Intrinsically Motivated Reinforcement Learning

Satinder Singh | Andrew G. Barto | Nuttapong Chentanez

Amna Liaqat

Motivation

- Extrinsic motivation
  - Reward comes from reaching an external goal
  - Develops specific competence
- Intrinsic motivation
  - Reward comes from within individual
  - Encourages play and exploration
  - Develops broad competence

Intrinsically Motivated RL

- Skills are building blocks used to find solutions to new problems
- Typical machine learning algorithms apply to a single problem
  - Not flexible enough to cope with new problems over time
  - Appropriate reward functions difficult to define
- Dopamine is the brain's intrinsic motivator
- Less dopamine fires when a behavior becomes familiar
  - Novelty is intrinsically rewarding
- Novelty can drive development of reusable skills to solve complex problems
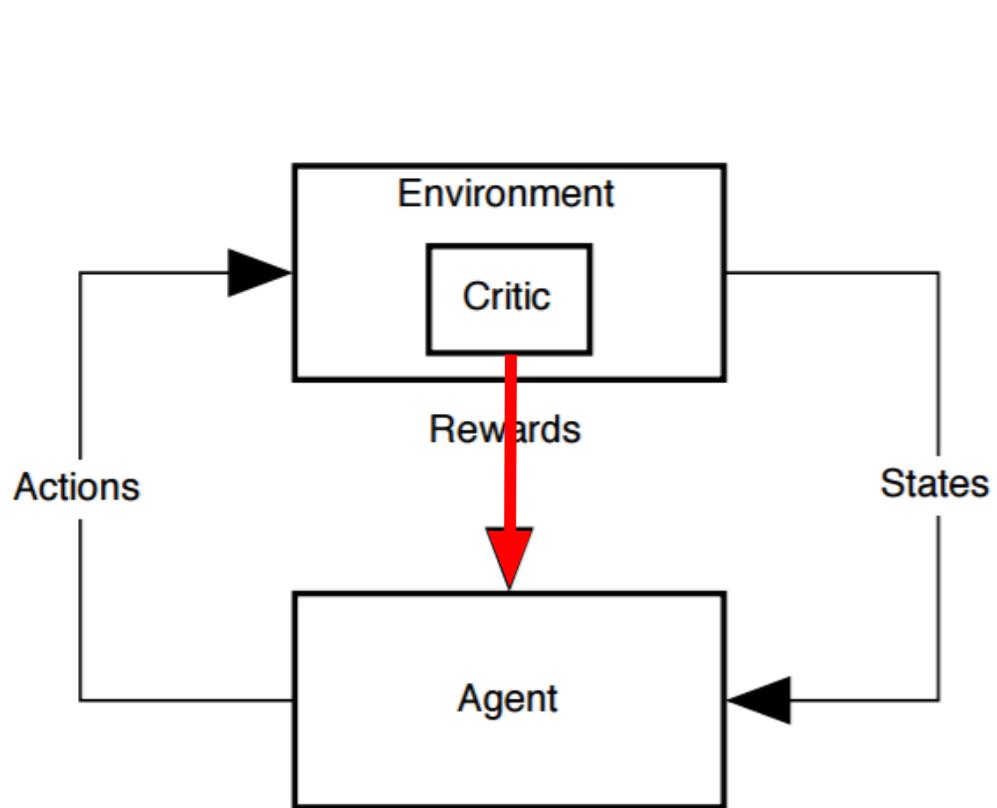
Rewards from external critic motivate agent
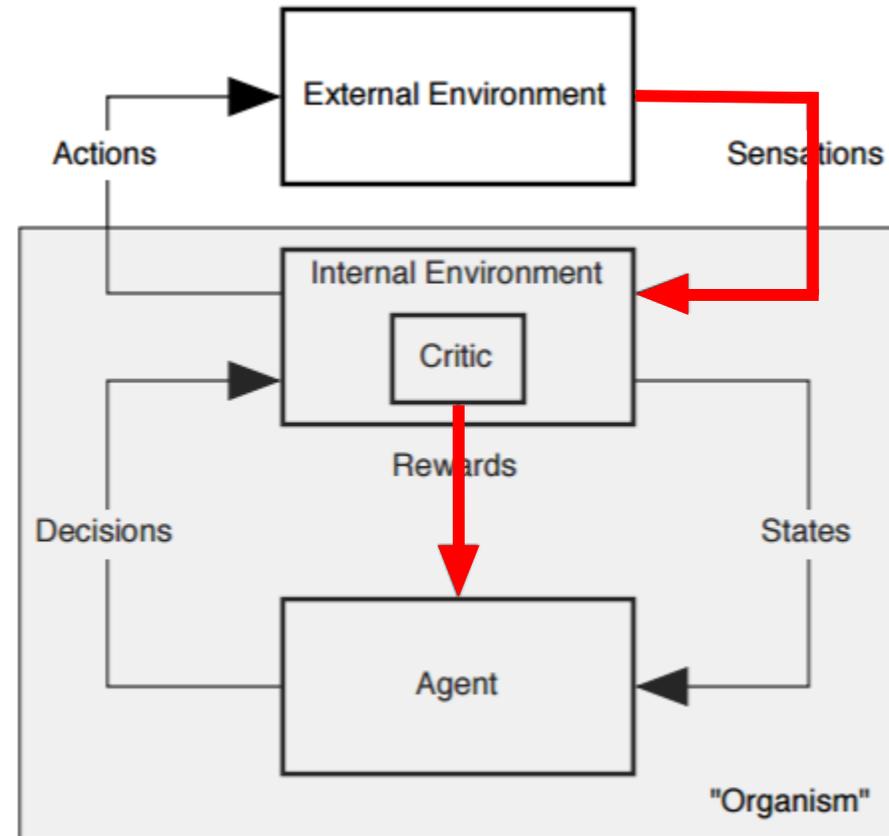
External sensations trigger rewards from internal critic motivate agent

