RePReL: Integrating Relational Planning and Reinforcement Learning for Effective Abstraction
Harsha Kokel, Arjun Manoharan, Sriraam Natarajan, Balaraman Ravindran, Prasad Tadepalli

Objective
Learn to act in relational domains, with varying task, number of objects and relations.

Motivation
Humans are quite natural at creating an abstract representation for efficient planning. They use different abstractions for different tasks. Can we leverage their domain knowledge and provide such abstractions to the RL agent?

RePReL
- Hierarchical Framework with Planner and RL
- Plan sequence of operators at high level and learn to execute each operator at lower level
- Advantages
  - Planner provides compositionality
  - Different RL agent for execution allows separate state representations
  - Abstraction is most useful when we have multiple relations and objects, so consider Relational MDP.
- Use D-FOCI statements to identify relevant objects and relations for a given task.

Toy Example
- Taxi domain with multiple passengers
- Acquire the sequence of passenger pickup and drop operators from planner
- Learn to execute the operators using RL
- Use different goal-conditioned RL for each operator
- Pickup operator only needs pickup location, drop location is irrelevant
- Similarly, drop only needs destination, pickup location becomes irrelevant

Model-agnostic Abstraction
Definition 4 (Li, Walsh, and Littman (2006)). A model-agnostic abstraction $\phi(s)$ is such that for any action $a$ and abstract state $s$, $\phi(s_1) = \phi(s_2)$ if and only if

$$\sum_{s_1'} \phi(s_1', a, s_1') = \sum_{s_2'} \phi(s_2', a, s_2')$$
$$\sum_{s_1''} R(s_1, a, s_1') = \sum_{s_2''} R(s_2, a, s_2')$$

This is the bisimulation condition, where the abstraction function preserves the transition and the reward distribution of ground MDP in the abstract MDP.

D-FOCI Statements
- First Order Conditional Influence (FOCI) statement is of the form “if condition then $X1$ influence $X2$”
- FOCI statements encode the information that literal $X2$ is influenced only by the literals in $X1$ when the stated condition is satisfied
- We present Dynamic FOCI (D-FOCI) statements

Sequences of Planning and RL
- Toy Example
  - Taxi domain
  - Office domain

Human knowledge can help provide effective abstractions that enable efficient learning and effective generalization across tasks and objects.

More details on https://starling.utdallas.edu/papers/RePReL