Count-Based Exploration in Feature Space for Reinforcement Learning

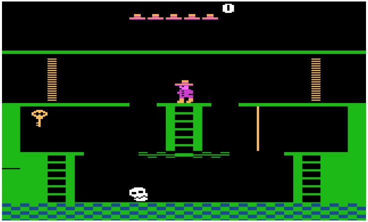
J. Martin S. Narayanan S. T. Everitt M. Hutter

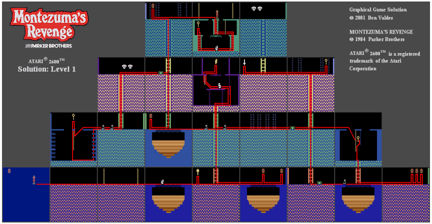
Research School of Computer Science
Australian National University

IJCAI, 2017

Efficient exploration is still an open problem in MDPs with:

- Large state spaces
- Sparse rewards





Novelty-Based Exploration in Large MDPs

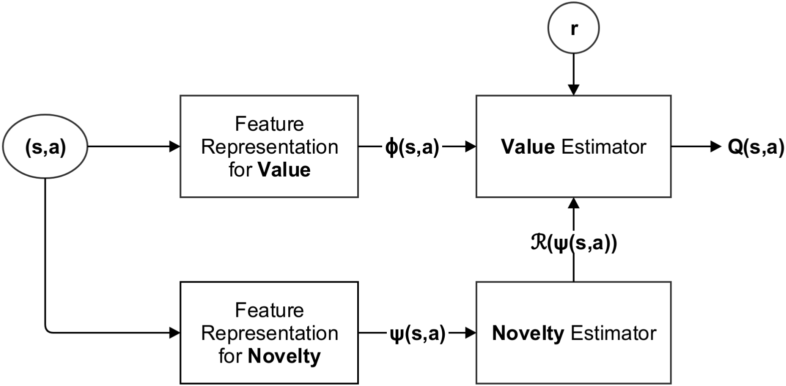
How do you explore efficiently?

- Encourage the agent to visit **novel** states to maximally reduce its uncertainty. How?
- Make your agent **curious about states with novel features**
    1. Choose a feature representation $\psi(s, a)$ of the state space
    2. Compute a visit pseudocount $\hat{N}(\psi)$
    3. Compute a novelty-based exploration bonus:

    $$\mathcal{R}(\psi) \propto \frac{1}{\sqrt{\hat{N}(\psi)}}$$

    4. Add the bonus to the reward r
    5. Train the agent with the augmented reward $r^+ = r + \mathcal{R}(\psi)$

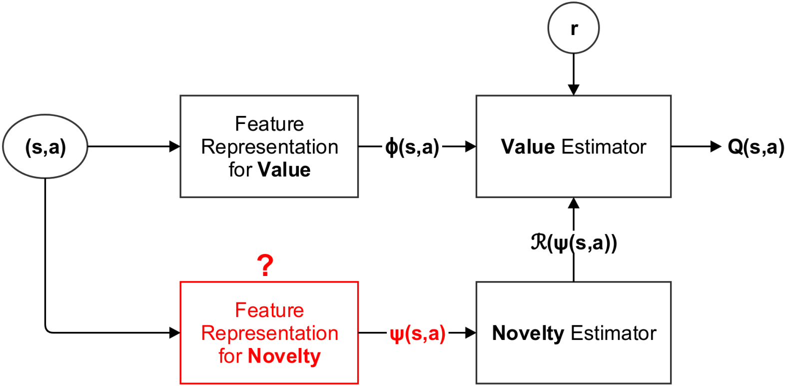
Novelty-Based Exploration in Large MDPs



Feature Representations for Novelty from previous work:

- Context-Tree Switching (CTS) Density Model (Google DeepMind) [1]
- #-Exploration (Berkeley) [4]
- Neural Density Model (Google DeepMind) [3]