What are Skills?

- Theory of options
  1. Option policy: Novelty is intrinsically rewarding
  2. Initiation Set: States in which the option can be initiated
  3. Termination Condition: Conditions under which the option terminates
- Not a sequence of actions
- Closed-loop control rule: respond to ongoing state changes
- Must be provided by system designer
- Development should result in collection of reusable skills

Options Framework

1. Option Models
   - Probabilistic effects of executing an option
   - Gives probability of option terminating at a state and total expected reward
   - Can be learned from experience
2. Intra-option Learning Methods
   - Allows for simultaneously updating option policies during agent interaction with environment

Developing Hierarchical Collections of Skills

- Children accumulate skills through intrinsically motivated play
- Behaviour that results in interesting consequences is remembered to potentially be used for future goals
- Behaviour is repeated to improve efficiency until the child is bored and moves on
- Claim: An option model can model the behavior of a child at play

Intrinsically Motivated RL

Behaviour
- Initially only primitive actions available to agent
- Skills and their models become available over time as action choices

Salient events
- Agent has intrinsic notion of interesting events

Reward
- Intrinsic reward generated by critic for salient events
- Similar to dopamine neurons firing for novel events
- Proportional to error in prediction of learning model

Skill-KB
- Knowledge base of skills

Playroom domain

Objects
- Light switch
- Ball
- Bell
- Two moveable blocks/buttons
- Toy monkey that makes sounds

Playroom domain

Agent
- Hand
- Eye
- Visual Marker

Playroom domain

Actions

- Move eye to hand
- Move eye to marker
- Move eye one step north, south, east or est
- Move eye to random object
- Move hand to eye
- Move marker to eye
- Natural operations also available
  - If hand and eye on light switch, then light can be flicked on/off
  - If hand and eye on ball, ball can be kicked in straight line to marker

Playroom domain

Environment

- Bell rings and makes a random step if ball is kicked to it

- Light switch controls room lighting

- Block colors only visible in light

- Blue block turns music on, red turns off

- Monkey makes noise if room is dark, music on, bell rings

Playroom domain

To get the monkey to cry out, the agent must:

1. get its eye to the light switch
2. move hand to eye
3. push the light switch to turn the light on
4. find the blue block with its eye
5. move the hand to the eye
6. press the blue block to turn music on
7. find the light switch with its eye

