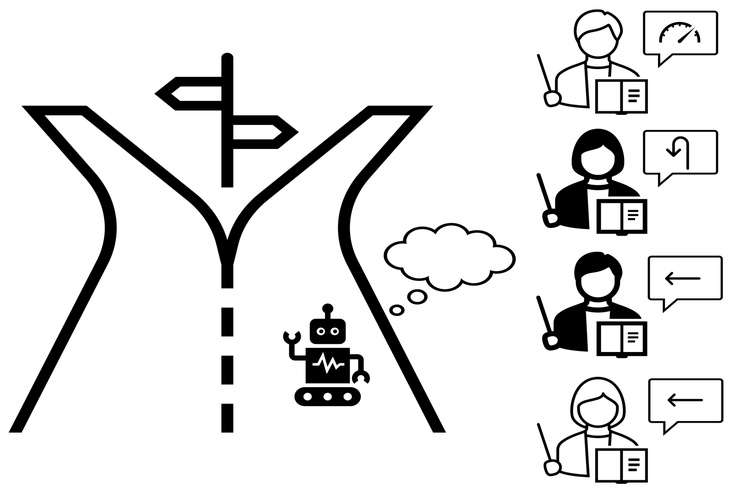
# Two-Level Actor-Critic Using Multiple Teachers

Su Zhang, Srijita Das, Sriram Ganapathi Subramanian, Matthew E. Taylor

Washington State University, Pullman, WA, USA
Vector Institute, Toronto, Canada
University of Alberta, Edmonton, Canada

## Problem Setup



**Objective:** Train the learning agent using action advice from the set of multiple teacher. And answer the question:

**When should the student listen to which teacher?**

## Research Questions

**R1:** Can the two-level actor-critic method handle **a mixture of sub-optimal teachers** relative to existing methods?

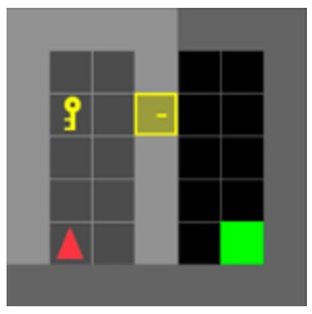
**R2:** Can the two-level actor-critic method incorporate **multiple partial teachers** with different areas of expertise?