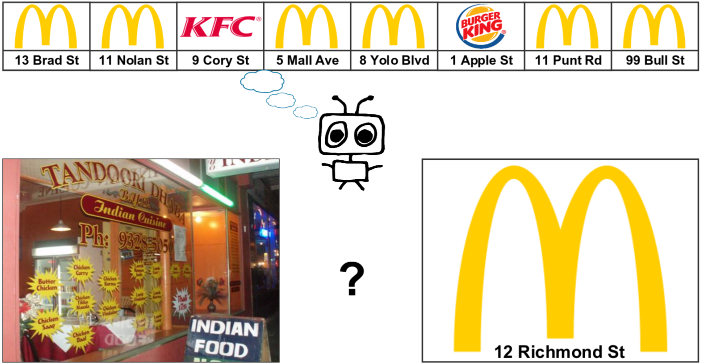
Novelty-Based Exploration in Large MDPs



Problem:

- **Which feature representation is appropriate for measuring the novelty of a state?**
- Previous works do not justify their choices

Which features are relevant when measuring novelty?

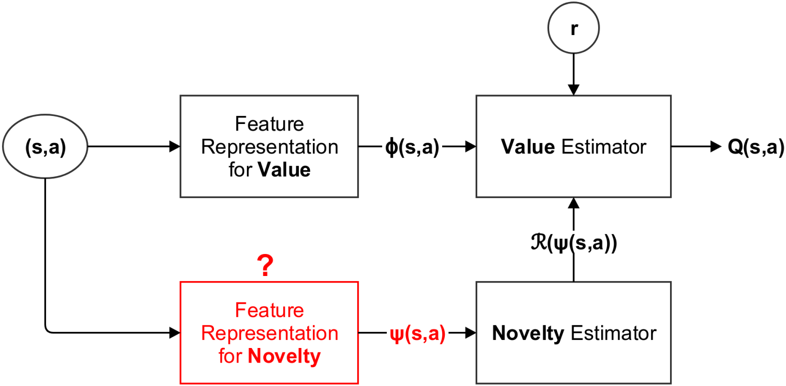


- Different flavours
- Different drinks menu

- Same flavours
- Same drinks menu

Irrelevant features: Wallpaper, Parking, Lighting, Floorspace, Address...

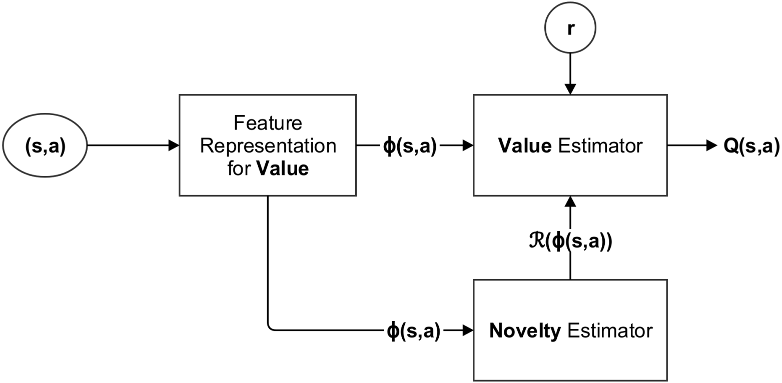
Previous Works do not use Value-Relevant Features



Problem:

- In this architecture, the feature representation used for novelty estimation may not capture **value-relevant features**
- So which features are relevant for maximising value?

The ϕ-Exploration Bonus Algorithm (ϕ-EB)



- Our novelty estimator assigns a high exploration bonus to states that have **novel, value-relevant features**
- Our ϕ-**Exploration Bonus** algorithm is simpler and less computationally expensive than previous approaches

The ϕ-Exploration Bonus Algorithm (ϕ-EB)

Require: β, t_{end}

while $t < t_{\text{end}}$ **do**

Observe r_t and features $\phi(s)$ for the current state s

Compute joint feature probability $\rho_t(\boldsymbol{\phi}) := \prod_i^M \rho_t^i(\phi_i)$

for i in $\{1,\ldots,\text{M}\}$ **do**

Update each probability ρ_{t+1}^i with observed feature ϕ_i

end for

Recompute joint probability $\rho_{t+1}(\boldsymbol{\phi}) := \prod_i^M \rho_{t+1}^i(\phi_i)$

