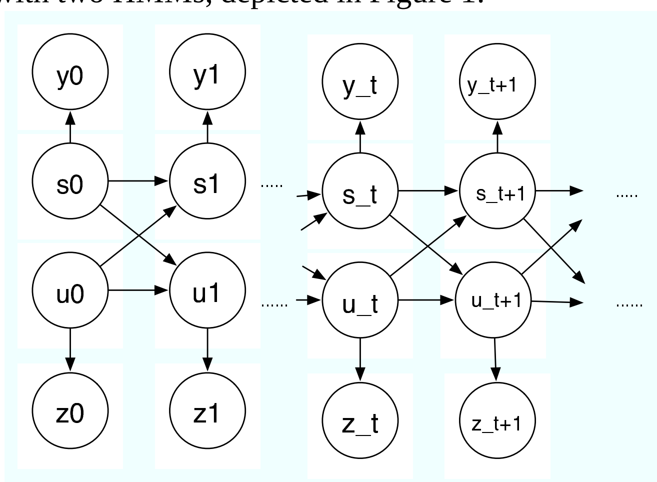
Probabilistic Reasoning in AI - Assignment 5

Due Friday April 4, 2008

1. [45 points] **Coupled Hidden Markov Models**

   We discussed in class several models for reasoning with sequences of data (trajectories). The HMM is the simplest such example, in which states are hidden, and we see observations that depend on the state. The Coupled Hidden Markov model (CHMM) is a similar kind of graphical model: we have several hidden Markov models running in parallel, and their states interact. This model is quite useful, for example, when you try to parse video, and you consider the observations as being sound and visual data, respectively.

   Consider a system with two HMMs, depicted in Figure 1:



   Here, s_i and u_i are the states of the two coupled HMMs, y_i and z_i are the observations coming from the two chains, and the two chains interact in the way depicted in the picture.

   (a) [5 points] Specify what are the parameters of this model.

   (b) [10 points] Derive an algorithm for computing the joint probability of a sequence of observations $(y_0, z_0), (y_1, z_1) \dots (y_T, z_T)$.

   (c) [10 points] Derive a forward algorithm that computes the most likely sequence of hidden states given a sequence of observations. You recall that in order to do this, in the case of a simple HMM, you maintain a "belief state", which gives the probability of each hidden state based on the observations seen so far. You can use a similar idea here. Alternatively, you may consider how you can apply the junction tree algorithm to this situation.

   (d) [10 points] Suppose that instead of the chains being coupled at every time step, the coupling only happens every k time steps (on time step 0, k $2k$ etc). For $k = 1$, you get the same model as above. If k is fairly large compared to the length of sequences, the chains are called **loosely coupled**. Describe how your model and the inference algorithms change in this case.

(e) [10 points] Suppose that you observe several sequences of two time series and you know that they come from a loosely coupled HMM; you know the number of possible states for each individual chain, but you do not know k. Describe a learning algorithm for this problem.

2. [30 points] **Gibbs sampling for partially observed Markov chains**

Consider a simple Markov chain where each state is 0 or 1. The initial value s_0 is drawn uniformly. The transition matrix is such that $p(s_{t+1} = s_t) = 0.9$ and $p(s_{t+1} \neq s_t) = 0.1$. Suppose that we observe $s_4 = 1$ and we want to compute $p(s_0|s_4 = 1)$.

(a) [10 points] Show how you would use Gibbs sampling in order to compute this conditional probability.

(b) [10 points] Extend your algorithm for the case in which we observe $s_t = 1$, and no other data is observed.

(c) [10 points] Describe what happens with the Gibbs sampling approach as t increases. If t was very large and you had to do approximate inference for such a problem, would you use Gibbs sampling or likelihood weighting? Justify your answer.

3. [25 points] **Expected Utility**

Data the android has been chosen to represent Starfleet in a power contest against the best Romulan android. The prize is $100 (old monetary units used on Earth in the 21st century). Data estimates that he has a 0.9 chance of winning if he is functioning correctly.

The night before the contest, Data observes a malfunction in his controls. Based on a quick diagnosis, Lt. Geordi LaForge tells him that there are three possible causes for the problem. With probability 0.2, Data has a damaged neurotramsmitter. In this case, his chance of winning drops to 50%. Also, in this case, with probability 40%, he will sustain further damage, valued at $70. If he does not enter the contest, the repair is valued at $20 and will fix the problem for sure. With probability 0.3, Data observed just a minor problem, which will not appear again. In this case, Data's chances of winning are not altered. With probability 0.5, there is a software malfunction, in which case Data will lose for sure. Geordi can later fix the problem at no charge.

(a) [5 points] Draw the decision graph for this problem.

(b) [5 points] Compute the expected utility of every choice of action, and explain what is the optimal decision.

(c) [5 points] Suppose that there is an entrance fee for the contest. How much should Data be willing to pay to enter?

(d) [5 points] The Ferengi have a software analyzer that can detect whether the software is ok or not. How much should Data be willing to pay for using the analyzer?

(e) [5 points] Draw the decision graph corresponding to the latter situation. Explain Data's optimal choices of action based on this graph.