Predicting Future Actions of Reinforcement Learning Agents

Stephen Chung[1] Scott Niekum[2] David Krueger[3]

1. Problem Formulation

How can we predict future actions and events for a trained agent?

- **Motivation:** Increasing application of RL in real world raises need to predict future agent actions and events
  - Intervention for Dangerous Behavior: Predicting an autonomous vehicle about to run a red light enables timely intervention
  - Improve Human-Agent Interaction: Helpful for passengers and other drivers to know if a nearby autonomous vehicle will turn left or right

- **Action prediction:** Predict the action distribution in the next L steps – $P(A_{t+1}, A_{t+2}, \dots, A_{t+L} | S_t, A_t)$
- **Event prediction:** Predict the probability of an event $E_{g,k} \coloneqq \{S_k, A_k | g(S_k, A_k) = 1\}$ defined by a binary function g occurring in the next L steps – $P\left(\cup_{l=1}^{L} \; E_{g,t+l} \middle| S_t, A_t\right)$

- Assuming a fixed (and trained) policy; policy from different types of RL algorithms are considered:
  1. **Non-planning agent** – Agents without an explicit world model and do not plan; e.g. PPO, IMPALA (Espeholt et al., 2018), Q-Learning, and most model-free RL algorithm
  2. **Implicit planning agent** – Agents without an explicit world model but exhibits planning-like behavior; e.g. DRC (Guez et al., 2019)
  3. **Explicit planning agent** – Agents with an explicit world model and plans with it; e.g. MuZero (Schrittwieser et al., 2020), Thinker (Chung, Anokhin, & Krueger, 2024)
- Assume we have a fixed number of transitions generated from the policy as training data

2. Methods - Inner-state Approach & Simulation-based Approach

Vanilla approach – Train a predictor with the state-action pair (S_t, A_t) as inputs and future action or event as target using supervised learning; but certain additional information can be provided as inputs:

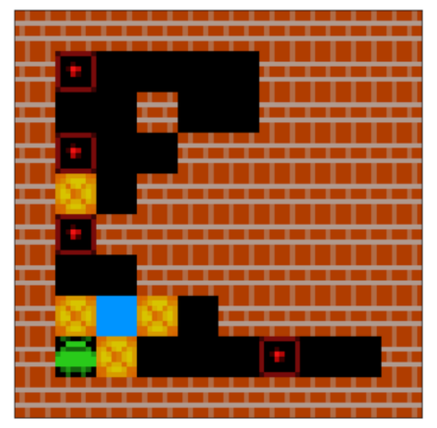
Inner-state Approach:

- Inner-state: All intermediate computations required to compute the agent's action
- In addition to the state-action pair, we select some inner states as inputs to the predictor, e.g.:
  - IMPALA – hidden layer activations
  - DRC – hidden state in the LSTM
  - MuZero – most visited rollout in simulations
  - Thinker – all rollouts in simulations
- Akin to probing the neuron activation of an animal's brain to predict its future action; if the animal is planning, better prediction accuracy can be expected

Simulation-based Approach:

- Train a world model and simulate the agent in this model to generate rollouts, which are provided as additional input to the predictor
- When the world model is perfectly accurate, the empirical distribution of rollouts are the same as the target distribution
- Akin to placing an animal in a virtual world to predict its future actions; if the virtual world closely resembles the real world, better prediction accuracy can be expected