## Experiments

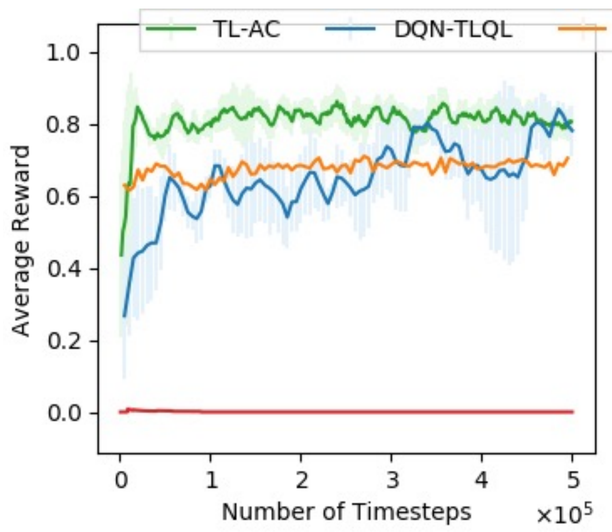
**Task:** MiniGrid DoorKey 7x7

**Teacher Sets:** Optimal teacher, Random teacher, Teacher L, Teacher R

**Baselines:** A2C, DQN-TLQL, AC-Teach

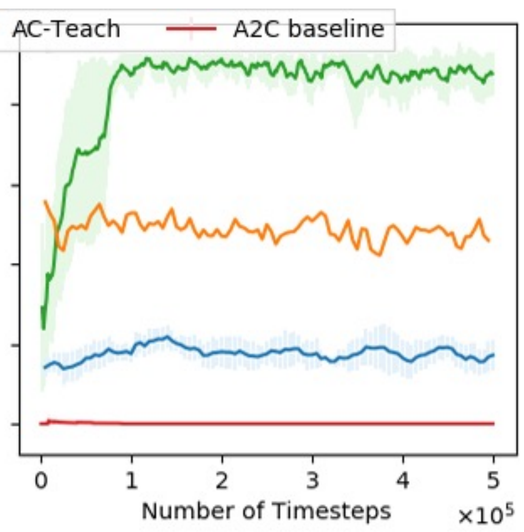


**FULL TEACHER**



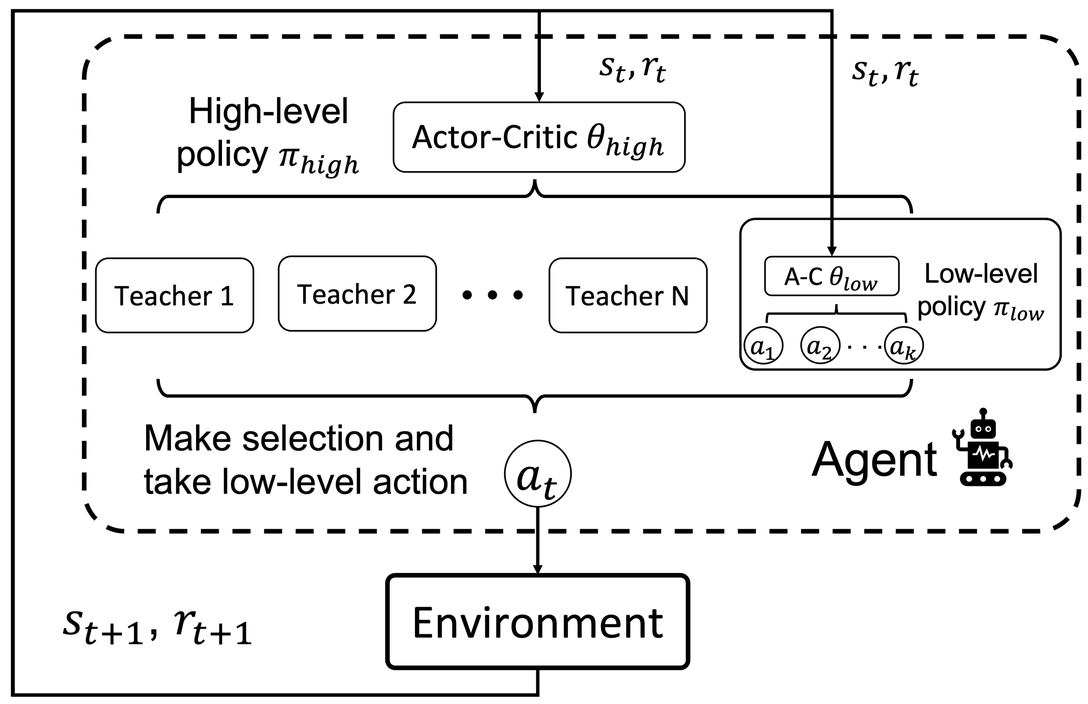

(a) Results with 1 optimal teacher and 2 random teachers

**PARTIAL TEACHER**




(b) Results with Teacher L and Teacher R

## Two-Level Actor-Critic Structure




I. **Low-level policy network: select action**
II. **High-level policy network: select teacher and take its action advice**

- Reward: environment reward when executing the chosen teacher's advice
- 2 networks use the same critic: $V_{\pi_{high}}(s) = V_{\pi_{low}}(s)$

## Advantages of our TL-AC:

- Lightweight and simple: use a **two-level** network structure, with a **single critic** for both levels
- Can easily switch between policies at every time-step to incorporate the best teacher's advice
- Full (complete) or partial teachers
- Sub-optimal teachers
- Can incorporate with any actor-critic algorithm

## Next Steps:

- Human teachers
- Uncertainty and confidence related scheme
- Limited presence of the teachers
- Budgeted version

WASHINGTON STATE UNIVERSITY

The Intelligent Robot Learning Laboratory

UNIVERSITY OF ALBERTA