Reinforcement Learning: Tutorial 5 (week from 3. 3. 2014)

1. How can particle filters be used in the context of robot localization?

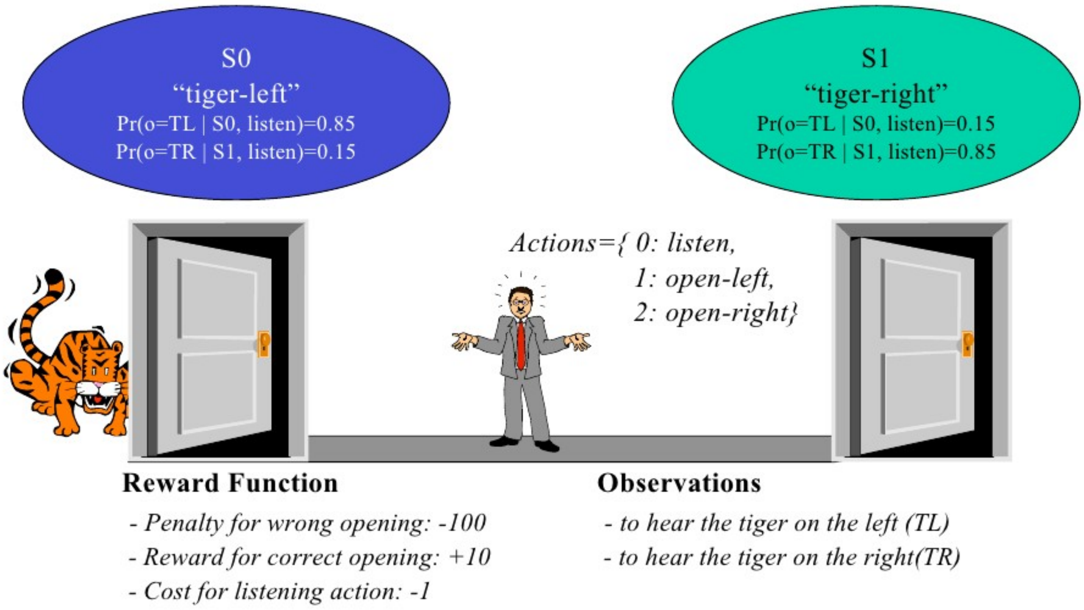
2. The "art" of importance sampling: We are sampling $P(x)$, which may be not cover the interesting aspect of the game. It is already interesting to consider sampling $X/L(x)$ based on the distribution $P(x)L(x)$. Why? How could it be useful in RL? Similarly, we can consider this scheme (see on the right ->) Why should one want to do this? What happens for $P(x)=W(x)$?

```
Algorithm Discrete_Bayes_filter( Bel(x),d )
η=0
If d is a perceptual data item z then
    For all x do
        Bel'(x) = P(z | x) Bel(x)
        η = η + Bel'(x)
    For all x do
        Bel'(x) = η^{-1} Bel'(x)
Else if d is an action data item u then
    For all x do
        Bel'(x) = Σ_{x'} P(x | u, x') Bel(x')
Return Bel'(x)
```

$$\overline{\mathcal{A}} = \frac{\sum_x P(x)\,\mathcal{A}(x)\,/\,W(x)}{\sum_x P(x)\,/\,W(x)} \,.$$

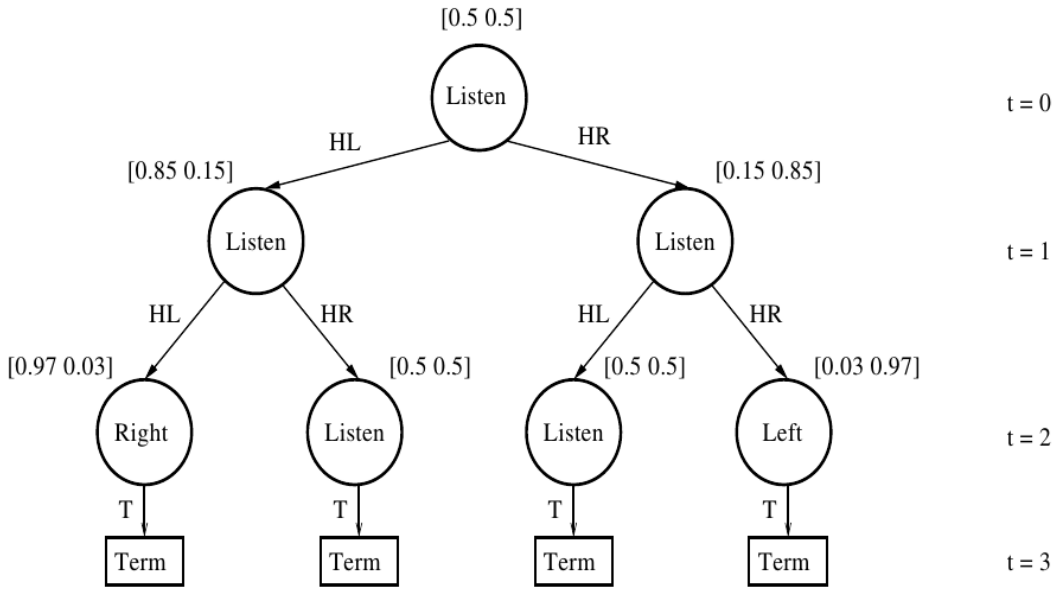
3. How are POMDPs and Hidden Markov Models (HMMs) related? Would a Viterbi algorithm be useful in POMDPs?

