PEORL: Integrating Symbolic Planning and Hierarchical Reinforcement Learning for Robust Decision Making

Fangkai Yang¹, Daoming Lyu², Bo Liu², Steven Gustafson¹

¹ Maana Inc., Bellevue, WA, USA  ² Auburn University, AL, USA

Introduction
Symbolic planning and reinforcement learning have both been used to create agents that behave intelligently in real world.

Planning-agent
• requires prior knowledge
• does not rely on trial-and-error
• is brittle to domain change and uncertainty

RL-agent
• doesn’t require prior knowledge
• Relies on trial-and-error from huge amount of experience
• is highly adaptive and robust to domain uncertainty

Can symbolic planning and reinforcement learning mutually benefit each other for decision making?
• Symbolic planning uses domain knowledge to guide RL for meaningful exploration
• RL helps symbolic planning to generate adaptive and robust plan to handle domain uncertainty and change

Background: Symbolic Planning
• Symbolic planning concerns on using a formal, logic-based language to describe prior knowledge of the dynamic domain, and automate reasoning and planning in the domain.
• Action languages based on Answer Set Programming, such as AC (Lee, Yang and Lifschitz, 2013), can be used to automated planning utilizing answer set solver such as Clingo.

Background: Reinforcement Learning
• Reinforcement learning is defined on a Markov Process $\langle S, A, P_{st}, r, \gamma \rangle$.
• The agent has no knowledge about the transition matrix and probability, and by interacting with the environment, it learns a policy $\pi: S \rightarrow A$ to accumulate maximal reward.
• R-learning (Schwartz, 1993; Mahadevan, 1996), different from Q-learning, concerns on long term average reward and is particularly suitable for planning and scheduling tasks.
• R-learning iterates on two values: long term average reward $R$, and gain reward $p$.

PEORL: Integrating Symbolic Planning with RL
• PEORL framework stands for Planning-Execution-Observation-Reinforcement Learning that features bi-directional communication between planning and learning.
• In PEORL, causal laws in action language has effect on cumulative plan quality.
• PEORL planning goal contains two parts: a logical constraint stating the goal state condition, and a linear constraint to enforce generating "better quality plan".
• Symbolic actions are mapped to “options”, in the sense of hierarchical RL to learn.

PEORL Architecture
We evaluate the framework in two scenarios: Taxi domain and Gridworld

• In Taxi Scenario 1, every movement (north, south, west, east) receives -1 reward, successful drop off: +10, improper pickup or drop off: -10. In this scenario, planning agent behaves best because planner favors shorter plan, yet PEORL agent converges to optimal after explored longer alternatives.
• In Taxi Scenario 2, we introduce an extra coupon at (4,4), leads to +10 reward. PEORL agent can adapt to the change by generating a longer, yet more rewarding plan.
• In Gridworld, the robot needs to navigate around a T-shaped bumping area and arrive at the door, then activate the door, open the door and enter.
• PEORL agent converges to the most rewarding plan and learns the policy of operating the door, effectively reduces execution failure in comparison to planning agent that cannot learn from experience.
• In all scenarios, PEORL agent converges to optimal behavior a lot faster than RL agent, thanks to equipped with domain knowledge.

Conclusion
We show that by integrating symbolic planning with hierarchical RL (hierarchical R-learning in particular), planning and RL can mutually benefit each other to make robust decisions. It is the first work of that features bi-directional communication between planning and RL.
• Future work involves integrating symbolic planning with deep RL, investigation on transferability, and integration with automatic option discovery.