Solving Phase Ordering with Off-Policy Deep Reinforcement Learning Algorithms

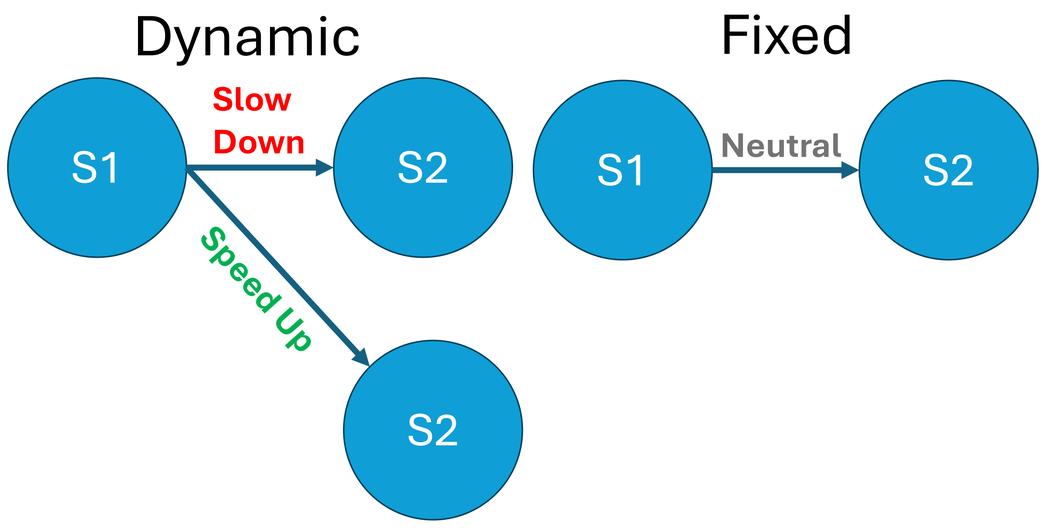
UC SANTA CRUZ BaskinEngineering

Oliver Chang (elochang@ucsc.edu)

Phase Ordering Problem

Motivation

- Modern compilers are equipped with many phase (optimizations) for intermediate representations
- Phase sequences could quickly grow exponentially so compiler must ensure there are no conflicts and are optimally sequenced
- Compiler-Gym is an accessible framework that provides an LLVM compiler, environments (programs), reward, and transition functions

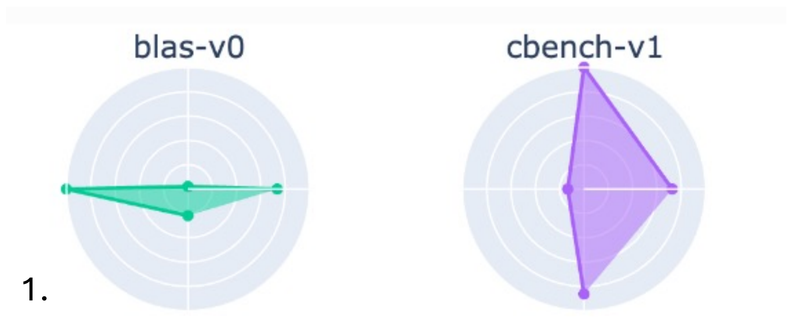


Challenges

- Optimizations could be interdependent
- Large search space
- Sequential tasks disproportionately represent data (non-IID)

Goal

- Apply DQN and Double-DQN to the phase ordering problem on BLAS and Cbench dataset environments by leveraging Compiler-Gym
- Demonstrate differences between using observation spaces (namely Autophase and InstrCount)
- Generalize Double-DQN to a larger action and observation space
- Reward Function: $R(s_t) = \frac{C(s_{t-1}) - C(s_t)}{C(s_t = 0) - C(s_b)}$



1.

