Honors in Machine Learning and Artificial Intelligence

Mini Project
Evaluating the performance of Deep Reinforcement Learning for solving the Cartpole Problem

Aboli Marathe

Roll No. 41301 | PRN. 71900008E | BE 3

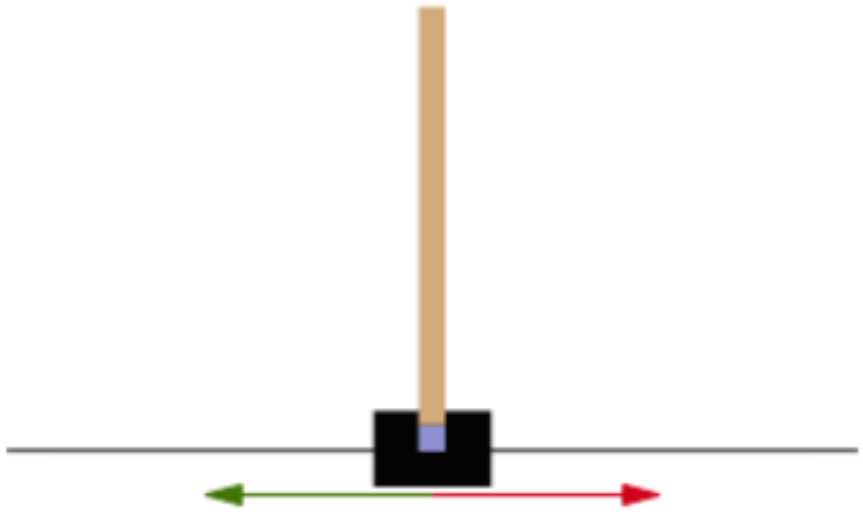
Department Of Computer Engineering
Pune Institute Of Computer Technology

Introduction

Reinforcement learning is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward.

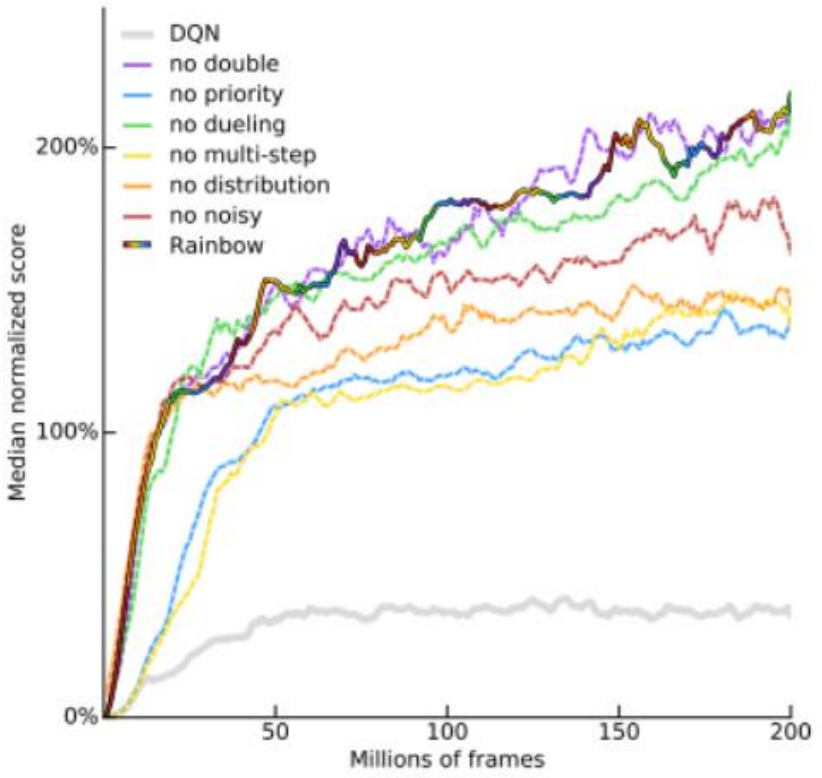
Cartpole problem

Cartpole - known also as an Inverted Pendulum is a pendulum with a center of gravity above its pivot point. It's unstable but can be controlled by moving the pivot point under the center of mass. The goal is to keep the cartpole balanced by applying appropriate forces to a pivot point.



