Bayesian Reinforcement Learning

Note Title 11/8/2005

RL: $\langle S, A, R \rangle$
we are missing T (transition dynamics)

MDP: $\langle S, A, T, R \rangle \Rightarrow \pi^*$
$\Rightarrow V^*$

Bayesian Q-Learning:
distributions over V^*

Model based Bayesian Exploration
distributions over T

Model-based Bayesian RL:

- distributions over $P_R(s'|s,a)$
- let $\theta^a_{s,s'} = P_R(s'|s,a)$

Prior $P(\theta^a_{s,s'}) = P_R(P_R(s'|s,a) = v)$

Posterior $P(\theta^a_{s,s'} \mid s \xrightarrow{a} s')$

Dirichlet distributions are conjugate priors of multinomial distributions

$Dir(\alpha_1, \ldots, \alpha_n) = k\,\theta_1^{\alpha_1 - 1}\,\theta_2^{\alpha_2 - 1} \cdots \theta_n^{\alpha_n - 1}$

Let $P(\theta^a_{s,s'}) = Dir(\alpha^a_{s,s'})$

then $P(\theta^a_{s,s'} \mid s \xrightarrow{a} s') = k P(s \xrightarrow{a} s' \mid \theta^a_{s,s'}) P(\theta^a_{s,s'})$

$= k P(s \xrightarrow{a} s' \mid \theta^a_{s,s'}) \prod_{s''} (\theta^a_{s,s''})^{\alpha^a_{s,s''}-1}$

$= k\, \theta^a_{s,s'} \prod_{s''} (\theta^a_{s,s''})^{\alpha^a_{s,s''}-1}$

$= k \prod_{s'' \neq s'} (\theta^a_{s,s''})^{\alpha^a_{s,s''}-1} (\theta^a_{s,s'})^{\alpha^a_{s,s'}}$

$= Dir(\alpha^a_{s,s'} + \delta(s', s''))$

State space: $S \times \{\theta^a_{s,s'}\}$

S: fully obs; $\{\theta^a_{s,s'}\}$: unobs

Hence we have a POMDP