A Motivating Example

Focus of Attention: Batch RL

Problem Setting
Given \( T = \{Q^*(s,a) \mid \forall s \in T, \forall a \in A\} \)
Induce \( \hat{\pi}^* \approx \pi^* \) (a classification problem)

An Importance-sensitive approach

\[
V(\pi^*) - V(\pi) \leq \frac{1}{1 - \gamma} \sum_{s \in S} (G^*(s) \cdot I(\hat{\pi}^*(s) \neq \pi^*(s)))
\]

Focus of Attention: Online RL

Problem Setting
Given online interaction/simulations
Compute \( \hat{\pi}^* \approx \pi^* \) (e.g., via policy iteration)

Basic learning framework
Classification-based approximate policy iteration with rollouts [Lagoudakis & Parr 03]

An Importance-sensitive approach

In policy iteration, \( \pi \rightarrow V^\pi \rightarrow \pi' \) (greedy policy)
Approximate greedy policy: \( \hat{\pi}' \)

\[
\text{If } \forall s \in S, \quad d_{\pi',D}(s) - d_{\hat{\pi}',D}(s) \leq \varepsilon \quad \text{then } V(\pi') - V(\hat{\pi}') \leq \frac{1}{1 - \gamma} \sum_{s \in S} G^*(s) \cdot I(\hat{\pi}'(s) \neq \pi'(s)) + \varepsilon \sum_{s \in S} G^*(s)
\]

The Philosophy

Not all states are equally important in terms of affecting the policy value

Focusing on important states to produce better policy
• higher policy value
• faster convergence

Notation

\( S \): state space; \( A \): action space;
\( \gamma \): discount factor; \( D \): start-state distribution
\( V^\pi(s), V^*(s), Q^\pi(s,a), Q^*(s,a) \): value functions
\( V(\pi) = E_{s \sim D} \{V^\pi(s)\} \): policy value
\( d_{\pi,D}(s) = \sum \gamma' \Pr[s_i = s \mid D, \pi] \)

Policy as a classifier:
state \( \xrightarrow{\pi} \) action

Experiments

• \( N \) by \( N \) grid world
• Randomly distributed rewards
• Two actions:
  • north-east
  • south-east

Experiment: Batch RL

Experiment: Online RL

Future Work

• Estimate \( d_{\pi,D}(s) \) online
• Multi-action case
• Application to real-world problems

Contributions

• focusing attention in reinforcement learning
• definition of state importance
• improved importance-sensitive algorithms
• better policy, faster convergence