10:06

Exercise Sheet 3
Computational Biology (Part 2), WS 12/13

Hand In: *Until Monday, 26.11.2012, 10:00 am, email to* `wild@cs...` *or in lecture.*

Exercise 8

4+6 Points

We consider the dice game of the *occasionally dishonest Casino* from the lecture and its associated Hidden Markov model with states $\{q_0, F, U\}$ and alphabet $\{1, 2, 3, 4, 5, 6\}$.

Since the casino recently changed its management, we are not sure if the dice used are still the same. However, we know from a reliable source that the new management was very interested in possible ways to maximize its luck. In this task we will try to find out as much as possible about their business model.

You may use an arbitrary programming language. Make sure to avoid *underflows* that render your results unusable. You may use existing libraries for HMMs.

For both parts, hand in thoroughly commented code and the final output. For part b), also present intermediate results of matrices δ and η that show the convergence of the entries.

a) By a "lucky coincidence" we know for one series of rolls which die was used for each of them. The data can be found in the file `dishonest-casino-a.txt`[1]. The first line contains the series of rolled numbers, the second line denotes the die used.

   Use this data to determine parameters δ and η of a HMM. Is the casino still using the unfair dice from the lecture?

   Afterwards, use the Viterbi algorithm to compute the probability $\Pr[\pi^*|S]$ of a most likely path π^*, where S is the first line of the file `dishonest-casino-a.txt`. Report π^* itself, as well. Is your π^* the same as the state sequence in the training data?

b) For a second table we do not have any information. We thus spend several nights writing down the dice rolls until security gets aware and removes us from the casino. The resulting data can be found in `dishonest-casino-b.txt`[2], each line representing one evening.

   Implement the Baum-Welch algorithm and train a HMM of the structure given above on the sequences given in the file until convergence. Use several random starting values for δ and η, the HMM from the lecture and your trained from a). Does the training converge to the same HMM for all starting values?

   Based on your experience from above, do both tables use the same set of dice?

[1] `http://wwwagak.cs.uni-kl.de/Veroffentlichungen/AdB-II/AdB-II-WS-12/13/dishonest-casino-a.html`
[2] `http://wwwagak.cs.uni-kl.de/Veroffentlichungen/AdB-II/AdB-II-WS-12/13/dishonest-casino-b.html`