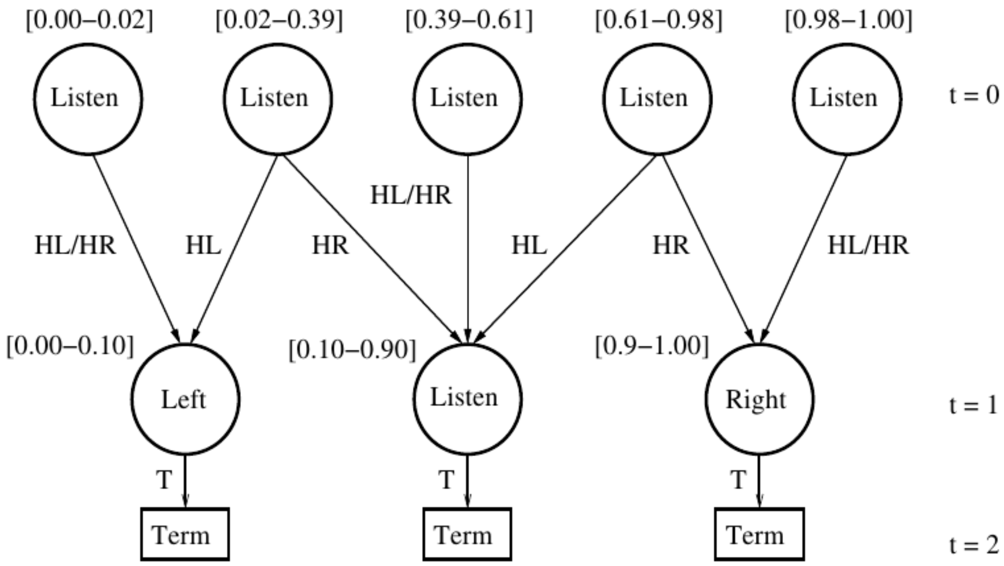
4. Discuss the tiger problem (from: Dr. Stephan Timmer "Introduction to POMDPs")




```
# This is the tiger problem of AAAI paper fame in the new POMDP
# format. This format is still experimental and subject to change

discount: 0.75
values: reward
```

```
states: tiger-left tiger-right
actions: listen, open-left, open-right
observations: tiger-left, tiger-right

Transitions:
listen -> identity
open-left -> uniform
open-right -> uniform

Observations
listen (in either state):     0.85 0.15
                              0.15 0.85
open-left: uniform
open-right: uniform

Rewards:
R:listen : * : * : * -1
R:open-left : tiger-left : * : * -100
R:open-left : tiger-right : * : * 10
R:open-right : tiger-left : * : * 10
R:open-right : tiger-right : * : * -100
```





5. Consider the application to a POMDP to the problem of controlling several elevators problem. For what definition of states does any uncertainty arise? Discuss the advantage of a POMDP over a state abstraction (that does not distinguish between states that can be confused). Compare to the original Barto&Crites approach (see final slides of the lecture RL09). Can the design of the elevator operation be changed such that this uncertainty is removed/reduced?

6. Recall the discussion of "afterstates" from a previous tutorial. Afterstates are an option to include the reaction of an opponent into the own policy. Under what conditions would it make sense to reformulate the problem as POMDPs.

7. Consider a robot moving down a hallway as a 1D problem with states being sections of the track of a length of 1m. The robot's speed is 1m/s +/- 0.1m/s (assume a uniform distribution of deviations). Discuss the belief propagation in standard POMDP vs. the corresponding effects in an augmented MDP or in QMDP. Think of a navigation task which is then to be solved by either of these methods.

8. Assume a robot moving in a dark environment where information is available only from touch sensors. The robot learns to move successfully using a POMDP. Now the lights are switched on and the robot can use again its excellent visual system. How can it use the information from POMDP for initialising a simpler RL method

9. Have a look at a review paper such as Anthony R. Cassandra (1998) A Survey of POMDP Applications. Discuss set-up, advantages and limitations of POMDP in the mentioned application problems (or do this simply for any of the examples above).