Playing Minecraft with Efficient Deep Q-Learning from Demonstrations

Fan Yang, Keyu Li, Chenming Li, Zhaoting Li, Lin Shao, Jiankun Wang
Team: CU-SF

Abstract

In MineRL Competition 2020, participants are challenged to solve complex and hierarchical tasks with sparse rewards using human demonstrations without having access to human-readable actions.

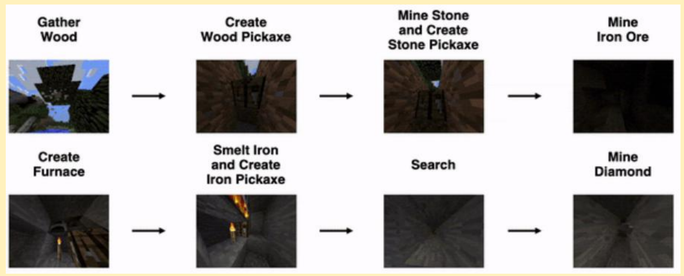
We approach this task by implementing some technical improvements to the DQfD algorithm, in order to better leverage the human demonstration data and realize efficient learning from demonstrations. The key techniques we use are:

- DQfD with some tricks in implementation, and
- Important action clustering.

We achieve 4-th place in Round 1.

Challenge Description

MineRL competition 2020:

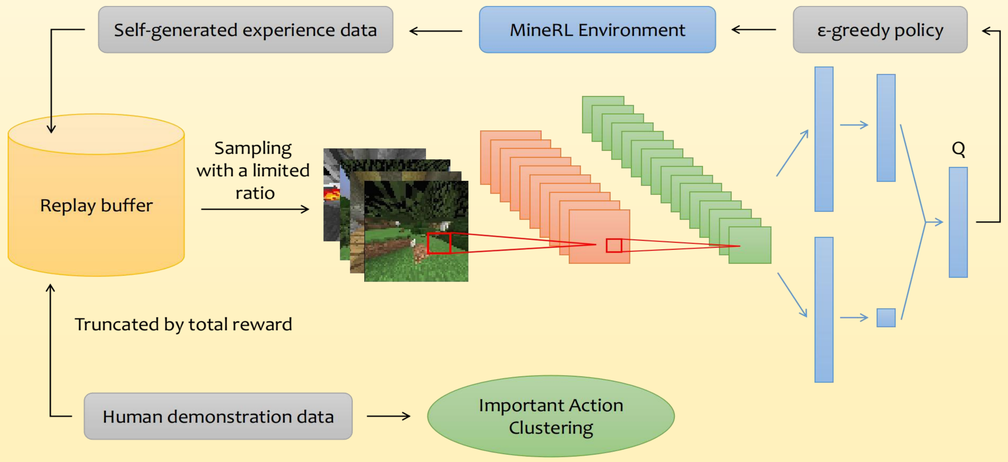


Participants compete to develop sample-efficient RL & IL algorithms to complete a hierarchical crafting-tree by harvesting resources and crafting items, and to finally obtain a diamond, using

- 8 million interactions in the game (provided by the MineRL simulator)
- 60 million frames of human demonstrations

No human-readable actions are provided in 2020.

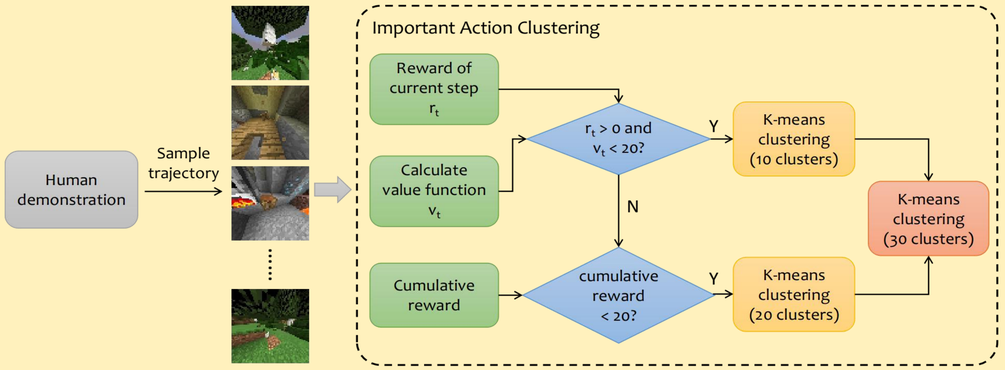
Deep Q-Learning from Demonstrations (DQfD)



Some technical improvements:

- Truncate the human demonstration data by total reward (<=70)
- Limit the sampling ratio between the agent's self-generated experience and demonstration data not to exceed 7:3
- Incorporate the ε-greedy policy in both the training and testing phase, to make the agent act randomly with a small probability

Important Action Clustering



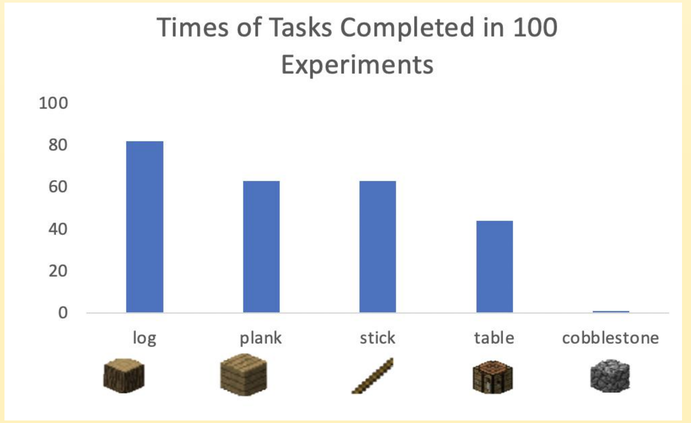
Process the human demonstration dataset to get important action set:

- Calculate the current step's reward, value function and cumulative reward
- Select the steps with instant reward >0 and value <20 to do K-means clustering → 10 important actions
- Select other steps with cumulative reward <20 to do K-means clustering → 20 less important actions
- The resulting 30 actions are clustered again to form the final action space

Performance

Avearge score: **6.47** (ranked **4-th** in Round 1)

In a total of 100 episodes, the agent can complete the following tasks:



Conclusion

On the basis of the DQfD algorithm, we incorporate some technical improvements to better leverage the human demonstration data and realize efficient learning from demonstrations, and achieve good performance in playing Minecraft.

香港中文大學 The Chinese University of Hong Kong
清華大學 Tsinghua University
Stanford University
南方科技大學 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Poster created by Keyu Li, PhD student at The Chinese University of Hong Kong (CUHK).