3. Experiment



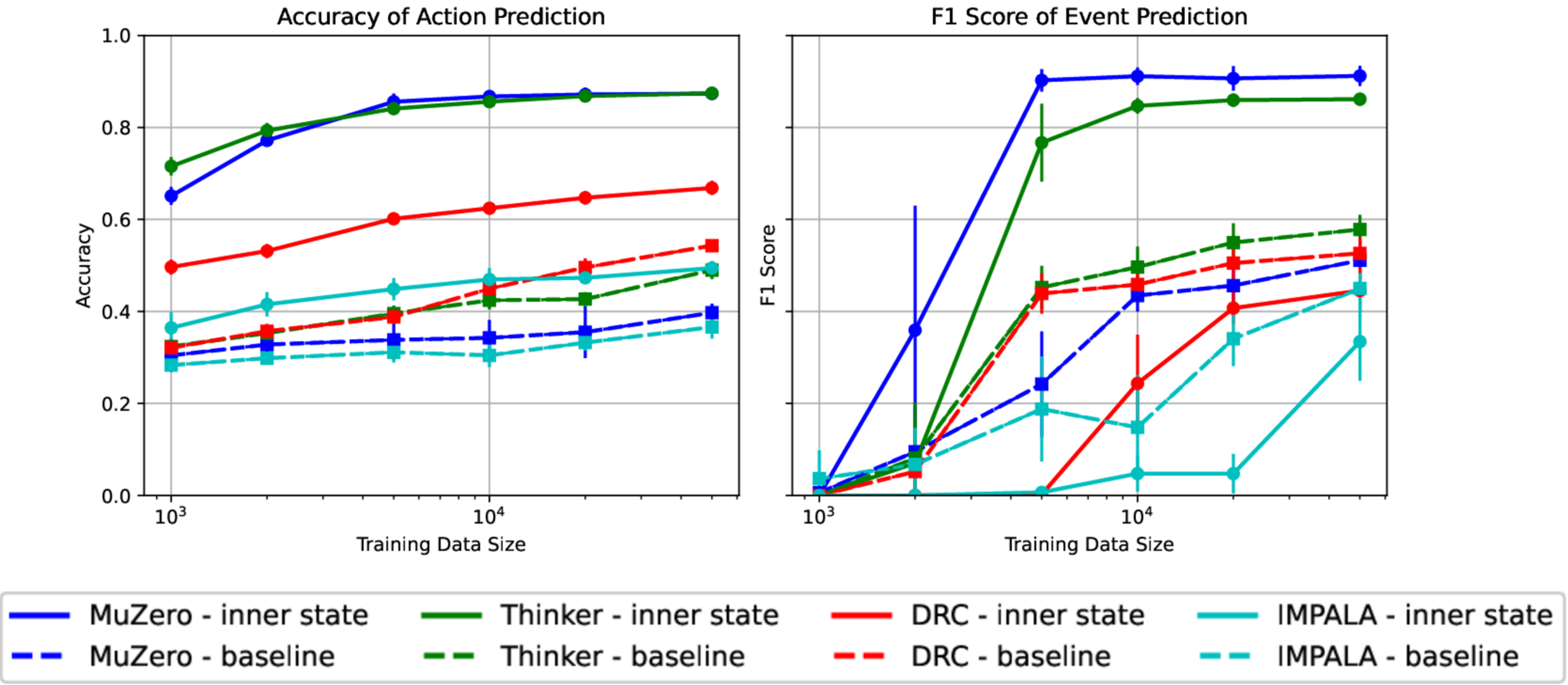
Environment – Sokoban

Goal: push the boxes ▨ to targets ■ ; can only push but not pull boxes

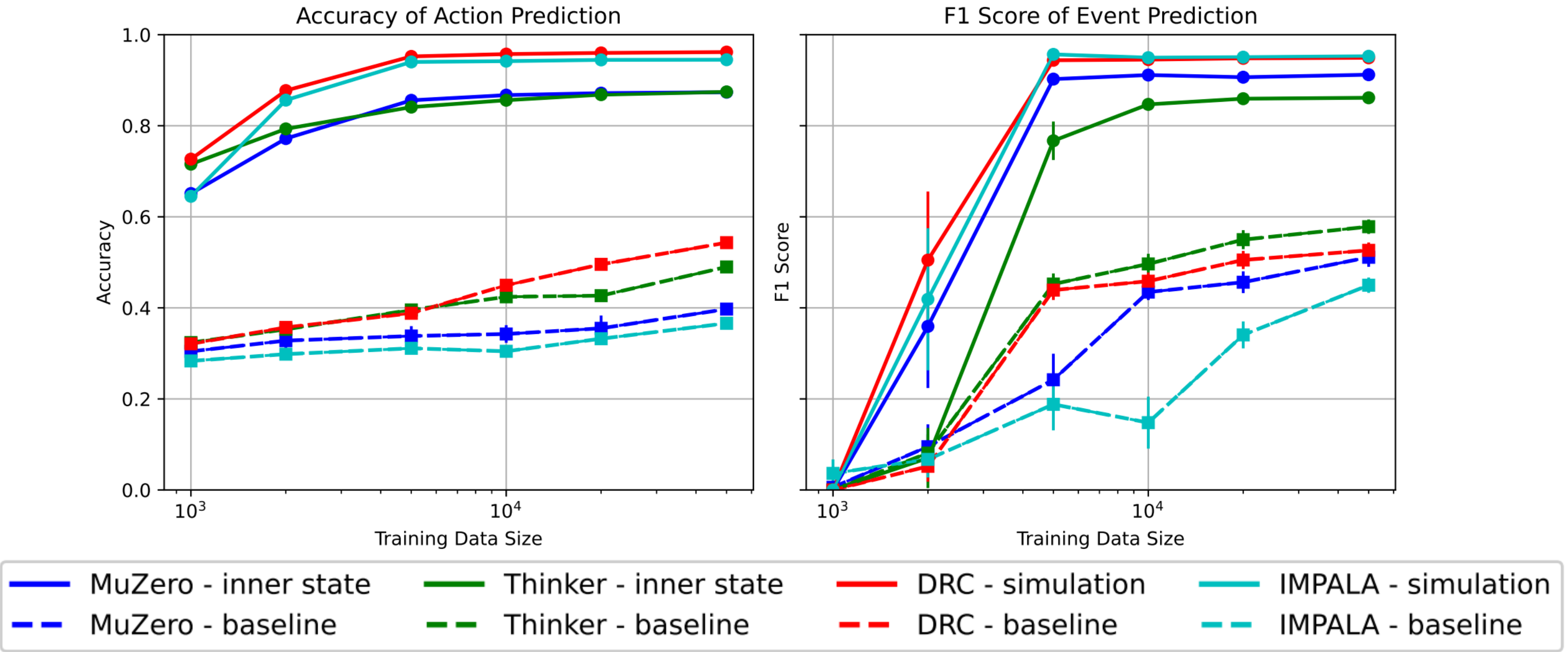
Action prediction - Predict the next 5 action

Event prediction – Predict if the agent will stand on the blue tile within 5 steps

Inner-state vs Vanilla Approach (baseline)



Simulation-based Approach vs Vanilla Approach (baseline)*



*We did not try the simulation-based approach in explicit planning agents, as it effectively requires two levels of simulation (one for the predictor and one for the agent), which is too difficult to learn.

- Both inner-state and simulation-based approaches are generally useful for predicting future actions and events
- The inner-state approach performs best with explicit planning agents, followed by implicit planning agents, and then non-planning agents
- The simulation-based approach works very well when an accurate world model is available, but is much less robust to the quality of world model in another ablation study

Conclusion – Use simulation-based approach when an accurate world model is available; use inner-state approach otherwise. Explicit planning agents are more predictable within the inner-state approach.