Double Dueling Agent for Dialogue Policy Learning

Yu-An Wang

https://github.com/MiuLab/E2EDialog

National Taiwan University

Microsoft Dialogue Challenge

- Double Dueling DQN

Movie Leaderboard

Rank	Model	Success Rate (Simulation)	Success Rate (Human)	Rating (Human)
1 Oct 25, 2018	Double Q *National Taiwan Unisversity*	41.8%	**31.1%**	**2.65/5**
1 Sep 20, 2018	DQN *single model*	**44.1%**	**30.8%**	**2.62/5**

Outline

- Variants of DQN
  - DQN
  - Double DQN
  - Dueling DQN
  - Prioritized DQN
  - Distributional DQN
- Exploration Strategies
  - Noisy DQN
  - Curiosity-based Exploration
- Experiments On Task-completion Dialogue Policy

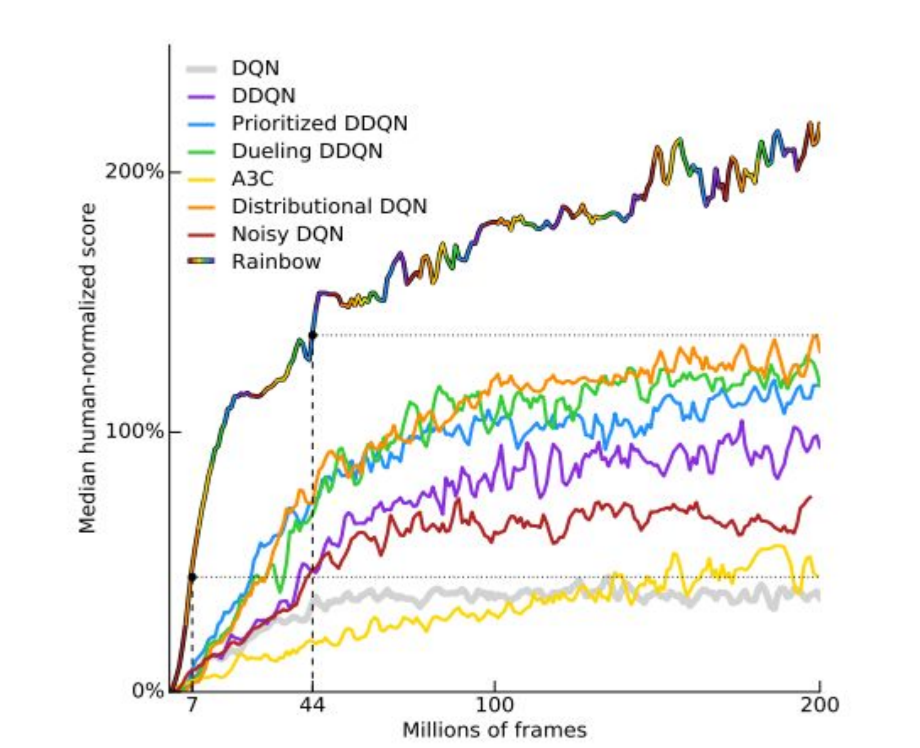
What is the BEST RL Algorithm for Dialogue Policy?

There are to many RL algorithms: Policy Gradient, Actor-Critic, DDPG, PPO, DQN, DDQN, Distributional DQN……etc

Rainbow

https://arxiv.org/pdf/1710.02298.pdf

- Combine 5 variants of DQN and test on Atari 2600



Deep Q-Networks (DQN)

- Value-based RL algorithm
- Learn a Q-Value function obeys a Bellman Equation

$$Q^*(s,a) = \mathbb{E}_{s'}\left[r + \gamma Q^*(s',a')|s',a'\right]$$

- Loss Function

$$L(\theta) = \mathbb{E}[(r + \gamma \max_{a'} Q(s',a',\theta') - Q(s,a,\theta))^2]$$

Double DQN and Dueling DQN

- Double DQN: Decouple selection and evaluation

$$L(\theta) = \mathbb{E}[(r + \gamma \max_{a'} Q(s', a', \theta') - Q(s, a, \theta))^2]$$

$$\downarrow$$

$$L(\theta) = \mathbb{E}[(r + \gamma Q(s', \arg\max_{a'} Q(s', a', \theta), \theta') - Q(s, a, \theta))^2]$$

- Dueling DQN: Split Q-value into advantage function and value function

$$Q(s, a) = A(s, a) + V(s) - \frac{1}{N_{actions}} \sum_i A(s, a_i)$$

Distributional DQN (Categorical DQN)

- Learn the distribution of value function
- Use a set of atoms to model a discrete distribution

$$\{z_i = V_{min} + i(\frac{V_{max} - V_{min}}{N-1}) | 0 \leq i \leq N\}$$

- Project the target distribution on the support vector, then minimize KL-divergence

$$L(\theta) = D_{KL}(\Phi \hat{\mathcal{T}} Z_{\theta'}(s,a) || Z_\theta(s,a))$$

Prioritized DQN

- Assign every transition a priority in replay buffer

$$p_i = \left| r + \gamma \max_{a'_i} Q(s'_i, a'_i, \theta') - Q(s_i, a_i, \theta) \right|^{\alpha}$$