Cartpole schematic drawing

Rainbow



Performance of Rainbow on Atari Benchmark

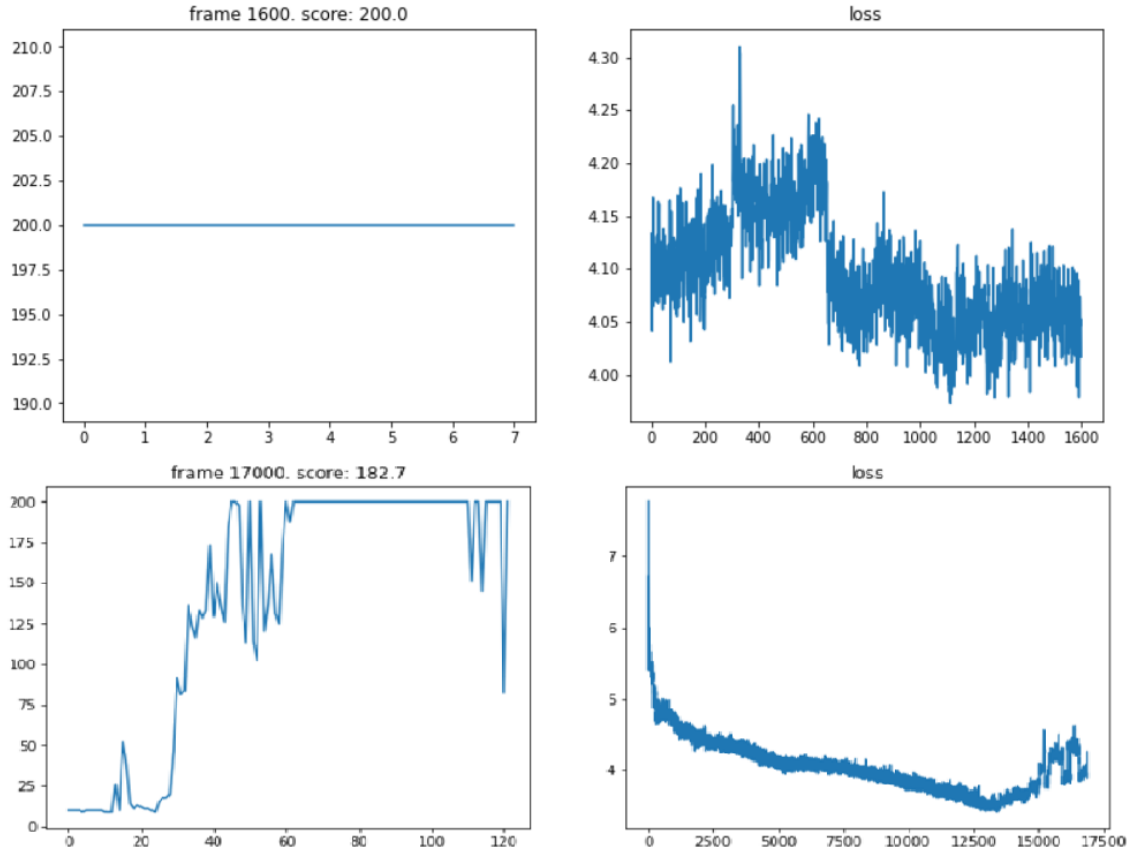
A selection of the below six extensions that each have addressed a limitation and improved overall performance were integrated into a single integrated agent, Rainbow for improved performance on RL tasks [8]:

