

Learning Montezuma's Revenge from a Single Demonstration (18.07)

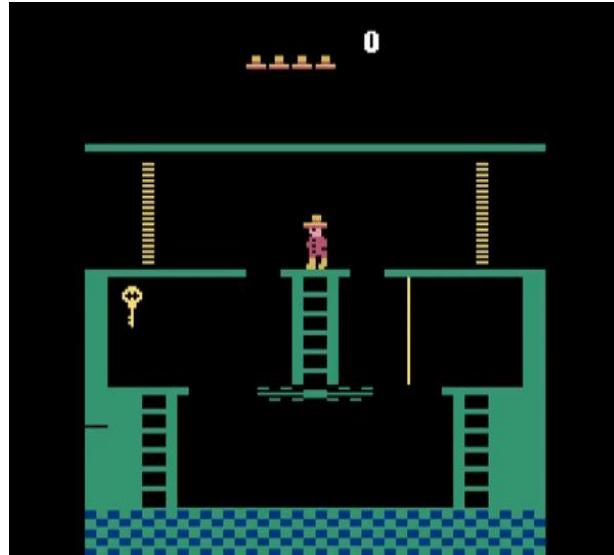
Ryan Lee

Exploration and Learning

- **Exploration:** Find action sequence with positive reward
- **Learning:** Remember and generalize action sequence
- Need both for a successful agent

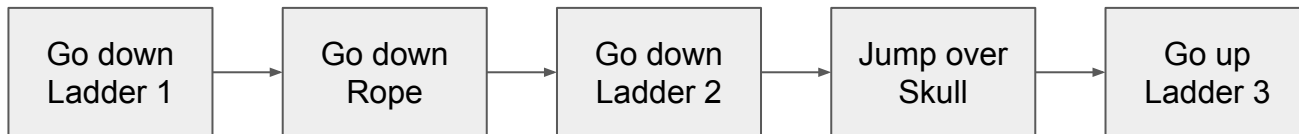
Montezuma's Revenge

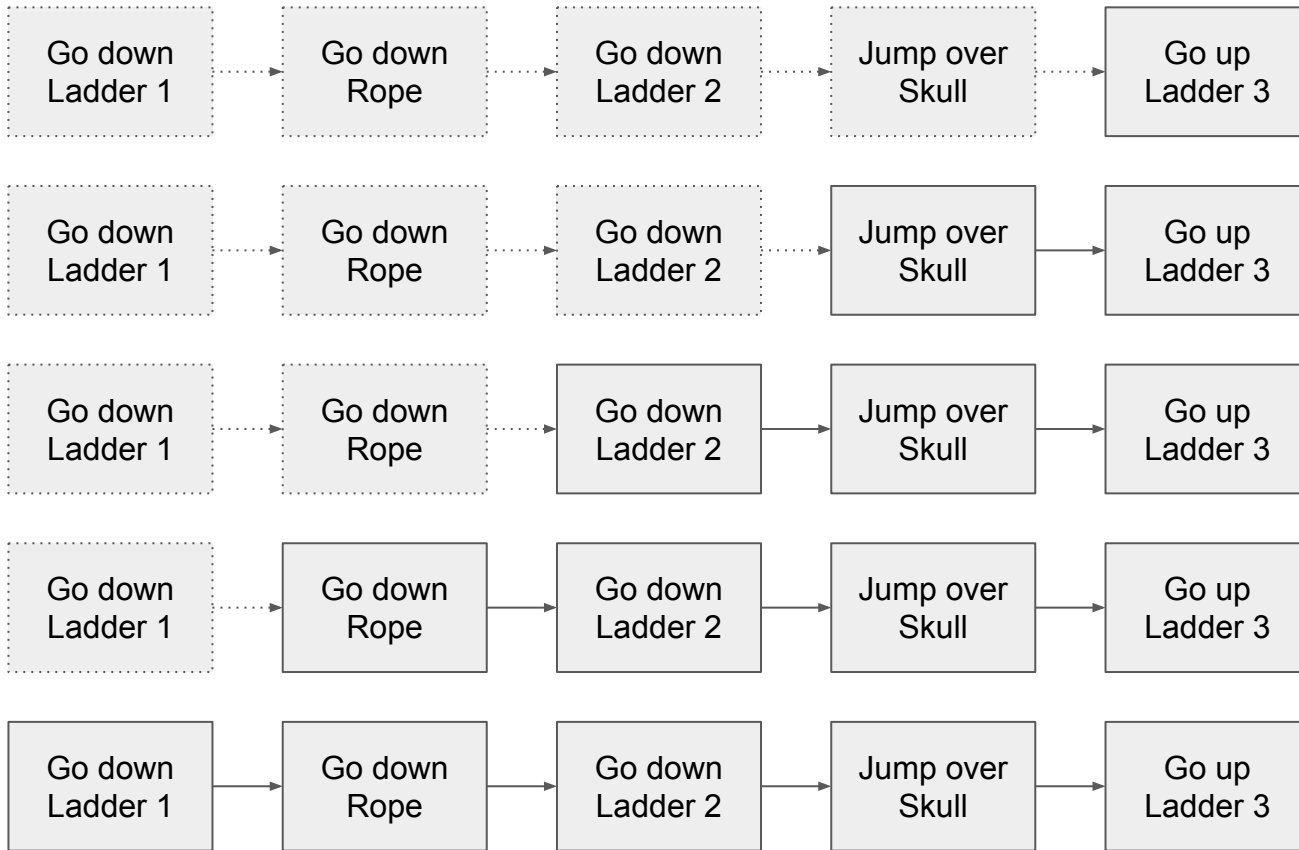
- One of the hardest games in Atari 2600
- Sparse rewards → Exploration is difficult

