

# COUNT-BASED EXPLORATION

*Jonathan Campbell*

*COMP-767*

*April 20, 2017*



# WHY EXPLORE?

- With exploration:
  - Better chance of finding optimal policy (e.g.: k-armed bandit)
  - All actions sampled infinitely often in limit: guarantees  $Q^*$  convergence.
- Simple approaches:  $\epsilon$ -greedy/softmax action selection
- Exploration is hard in environments with:
  - noisy rewards
  - nonstationarity
  - very large state spaces



# DELAYED Q-LEARNING W/ INTERVAL ESTIMATION

- Keep running average of update to  $Q(s, a)$ .
- If difference between current  $Q(s, a)$  and average is larger than  $\epsilon$ , install update and reset average.
- Add exploration bonus of form  $\frac{\beta}{\sqrt{\#(s, a)}}$
- If  $m$  samples of  $(s, a)$  are reached without installing, reset average as well as  $\#(s, a)$  to 0.
- Performs better with larger  $m$ .

Alexander Strehl

Probably Approximately Correct (PAC) Exploration in Reinforcement Learning (2007)



# STATE HASHING

- With continuous and/or large state spaces, can't keep counts.
  - May never observe the same exact state more than once.
- Possible solution:
  - Use hash fn. to discretize state and maintain counts of  $\Phi(s)$ .
  - SimHash:  $\phi(s) = \text{sgn}(Ag(s)) \in \{-1, 1\}^k$
  - Add exploration bonus to reward and use regular Q/etc alg.:

$$r^+(s, a) = \frac{\beta}{\sqrt{n(\phi(s))}}$$

Tang et al.

#Exploration: A Study of Count-Based Exploration for Deep Reinforcement Learning (2017)



# DENSITY MODELS

- Other solution:
  - Replace  $\#(s, a)$  with density model over state space.
    - Density model gives probability  $p(x)$  for state  $x$ , and  $p'(x)$ : prob. of  $x$  after observing a new occurrence of  $x$  ('recoding').
    - Pseudo-count: 
$$\hat{N}(x) = \frac{p(x) (1 - p'(x))}{p'(x) - p(x)}$$
    - Add exploration bonus to reward (same as last slide).
- Difficulty with these approaches:
  - State must first be visited once before bonus can be applied.

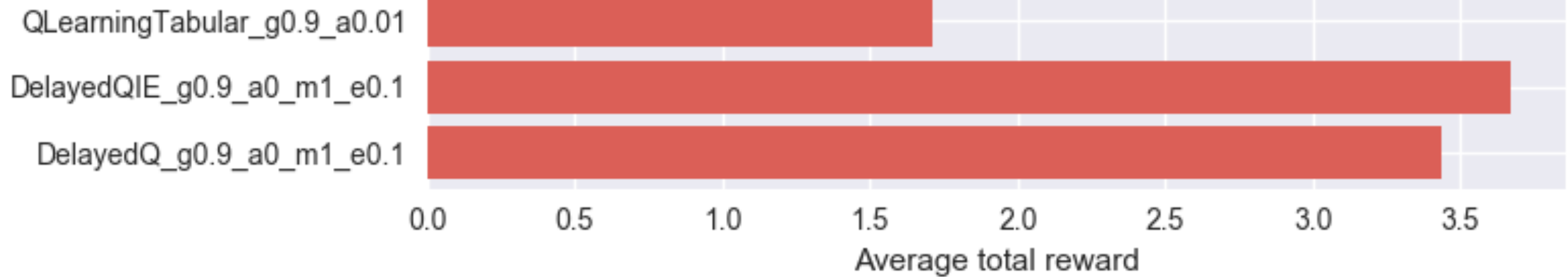
*Bellemare et al.*

*Unifying Count-Based Exploration and Intrinsic Motivation (2016)*



# SAMPLE RESULTS

nethack\_combat\_giantbat-arrow



nethack\_combat\_giantbat-arrow

