Frame skipping in Deep Reinforcement Learning

Karan Agarwalla Rishi Agarwal

IIT Bombay

October 22, 2021

Outline

Introduction

- Markov Decision Processes
  - MDPs are used to model sequential decision tasks.
  - Formally, an MDP 'M' is a 5-tuple (S, A, T, R, γ).
- Learning agents
  - Agents interact with the MDP, take actions and transition from one state to another.
  - The agent receives a "reward" at each step which objectively evaluates its performance.
  - Traditionally, an agent senses states and decides actions at every time step.
- Frame-skipping and action repetition
  - **Reduced sensing**: agent senses states periodically at every "d" time steps. For the "d-1" time steps, agent is free to perform any set of actions. (a.k.a frame-skipping)
  - **Action repetition**: agent repeats the last executed action till it senses the next state.

Motivation

- Reduced sensing consumes **lesser computational** resources.
- Lack of theoretical insights in frame-skipping and action repetition.
- Experimentally determine the impact of frame-skipping on **Atari-2600 games**.
- Validate our hypothesis that there exists a relation between discount factor 'γ' and frame-skip parameter 'd', such that higher frameskip resembles lower discount factor.

Related Work

- Deep Q-Networks[3] introduced in the domain of Atari Games matched human experts using a single architecture
- Recent developments include use of duelling networks[5] and macros[1], pre-defined action sequences
- A policy gradient approach,FiGAR[4], tunes the frameskip online
- Upper bounds on error due to action repetition[2]

Problem Description

- Empirically find out the effects of:
  - Frameskip parameter, and
  - Discount factor on the agent's performance
- Determine if there exists any relation between the frameskip parameter and the discount factor.
- Generate plots for average reward achieved by the agent vs the number of training steps for different values of 'd' and 'γ'.
- We use the DQN algorithm and run experiments on two Atari games - Seaquest and Enduro.
- We model the two games as continuous MDPs, and the objective is to maximize the long term reward obtained by the agent.

Our Approach

- Based on the same low level convolutional structure of DQN[3]
- Experience replay based approach to prevent correlating frames
- Used duelling architecture to quickly identify valuable states
- Rewards clipped from -1 to 1 to prevent exploding gradients

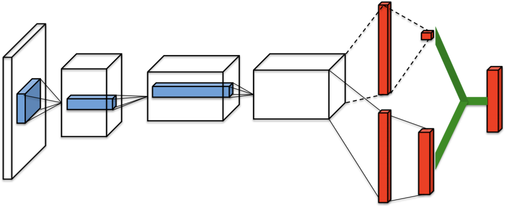


Figure 1: Duelling DQN Network[5]

Experiments

- We have trained agents on two games - Seaquest and Enduro.
- Training has been performed till 50 million time steps
- We used different values of frameskip - 4, 8, 16, 20 and discount factor - 0.9 and 0.99.
- To compare different values, we plot the moving average of rewards obtained by the agents with the number of time steps processed by the environment.
- We also present a table which shows the maximum moving average reward for a given frameskip and discount factor.

Results

FS	0.99	0.9
4	**716**	452
8	490.85	**493.75**
16	303.9	222.8
20	155.5	135.45

(a) Enduro

FS	0.99	0.9
4	**5179.5**	570
8	3452	**4697**
16	1819	2912.5
20	1323	1879.5

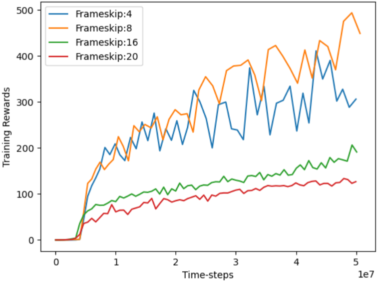
(b) Seaquest

Table 1: Training Rewards

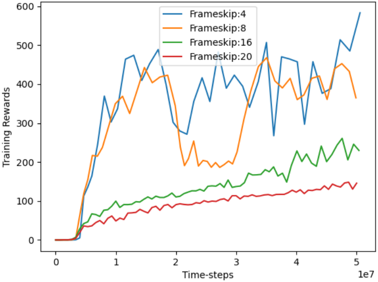
Game	Human	Linear	DQN	Our Approach
Enduro	309.6	129.1	301.8	716
Seaquest	20182	664.8	5286	5179.5

Table 2: Base-Line Comparison[3]

Results



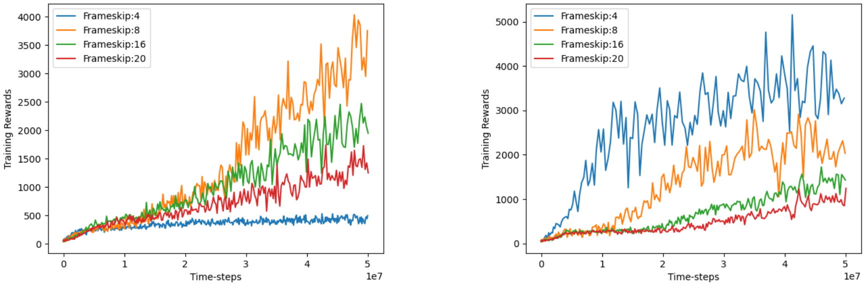
(a) Enduro with $\gamma = 0.9$



(b) Enduro with $\gamma = 0.99$

Figure 2: Plot of training rewards vs time-steps.

Results



(a) Seaquest with $\gamma = 0.9$ (b) Seaquest with $\gamma = 0.99$

Figure 3: Plot of training rewards vs time-steps.

Some Observations

- Best score for Enduro is 716 using frameskip 4 and $\gamma = 0.99$
- Best score for Seaquest is 5179.5 using frameskip 4 and $\gamma = 0.99$
- Best frameskip parameter for $\gamma = 0.99$ is 4 and for $\gamma = 0.9$ is 8
- Informal experiments indicate that the performance without frame-skipping is lower than that of higher frameskip values

Conclusion and Discussion

- We validate the significant benefits of frame-skipping, by running extensive experiments on Atari-2600 games.
- We ran experiments with different values of discount factor to gain insights about our hypothesis. Results so far do not support our hypothesis.
- However, we believe that more experimentation is required to arrive at a reasonable conclusion.
- A possible direction for future research is to explore effects of frameskip parameter and discount factor on a wide variety of algorithms like policy gradient based methods, online RL and other sequential tasks.

References I

Ishan P. Durugkar, Clemens Rosenbaum, Stefan Dernbach, and Sridhar Mahadevan.
Deep reinforcement learning with macro-actions.
CoRR, abs/1606.04615, 2016.

Shivaram Kalyanakrishnan, Siddharth Aravindan, Vishwajeet Bagdawat, Varun Bhatt, Harshith Goka, Archit Gupta, Kalpesh Krishna, and Vihari Piratla.
An analysis of frame-skipping in reinforcement learning.
CoRR, abs/2102.03718, 2021.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller.
Playing atari with deep reinforcement learning.
CoRR, abs/1312.5602, 2013.

References II

Sahil Sharma, Aravind S. Lakshminarayanan, and Balaraman Ravindran.
Learning to repeat: Fine grained action repetition for deep reinforcement learning.
CoRR, abs/1702.06054, 2017.

Ziyu Wang, Nando de Freitas, and Marc Lanctot.
Dueling network architectures for deep reinforcement learning.
CoRR, abs/1511.06581, 2015.

Thank You!