



SkiMaster

CS221-Spring 2019

Erick Cardenas
ecasas@stanford.edu

Troy Ling
troyling@stanford.edu

Emilio Unda
eunda@stanford.edu

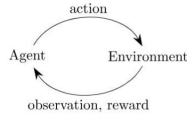
<https://github.com/ercardenas/ski-master>
<https://youtu.be/vo73nmA10y4>

Introduction

Reinforcement Learning (RL) in games has gained popularity in recent years. Inspired by various AI gaming bots, we developed a RL algorithm and a heuristic based agent to master the game of Skiing, a classic video game for the Atari console. In this game, the player needs to ski down a slope by passing between pairs of flags, while making best efforts to avoid obstacles along the way. The evaluation metric for success is to pass as many flags as possible in the shortest amount of time until the final pair of flags.

Environment

- Using the Atari 2600 environment from Gym
 - A toolkit for developing RL algorithms by OpenAI [1]
- Program an agent in an event loop
 - Observe
 - Obtain Reward
 - Perform an Action



Problem Definition

State (Ram):

- The RAM of the Atari 2600 environment

State (Image):

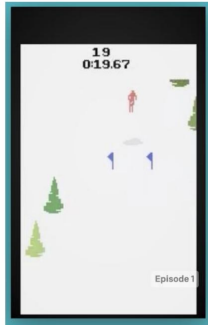
- An image of emulator screen. (250, 160, 3)

Actions:

'noop', 'right', 'left'

Rewards:

- 0; if not game_over
- (-number_of_frames - k * uncaptured_flags); otherwise



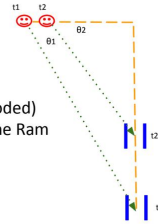
Approaches

Ram Based Q-Learning:

- Model
 - Linear Classifier
- Features:
 - Direction to take (one hot encoded)
 - The value at each position in the Ram

Image Based Q-Learning

- Model
 - Convolutional Neural Network
 - 4-6 Layers Deep
- Inputs
 - Direction to take (one hot encoded)
 - The image (we tried several approaches)
 - Raw image
 - Raw Image + Dense Optical Flow (prev, current)



Heuristic Based Model

- Model
 - Hand crafted rules (angles)
- Features
 - Position of the Skier
 - Position of the next set of flags

Results

| Model | Captured Flags | Time | Average Sum of Rewards |
|------------------------|----------------|-------------|------------------------|
| Random Actor Baseline | 3 | 1:07 | -17,000 |
| Human Player Oracle | 20 | 0:39.5 | -4,500 |
| Ram Based Q-Learning | 0 | 4:59 | -30,000 |
| Image Based Q-Learning | 0 / 2 | 4:59 / 1:44 | -30,000 / -14,000 |
| Handcrafted Heuristics | 20 | 0:37 | -3,800 |

Discussion

Our best model, handcrafted heuristics, finished the game in 37 seconds by passing through all 20 flags. This is better than the human player's (oracle) performance of 39.5 seconds (-6% latency reduction).

On the other hand, neither of the Q-Learning models were able to learn any meaningful strategy. We suspect that the Linear model was not able to derive any meaningful relationships between the features and the action to take. The Image based model did not seem able to understand the value of a given state. It did not understand the spatial relationships between the skier and the flags, even when we added the flow vectors.

Challenges

- Q-Learning CNN model could not learn how to play the game.
 - The weights would sometimes overflow to Nan
- Figuring out the best heuristic of the skiing game based on the position of the player.

References

- Gym Open AI, 2016, <https://gym.openai.com/>
- Stella, A Multiplatform Atari 2600 Emulator, <https://stella-emu.github.io/>; Stephen Anthony, Christian Speckner, Thomas Jentzsch, Eckhard Stolberg
- ALE, The Arcade Learning Environment: An Evaluation Platform for General Agents, Bellemare M.-G., Naddaf Y., Veness J., Bowling M., Journal of Artificial Intelligence Research, 2013, June volume 47, pages 253-279, <https://github.com/mgbellemare/Arcade-Learning-Environment>
- Solve Skiing-v0 by Using Deep Reinforcement Learning, Github: jaywalnut310 <https://github.com/jaywalnut310/rl-atari-skiing>
- Super Mario Bros Game Playing with Deep Reinforcement Learning, Dapeng Hong (hongdp), Billy Wan (xwan), Yaqi Zhang (yaqiz), Stanford CS221 2018
- Gym, Skiing-v0, <https://gym.openai.com/envs/Skiing-v0>