Bowling with Deep Learning

Stanford University

Zizhen Jiang

Motivation

- DeepMind

[This image is from Nature]

- Games

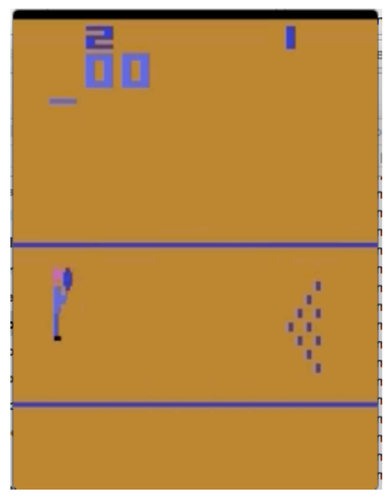
[This image is in the public domain]

Problem Statement

- Atari Games
  - agents
    - observation: image

      $x_t \in \mathbf{R}^d$
    - actions

      $a_t \in A = \{1, 2, \ldots, K\}$



Actions:
- up
- down
- throw

- Q(s, a), quantity of a state-matrix combination

$$Q(s_t, a_t) \leftarrow \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha_t}_{\text{learning rate}} \Big(\overbrace{\underbrace{r_{t+1}}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \underbrace{max_a Q(s_{t+1}, a)}_{\text{optimal future value}}}^{\text{learned value}} - \underbrace{Q(s_t, a_t)}_{\text{old value}} \Big)$$

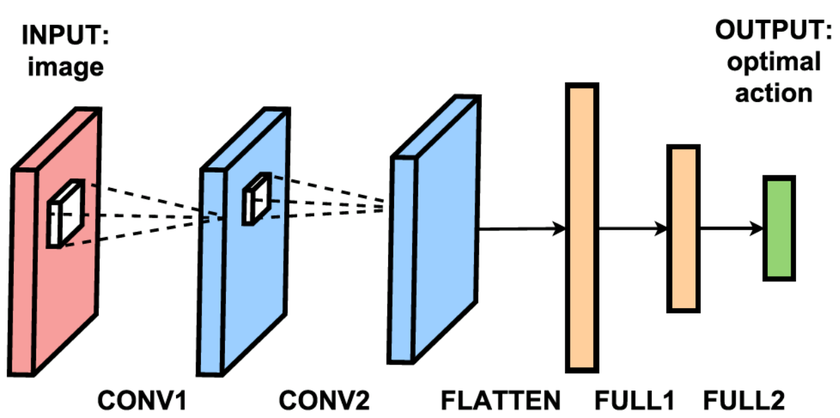
$$\theta_{t+1} \leftarrow \theta_t + \alpha_t (r_{t+1} + \gamma max_a Q_\theta(s_{t+1}, a) - Q_\theta(s_t, a_t)) \nabla_\theta Q(s_t, a)$$

- loss

$$L(\theta) = \mathbf{E}_{s_t, a, r, s_{t+1}} \left[r + \gamma max_{a \in A} Q_\theta(s_{t+1}, a) - Q_\theta(s_t, a_t) \right]^2$$

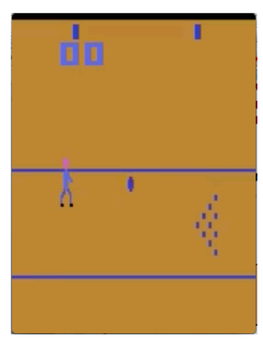
Technical Approach

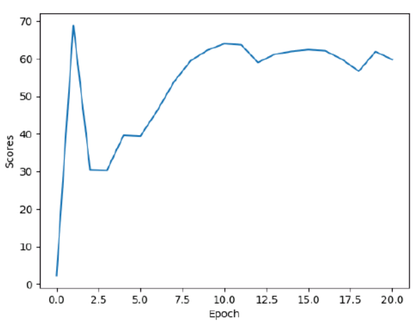
- Implementation
  - CS 234 template
  - Open.AI: environments
  - DeepMind: guideline
- Architecture



Results & Discussion

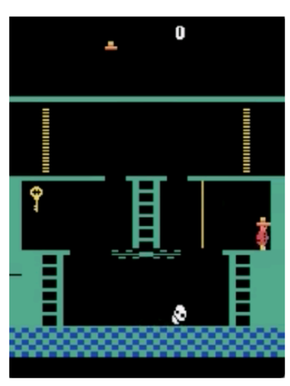
- Test
  - Bowling-v0
    - improved
    - overfitting

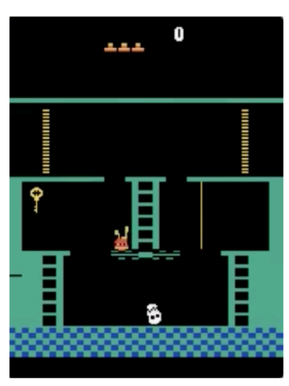




Game	Random	Human	DQN	Double DQN	My DQN
Bowling	23.10	154.80	42.40	70.50	59.80

  - MontezumaRevenge-v0
    - not improved
    - agent
      - > died
      - > stuck





On-Going Changes

- Smaller learning rate
- Change network architecture
  - change parameters
  - add more layers
    - batch normalization
    - dropout
- Double deep Q-learning

References

1. cs231n.stanford.edu
2. cs234.stanford.edu
3. https://en.wikipedia.org/wiki/Q-learning
4. V. Mnih, *et al.*, "Human-Level Control through Deep Reinforcement Learning," *Nature* 518, 519-533
5. V. Mnih, *et al.*, "Playing Atari with Deep Reinforcement Learning," *NIPS*, 2013
6. H. v. Hasselt, *et al.*, "Deep Reinforcement Learning with Double Q-Learning," AAAI, 2016