Playroom domain

8. move hand to eye

9. press light switch to turn light off

10. find the bell with its eye

11. move the marker to the eye

12. find the ball with its eye

13. move its hand to the ball

14. kick the ball to make the bell ring

Playroom domain

- Final task of getting monkey to cry out is easier if other skills are learned

- Ex. Turn light on/off, turn music on, ring bell

- Changes in light and sound are salient

- Initial action value function is uninformative, so exploration is random

- Salient events initiates learning of option and option model

- Agent repeats salient event, which improves its policy and option model

- When intrinsic reward diminishes, agent moves on

Playroom domain results

- Simpler salient events occur earlier
- Skills for simpler tasks mastered earlier
- Events occur despite diminished reward as they are needed for more complex events
- Complex skills learned quickly once simple skills are mastered – skills are built upon
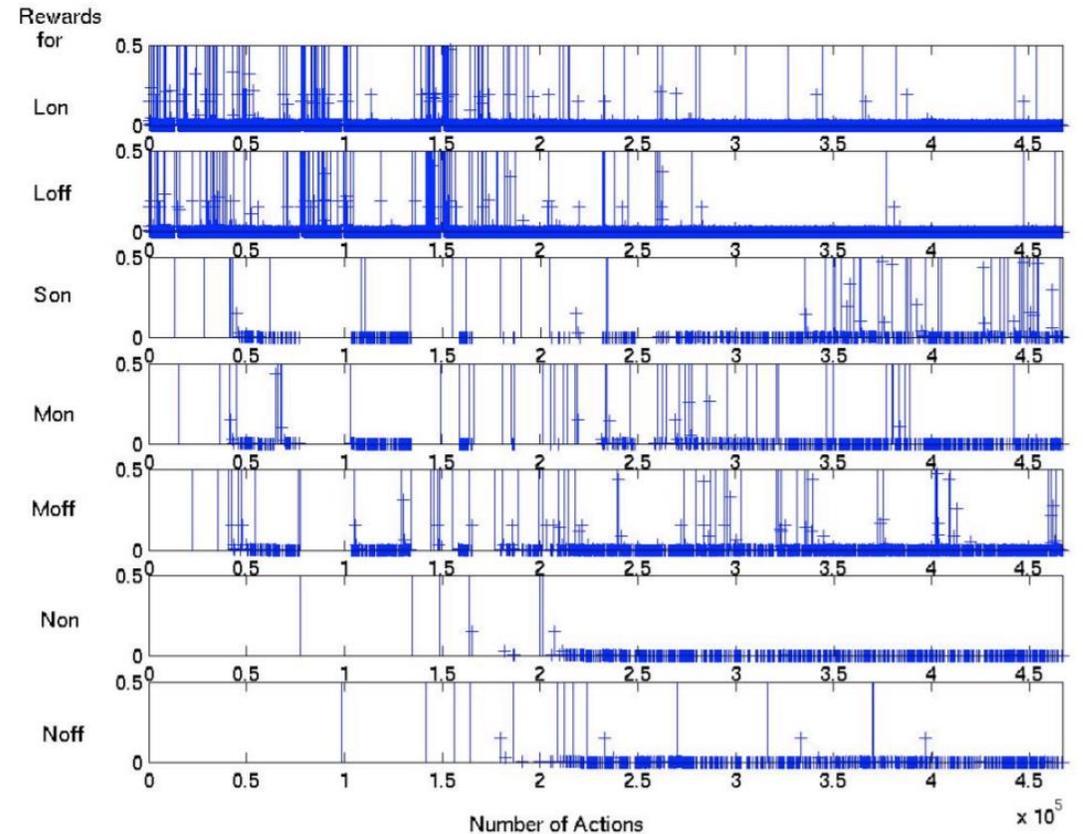
Discussion

- Agent achieves goal much faster when it has both intrinsic and extrinsic motivation

- Future work should draw from psychological, statistical and neuroscience literature to implement other conceptualizations of exploration, interest, and intrinsic motivation