Beating the World's Best at Super Smash Bros. Melee with Deep Reinforcement Learning

Firoiu, V., Whitney, W. F., & Tenenbaum, J. B.

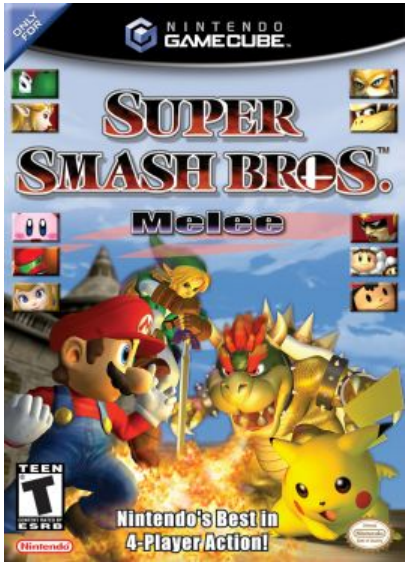
arXiv preprint arXiv:1702.06230, 2017.

Introduction

- Goal

  Investigate the performance of deep learning AI on SSBM

- Super Smash Bros. Melee (SSBM)
  - Released in 2001
  - Popular fighting game
  - There is still active tournament



SSBM as RL environment

- Rules
  - Send the opponent out of screen to win
  - Build up Percentage to send them far

- Challenges
  - Large game state and complex transition
  - Delayed rewards
  - Great deal of diversity (26 characters and multiple stages)

Q-Learning

- Model-free reinforcement learning algorithm
- Learn a mapping from state to future rewards of each action

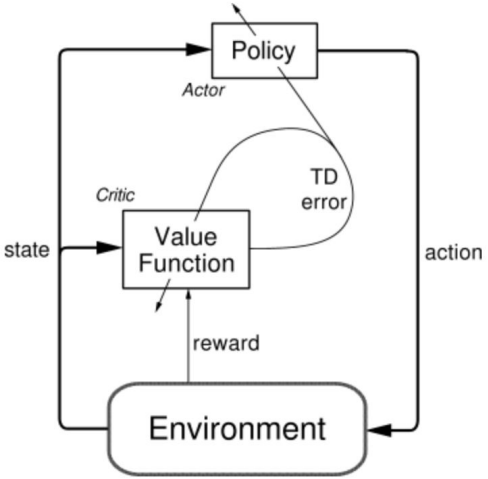
- Details
  - Neural network instead of Q table
  - Gradient descent based on the loss (objective) function below
  - Choose action with ε-greedy and proportion of Q values

$$L = (Q(s_t, a_t) - [r_t + \lambda r_{t+1} + \cdots + \lambda^n Q(s_{t+n}, a_{t+n})])^2$$

Policy Gradient Method

- Actor-Critic

  Actor selects action and Critic criticizes the actions made by Actor

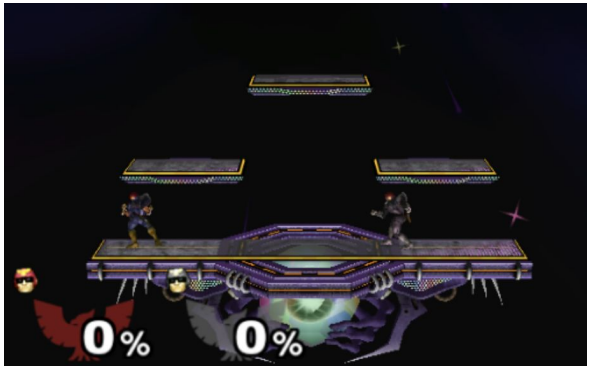


- Details
  - Update policy using the gradient below
  - Entropy scale h enables agent to explore effectively

$$\Delta\theta = \alpha(A(s,a) - h)\nabla_\theta \log \pi_\theta(s,a)$$
$$A(s,a) = Q(s,a) - V(s)$$

Training

- Infinite time mode
- Character: Captain Falcon, stage: Battle Field



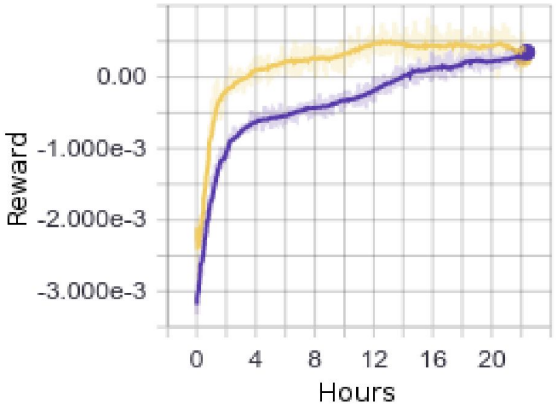
- Rewards
  - win +1, lose -1,
  - (damage dealt - damage taken)*constant
- limitations
  - No information about projectile
  - Less input
- Generate experiences in parallel

Result: against in-game AI

- Both AI defeted the strongest in-game AI

- The Actor-Critic AI learned strategy that's similar to human players
- The Q-Learning AI learned tricky policy

  - "Starter agent" (provided by OpenAI as a baseline RL AI) couldn't learn well in the same setup.



Result: Self-play

The agents trained against the in-game AI was not strong enough to defeat experienced players… >>> Self-play

- Train AI against old version of itself
- Became so strong that the AI can beat top players in competetion

But it had some weaknesses,

- Crouching at the edge of the stage caused the AI to suicide by falling off the edge.
- The AI performed much worse when it played against other characters.

Agent diversity and character transfer

- Train another five characters
  - Different network for different characters
  - Self-play against different characters

> This diversity fixed the odd behaviour.

- Transfer learning
  - Initializing weights of network with other character’s trained network makes learning faster.

Opponent	Rank	Kills	Deaths
S2J	16	4	2
Zhu	31	4	1
Gravy	41	8	5
Crush	49	3	2
Mafia	50	4	3
Slox	51	6	4
Redd	59	12	8
Darkrain	61	12	5
Smuckers	64	8	5
Kage	70	4	1

The AI after self-play training surpasses top SSBM players.

	Scratch	Sheik	Marth	Fox	Falco	Peach	Falcon
Sheik	36	0	4	7	7	3	9
Marth	40	5	0	11	10	7	10
Fox	31	8	6	0	2	6	7
Falco	35	9	6	2	0	7	5
Peach	26	2	4	5	5	0	6
C. Falcon	53	9	11	13	12	10	0

Transfer learning makes learning much faster.

Discussion

- Q-Learning performed reasonably well against fixed opponent (i.e. in-game AI). However, it didn’t perform well when the opponent’s policy changes during training (i.e. self-play and against other characters).

- AI’s reaction speed (33ms) can be cheating. But their attempt to add action delay wasn’t successful. This remains interesting and challenging problem.

Conclusions

In this paper, they

- Introduce a new environment, SSBM, to RL community.
- Analyze the difficulties when adapting traditional RL algorithm to such environment.
- Demonstrate an agent based on deep learning which even surpasses top human players.

Links

“A Super Smash Bros-playing AI has taught itself how to stomp professional players”

https://qz.com/917221/a-super-smash-bros-playing-ai-has-taught-itself-how-to-stomp-professional-players

Their AI vs. Mafia (one of the top players)

https://www.youtube.com/watch?v=dXJUlqBsZtE&feature=emb_title&ab_channel=vladfi1

Github

https://github.com/vladfi1/phillip