1. Double Q-learning
2. Prioritized replay
3. Dueling networks
4. Multi-step learning
5. Distributional RL
6. Noisy Nets

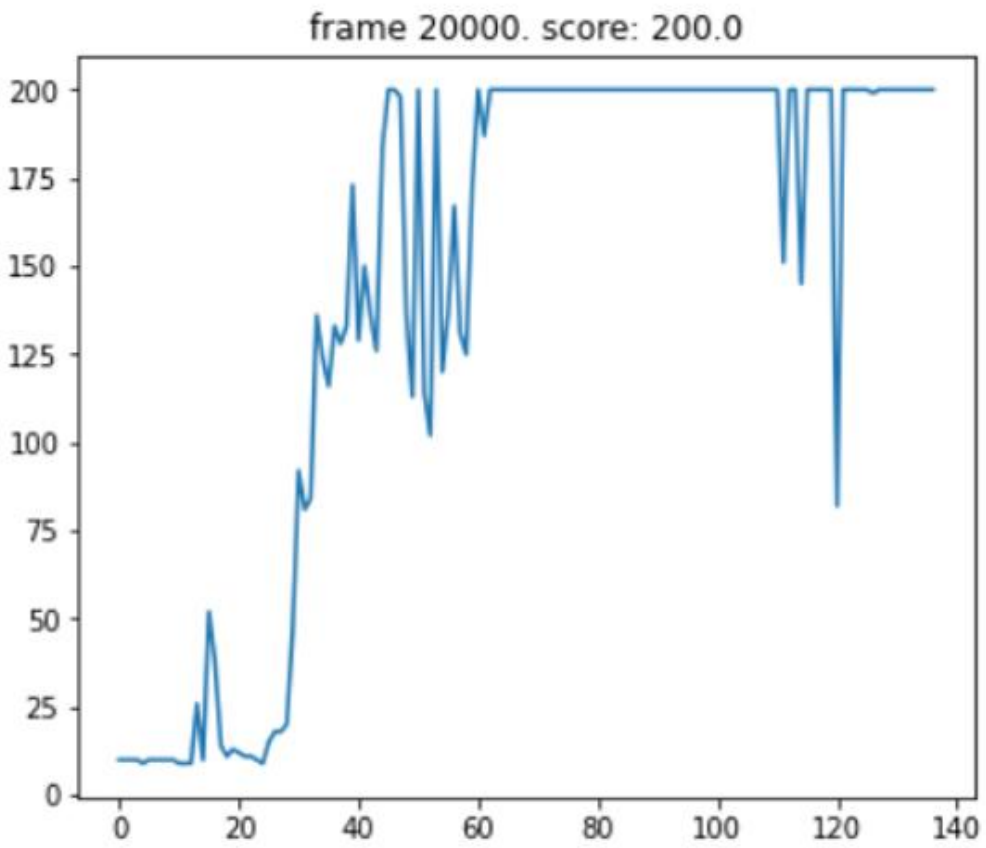
Rainbow Agent

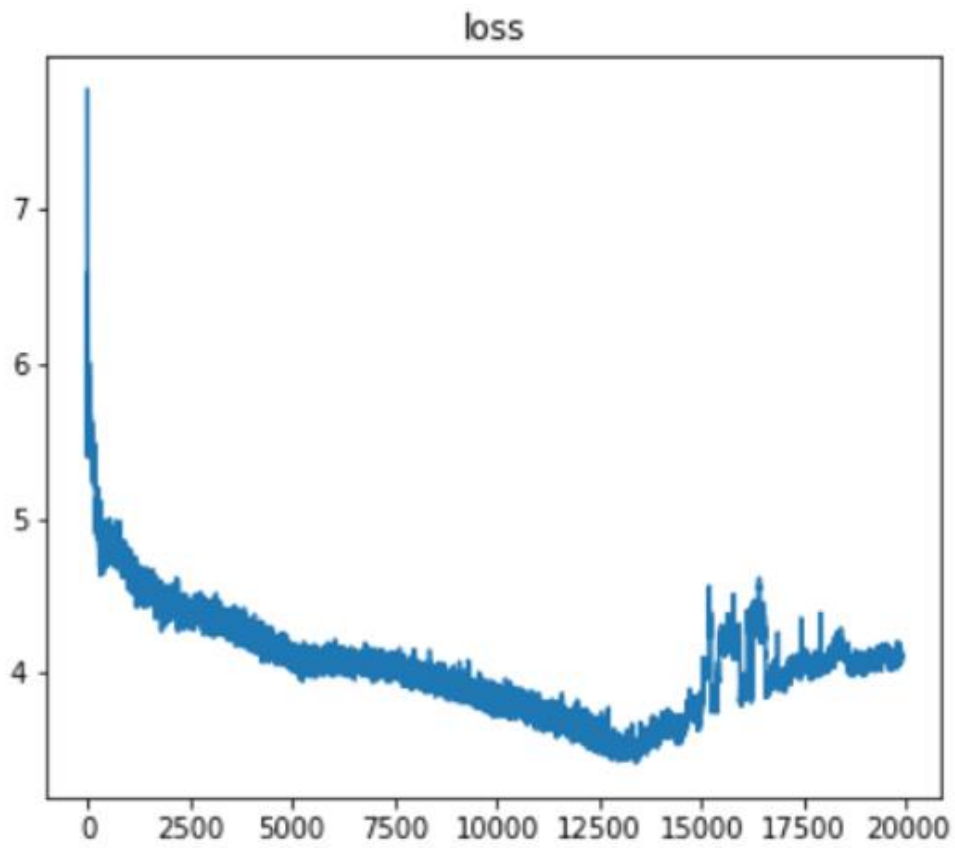
Method	Note
select_action	select an action from the input state.
step	take an action and return the response of the env.
compute_dqn_loss	return dqn loss.
update_model	update the model by gradient descent.
target_hard_update	hard update from the local model to the target model.
train	train the agent during num_frames.
test	test the agent (1 episode).
plot	plot the training progresses.

