# Count-Based Exploration in Feature Space for Reinforcement Learning

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

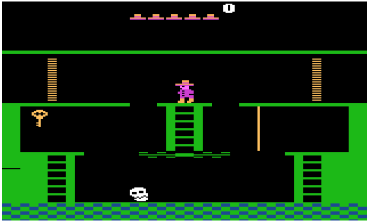
Research School of Computer Science
Australian National University

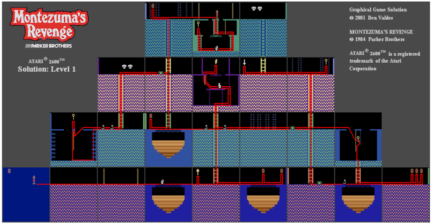
SURL workshop, ECML PKDD
September, 2017

# The Exploration/Exploitation Dilemma

Efficient exploration is still an open problem in MDPs with:

- Large state spaces
- Sparse rewards





# Novelty-Based Exploration in Large MDPs

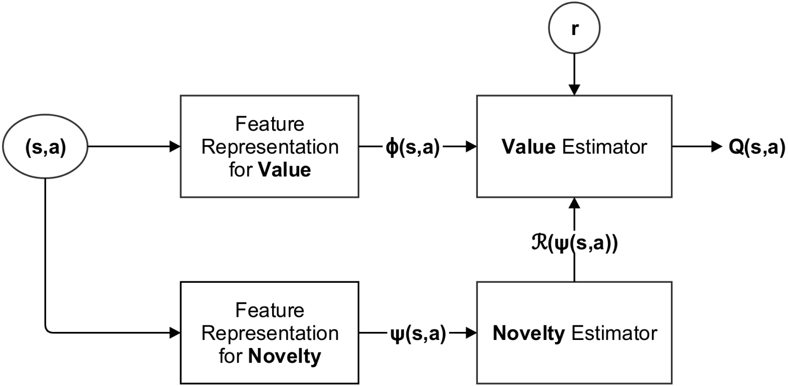
How do you explore efficiently?

- 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 $r^{+} = r + \mathcal{R}(\psi)$

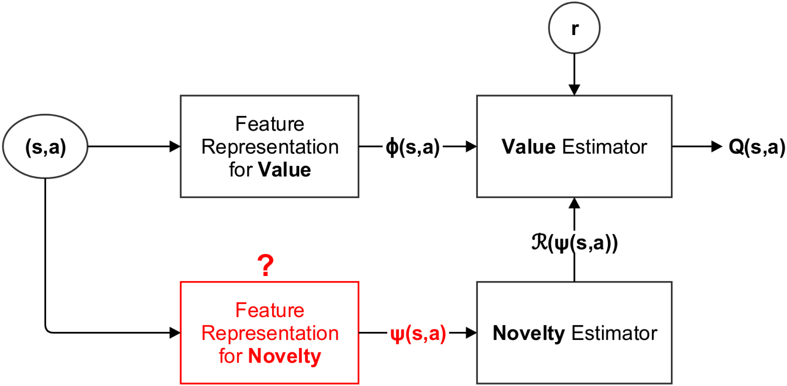
# Novelty-Based Exploration in Large MDPs


r
(s,a)
Feature Representation for **Value**
ϕ(s,a)
**Value** Estimator
Q(s,a)
ℛ(ψ(s,a))
Feature Representation for **Novelty**
ψ(s,a)
**Novelty** Estimator

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]

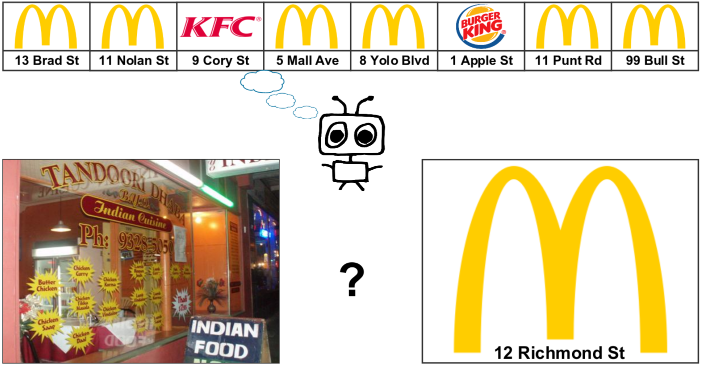
# Novelty-Based Exploration in Large MDPs