Compute the ϕ-pseudocount $\hat{N}_t^{\phi}(s) := \dfrac{\rho_t(\boldsymbol{\phi})(1-\rho_{t+1}(\boldsymbol{\phi}))}{\rho_{t+1}(\boldsymbol{\phi})-\rho_t(\boldsymbol{\phi})}$

Compute the exploration bonus $\mathcal{R}_t^{\phi}(s,a) := \dfrac{\beta}{\sqrt{\hat{N}_t^{\phi}(s)}}$

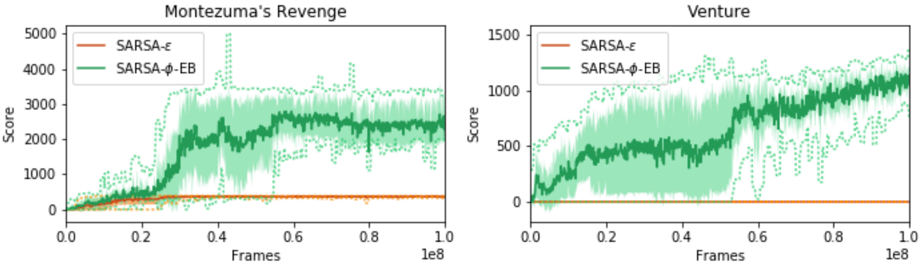
Add the bonus to the reward $r_t^+ := r_t + \mathcal{R}_t^{\phi}(s,a)$

Pass $\phi(s)$, r_t^+ to RL algorithm to update $\boldsymbol{\theta}_t$

end while

return $\boldsymbol{\theta}_{t_{\text{end}}}$

Empirical Evaluation



	Venture	**Montezuma**
Sarsa-ϕ-EB (100M)[2]	1169.2	2745.4
Sarsa-ϵ (100M)	0.0	399.5
DDQN-PC (100M)[1]	86.4	**3459**
A3C+ (200M)[1]	0	142
TRPO-Hash (200M)[4]	445	75

Trained ϕ-EB agent playing Atari

Summary

The ϕ-Exploration Bonus algorithm:

- A new count-based optimistic exploration algorithm for RL
- Feasible in **high-dimensional** state spaces
- **Simpler** and **cheaper** than competitive methods
- Near **state-of-the-art results** across the ALE

- Our ϕ-pseudocount measures uncertainty/novelty by exploiting the **same features used for value function approximation**

Outlook

- Use with **nonlinear** function approximation
- Combination with Deep RL algorithms

Further Reading I

Marc G. Bellemare, Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Rémi Munos.
Unifying count-based exploration and intrinsic motivation.
CoRR, abs/1606.01868, 2016.

Jarryd Martin, Suraj Narayanan S., Tom Everitt, and Marcus Hutter.
Count-Based Exploration in Feature Space for Reinforcement Learning.
In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 2017.

Georg Ostrovski, Marc G. Bellemare, Aäron van den Oord, and Rémi Munos.
Count-based exploration with neural density models.
CoRR, abs/1703.01310, 2017.

Further Reading II

Haoran Tang, Rein Houthooft, Davis Foote, Adam Stooke, Xi Chen, Yan Duan, John Schulman, Filip De Turck, and Pieter Abbeel.
#Exploration: A study of count-based exploration for deep reinforcement learning.
CoRR, abs/1611.04717, 2016.

Contact

- **Email:** mail@jarrydmartin.com
- **Paper Title:** Martin et. al. "Count-Based Exploration in Feature Space for Reinforcement Learning" (2017)
- **Available at:** IJCAI-17 proceedings, arxiv

- **Shameless self-promotion:** This was my MSc dissertation, I am currently applying for **PhDs** and **Research Internships**. Hire me!