Training


frame 1600. score: 200.0
210.0
207.5
205.0
202.5
200.0
197.5
195.0
192.5
190.0
0
1
2
3
4
5
6
7
loss
4.30
4.25
4.20
4.15
4.10
4.05
4.00
0
200
400
600
800
1000
1200
1400
1600
frame 17000. score: 182.7
200
175
150
125
100
75
50
25
0
0
20
40
60
80
100
120
loss
7
6
5
4
0
2500
5000
7500
10000
12500
15000
17500

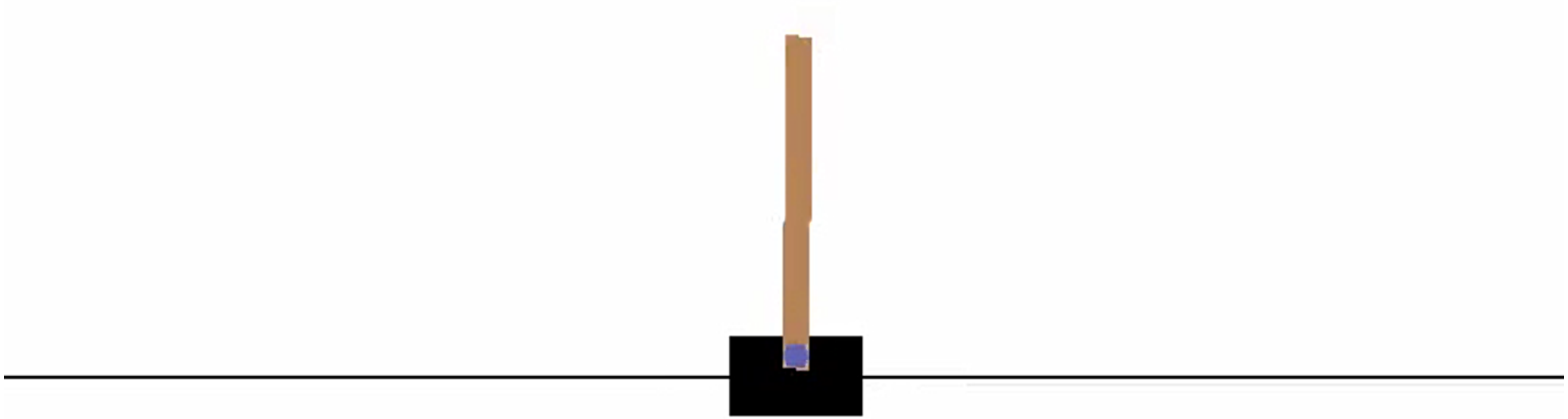
Results


frame 20000. score: 200.0
200
175
150
125
100
75
50
25
0
0
20
40
60
80
100
120
140


loss
7
6
5
4
0
2500
5000
7500
10000
12500
15000
17500
20000

Conclusion

- The model is able to reach a score of 200 with a minimal training time of 5 minutes. We selected a small sample of training to demonstrate in the project and showed the model after training during this period.
- Thus in this project we have applied and demonstrated the efficiency of the Rainbow method of Reinforcement Learning and used it for solving the cartpole problem.
- The final video and performance of model can be seen in next slide and is attached as a part of the submission.



References

1. V. Mnih et al., "Human-level control through deep reinforcement learning." Nature, 518 (7540):529–533, 2015.

2. van Hasselt et al., "Deep Reinforcement Learning with Double Q-learning." arXiv preprint arXiv:1509.06461, 2015.

3. T. Schaul et al., "Prioritized Experience Replay." arXiv preprint arXiv:1511.05952, 2015.

4. Z. Wang et al., "Dueling Network Architectures for Deep Reinforcement Learning." arXiv preprint arXiv:1511.06581, 2015.

5. M. Fortunato et al., "Noisy Networks for Exploration." arXiv preprint arXiv:1706.10295, 2017.

6. M. G. Bellemare et al., "A Distributional Perspective on Reinforcement Learning." arXiv preprint arXiv:1707.06887, 2017.

7. R. S. Sutton, "Learning to predict by the methods of temporal differences." Machine learning, 3(1):9–44, 1988.

8. M. Hessel et al., "Rainbow: Combining Improvements in Deep Reinforcement Learning." arXiv preprint arXiv:1710.02298, 2017.

9. https://github.com/Curt-Park/rainbow-is-all-you-need

10. Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." arXiv preprint arXiv:1312.5602 (2013).