**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?

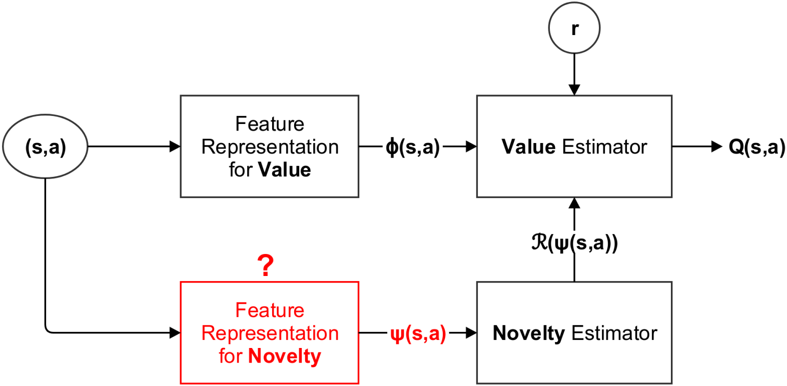


- Different flavours
- Different drinks menu

- Same flavours
- Same drinks menu

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

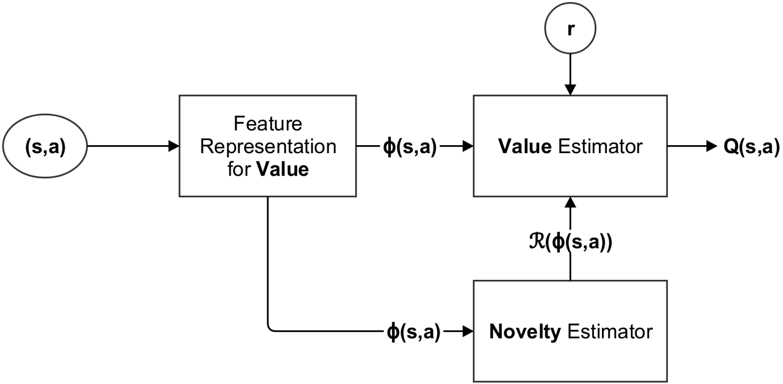
# Previous Works do not use Value-Relevant Features


r
(s,a)
Feature Representation for **Value**
ϕ(s,a)
**Value** Estimator
Q(s,a)
ℛ(ψ(s,a))
?
Feature Representation for **Novelty**
ψ(s,a)
**Novelty** Estimator

**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 $\phi$-Exploration Bonus Algorithm ($\phi$-EB)

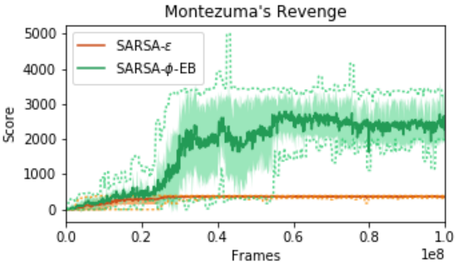


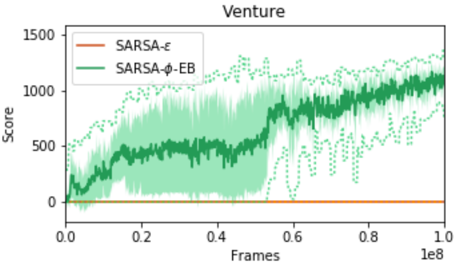
- Our novelty estimator assigns a high exploration bonus to states that have **novel, value-relevant features**
- Our $\phi$-**Exploration Bonus** algorithm is simpler and less computationally expensive than previous approaches

# The $\phi$-Exploration Bonus Algorithm ($\phi$-EB)

```
Require: $\beta$, $t_{\text{end}}$
  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 $\{1,\ldots,M\}$ do
          Update each probability $\rho_{t+1}^i$ with observed feature $\phi_i$
      end for
      Recompute joint probability $\rho_{t+1}(\boldsymbol{\phi}) := \prod_i^M \rho_{t+1}^i(\phi_i)$
      Compute the $\phi$-pseudocount $\hat{N}_t^{\phi}(s) := \frac{\rho_t(\boldsymbol{\phi})(1 - \rho_{t+1}(\boldsymbol{\phi}))}{\rho_{t+1}(\boldsymbol{\phi}) - \rho_t(\boldsymbol{\phi})}$
      Compute the exploration bonus $\mathcal{R}_t^{\phi}(s,a) := \frac{\beta}{\sqrt{\hat{N}_t^{\phi}(s)}}$
      Add the bonus to the reward $r_t^+ := r_t + \mathcal{R}_t^{\phi}(s,a)$
      Pass $\boldsymbol{\phi}(s)$, $r_t^+$ to RL algorithm to update $\boldsymbol{\theta}_t$
  end while
   return $\boldsymbol{\theta}_{t_{\text{end}}}$
```

# Empirical Evaluation





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

# Trained $\phi$-EB agent playing Atari

Switch to dedicated video-player, if flash fails to load video.

# 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.