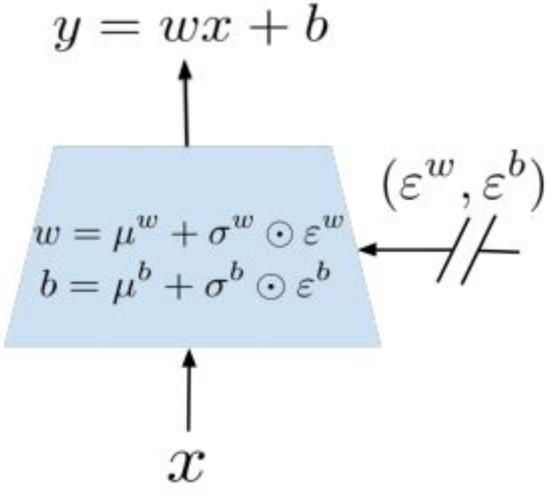
- Sample transitions with probability according to priorities

Exploration Strategies

- Noisy Network
- Curiosity-based Exploration

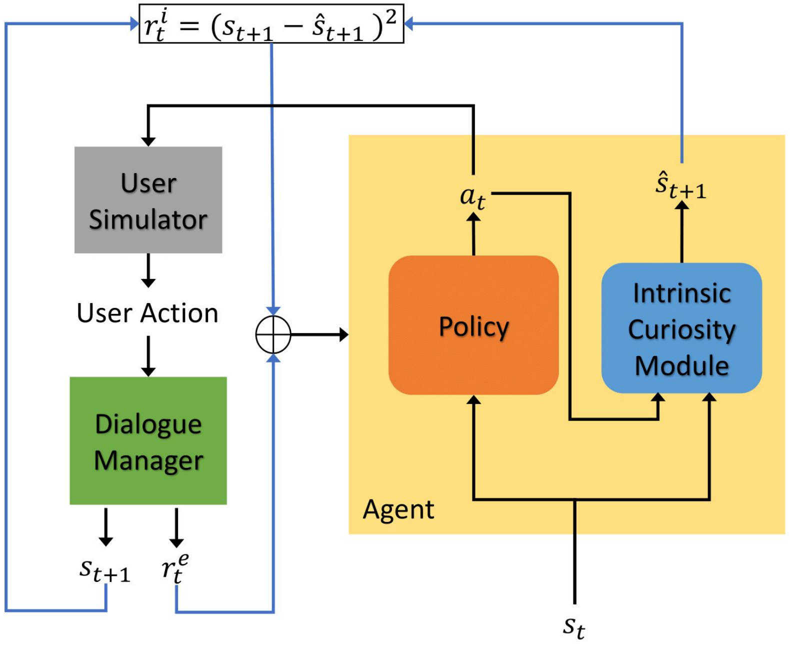
Noisy DQN

- Add noise in linear layer to induce stochastic exploration



Curiosity-based Exploration

- Use error of next state's prediction as intrinsic reward
- High error -> the state is novel for the agent



Experiments

- Variants of DQN
- Exploration strategies

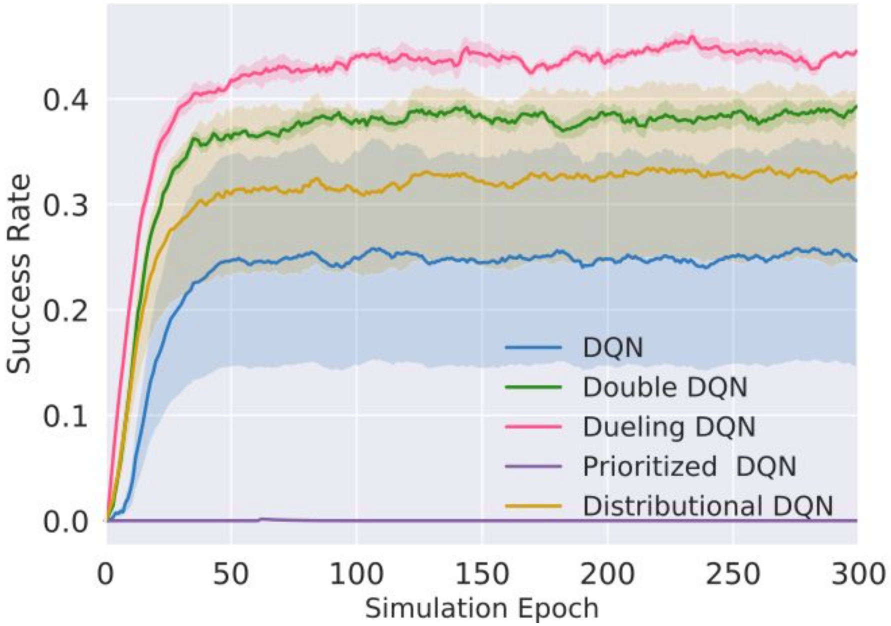
Setup

- Task: Movie-Ticket Booking
- Each model trained 5 times with different random seeds