Reinforcement Learning Algorithms

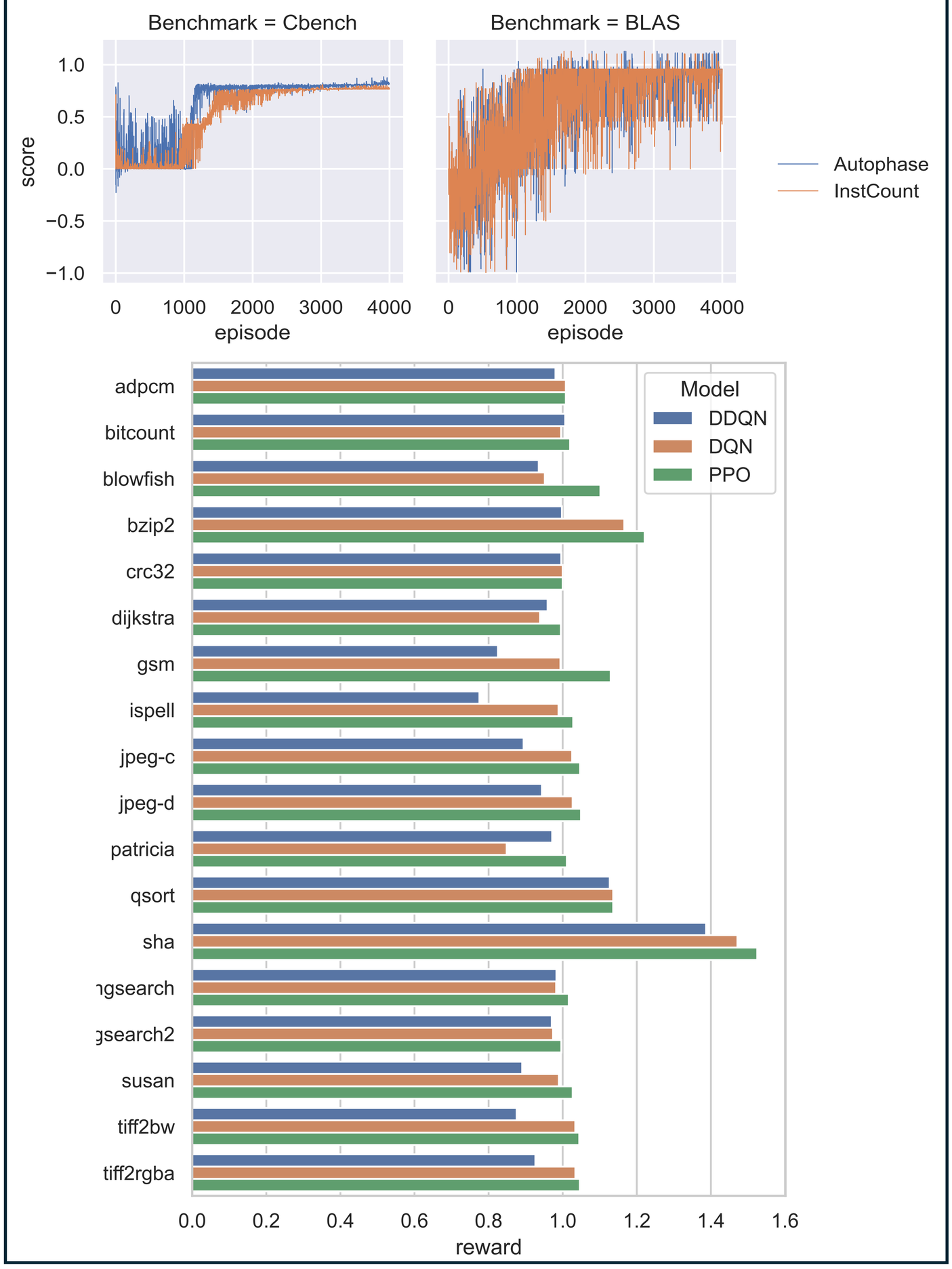
- DQN: combines Q-learning with deep neural networks.
- Double-DQN: addresses overestimation bias from DQN, resulting in a more stable performance – especially in high-action spaces
- PPO: policy-gradient, on-policy method that uses a clipped-surrogate objective function to limit policy updates

1. https://compilergym.com/llvm/index.html#action-space

Setup

- 124-dimension action space
- Observations are Autophase (56-dimension) feature vector for static IR states or LLVM-IR (70-dimension) feature vector
- 3-layer MLP; each layer contains 512 neurons

Results



Take-aways

- In certain programs, Double-DQN performs on-par with PPO and DQN with a large action and observation space (the plot above has Double DQN trained on 124-dim action space vs 14-dim)
- An alternative approach to phase-ordering problem

Future Work

- Apply Soft-Actor Critic (SAC)
- Experiment with difference reward functions and state representations
- Try hierarchical or multi-task RL to increase phase sequence lengths