Policy Iteration for Relational MDPs

Chenggang Wang and Roni Khardon, Computer Science Department, Tufts University, {cwan | roni}@cs.tufts.edu

First Order Decision Diagrams

Domain definition

4 predicates: \( p_1(x), p_2(x), q_1(x), q_2(x) \).
3 actions: \( A_1: \text{no-op}; \ A_2(x^*): q_2(x^*) \rightarrow p_2(x^*); \)
\( A_3(x): A_3(x) \) with probability 0.9 \( q_1(x^*) \rightarrow p_1(x^*); \)
\( A_4(x): \) with probability 0.1 (no-op).

Domain dynamics captured by Truth Value Diagrams (TVD)

An example of Rel-regress-policy

Observation 1: Existential relational policies are not fully specified. As a result, their value functions may not be well defined.

Observation 2: There exist domains and well defined existential relational policies such that

1. \( \text{regress-policy} \) cannot be expressed within an existential language.
2. \( w_{i+1} = Q_{w_{i+1}} > Q_w^{e} \), i.e., Rel-regress-policy may underestimate \( Q_w^{e} \).

Theorem: The sequence of value functions obtained from relational modified policy iteration converges monotonically to the optimal value function.