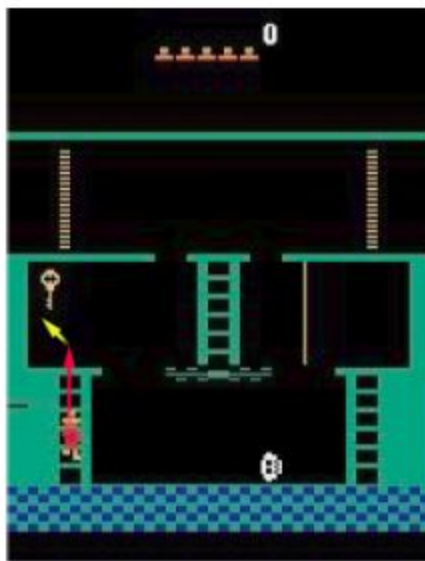


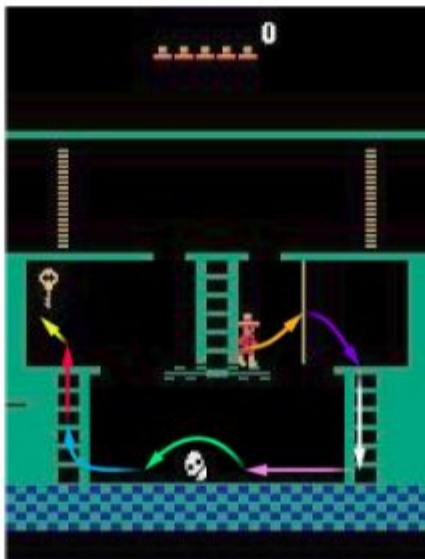
Simplifying Exploration with Demonstrations

- Solution: Shorten the episode
 - Start the agent near the end of demonstration
 - Train agent until it ties or beats the demonstrator's score
 - Gradually move starting point back in time









Result

- 74500 points on Montezuma's Revenge (State of the Art)
- Surpasses demo score of 71500
- Exploits emulator flaw



Comparison with DeepMind's approach

- DeepMind's approach
 - Less control over environment needed
 - Agents imitate the demo
- This approach
 - Need full game states in demo
 - Directly optimize game score → Less overfitting for sub-optimal demo
 - Better in multiplayer games where performance should be optimized against various opponents

Remaining Challenges

- Agent cannot reach exact state in demo
 - Agent needs to generalize between similar states
 - Problematic in *Gravitar* or *Pitfall*
- Careful hyperparameter tuning needed
- High variance in each run
- NN does not generalize as well as human

Thank you!

Original content by OpenAI

- [Learning Montezuma's Revenge from a Single Demonstration](#)

You can find more content in

- github.com/seungjaeryanlee
- www.endtoend.ai