Movie-Ticket Booking Task
usr: Can I get tickets for zoolander 2 tomorrow?
agt: Which city would you like?
usr: I want to watch at seattle.
agt: How many tickets do you need?
usr: I want 2 tickets please!
agt: 9:25 pm is available.
usr: I want to watch at regal meridian 16.
agt: Great - I was able to purchase 2 tickets for you to see zoolander 2 tomorrow at regal meridian 16 theater in seattle at 9:25 pm.
usr: Thank you.
agt: Thank you.
Success

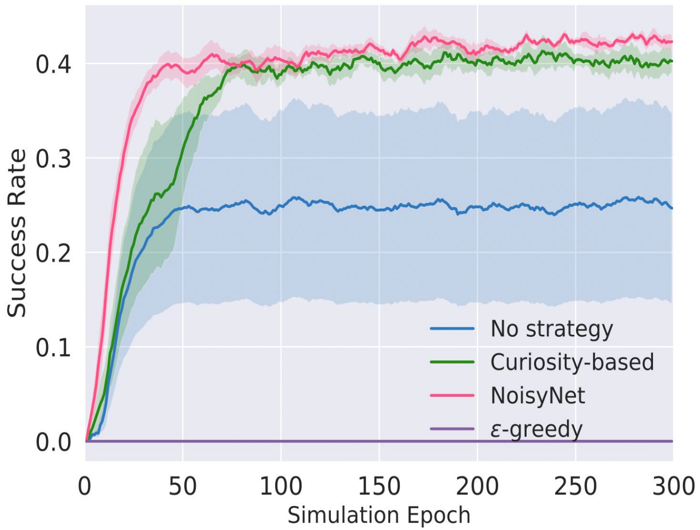
Variants of DQN

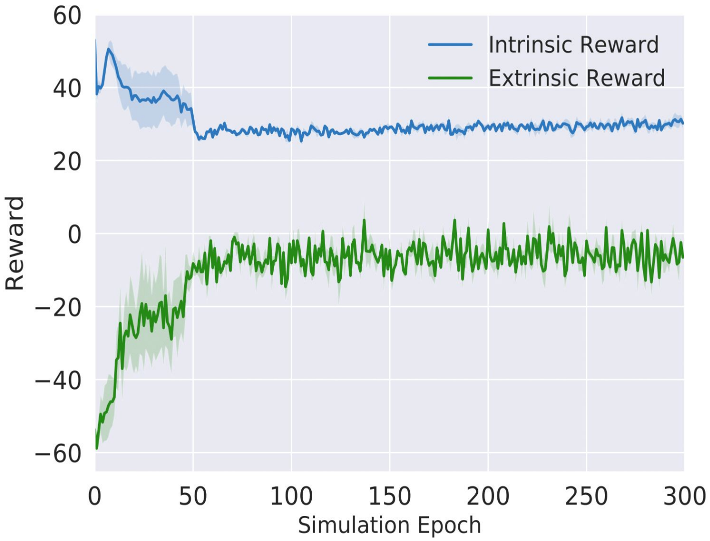
- Dueling DQN performs best
- DQN and Distributional DQN sometimes fail
- Prioritized DQN always fails
- Final choice: Double + Dueling



Exploration Strategies

- Choosing a suitable exploration strategy can make training more stable





Conclusions

- Dueling DQN performs best in this task
- Suitable exploration strategies can make training more stable

Thanks for Listening

The code is available here: https://github.com/MiuLab/E2EDialog

The paper with more details *Investigating Variants of Deep Q-Networks for Task-Completion Dialogue Policy* will be available on arxiv soon.

References

- Hessel, Matteo, et al. "Rainbow: Combining improvements in deep reinforcement learning." arXiv preprint arXiv:1710.02298 (2017).
- Van Hasselt, Hado, Arthur Guez, and David Silver. "Deep Reinforcement Learning with Double Q-Learning." AAAI. Vol. 2. 2016.
- Wang, Ziyu, et al. "Dueling network architectures for deep reinforcement learning." *arXiv preprint arXiv:1511.06581* (2015).
- Schaul, Tom, et al. "Prioritized experience replay." *arXiv preprint arXiv:1511.05952* (2015).
- Bellemare, Marc G., Will Dabney, and Rémi Munos. "A distributional perspective on reinforcement learning." *arXiv preprint arXiv:1707.06887* (2017).
- Fortunato, Meire, et al. "Noisy networks for exploration." *arXiv preprint arXiv:1706.10295* (2017).
- Pathak, Deepak, et al. "Curiosity-driven exploration by self-supervised prediction." *International Conference on Machine Learning (ICML)*. Vol. 2017. 2017.