Modeling Sensorimotor Behavior through Modular Inverse Reinforcement Learning with Discount Factors

Ruohan Zhang, Shun Zhang, Matthew H. Tong,
Mary M. Hayhoe, and Dana H. Ballard

The University of Texas at Austin

August 27, 2017

Research Question

- How do humans make decisions in a multitask environment?[1]
- At least two variables matter: reward and planning horizon.
- How do we estimate these variables from behavior data?

[1] Featured image credit: Matt Cheeham (a photo of the pedestrian scramble at Londons Oxford Circus)

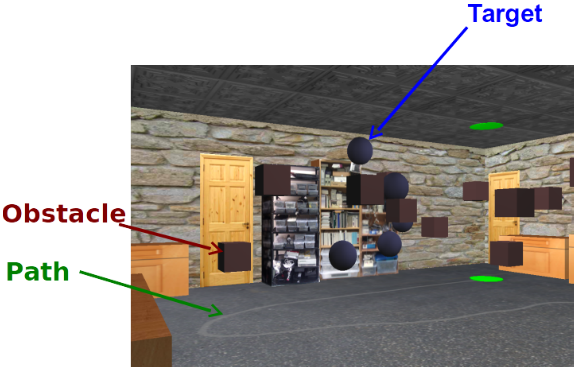
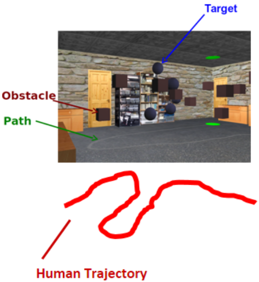


Figure: A subject wears a head mounted display and trackers for eyes, head, and body. Subjects are instructed to do a combination of **following a path, collecting targets, and avoiding obstacles** (designed by Matthew H. Tong).

Research Outline



- Given observed environment states and human actions (data)
  - **Modeling**: hypothesize a decision model
  - **Learning**: estimate the decision variables
  - **Imitation**: reproduce end-to-end behaviors

Modeling: A Reinforcement Learning Framework

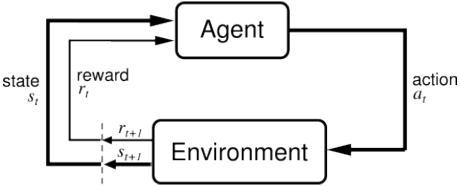


Figure: Agent-Environment Interaction

Modeling: A Reinforcement Learning Framework

In our problem, the observed data $\langle s_t, a_t \rangle$:

- State s_t: distances and angles to obstacles, targets, and the path.
- Action a_t: 16 discrete orientations.

Modeling: A Reinforcement Learning Framework

In our problem, the observed data $\langle s_t, a_t \rangle$:

- State s_t: distances and angles to obstacles, targets, and the path.
- Action a_t: 16 discrete orientations.

The variables to be estimated:

- Reward $\mathcal{R}$: scalar rewards for an obstacle, a target, and the path.
  - Known as the inverse reinforcement learning (IRL) problem [1].

Modeling: A Reinforcement Learning Framework

In our problem, the observed data $\langle s_t, a_t \rangle$:

- State s_t: distances and angles to obstacles, targets, and the path.
- Action a_t: 16 discrete orientations.

The variables to be estimated:

- Reward $\mathcal{R}$: scalar rewards for an obstacle, a target, and the path.
  - Known as the inverse reinforcement learning (IRL) problem [1].
- Discount factor γ: $\gamma \in [0, 1)$,
  - How much a future reward matters compared to the current reward.

Modeling: Modular Reinforcement Learning with γ

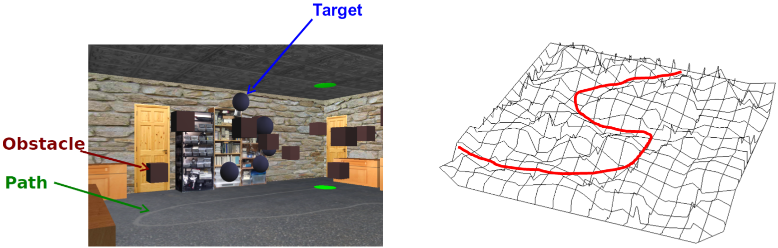
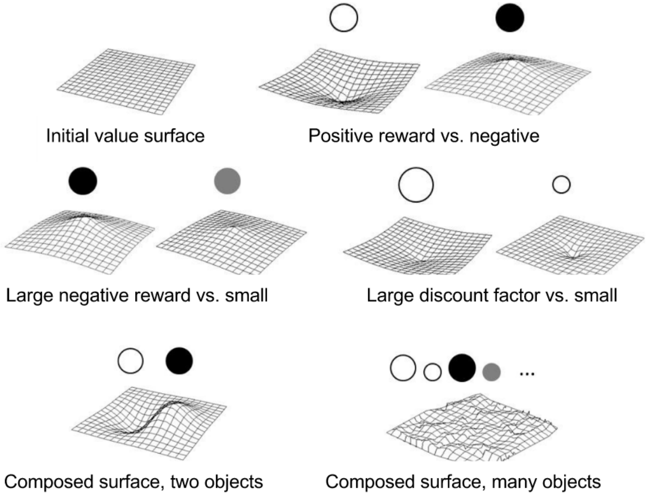
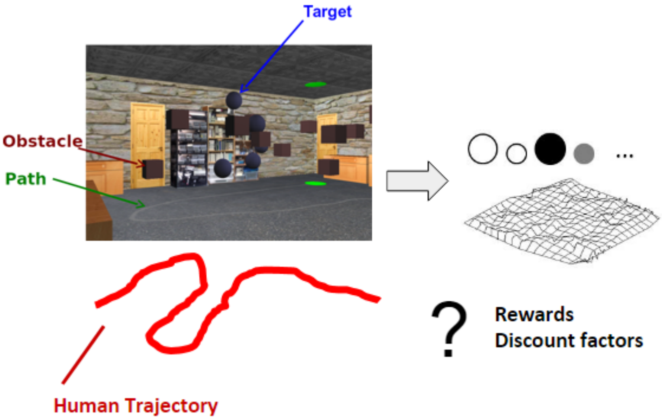


Figure: A human subject chooses action based on her value function, visualized as a value surface.

Modeling: Modular Reinforcement Learning with γ



Learning: Modular Inverse Reinforcement Learning



Learning: Modular Inverse Reinforcement Learning

Maximum Likelihood Inference (Rothkopf and Ballard, 2013)

- Learn rewards and discount factors to maximize the likelihood of observing the actual human actions.

Learning: Modular Inverse Reinforcement Learning

Maximum Likelihood Inference (Rothkopf and Ballard, 2013)

- Learn rewards and discount factors to maximize the likelihood of observing the actual human actions.
- An improved algorithm based on (Rothkopf and Ballard, 2013) [2]
  - Also learns the discount factor.
  - Learns which object to attend to when multiple objects are nearby

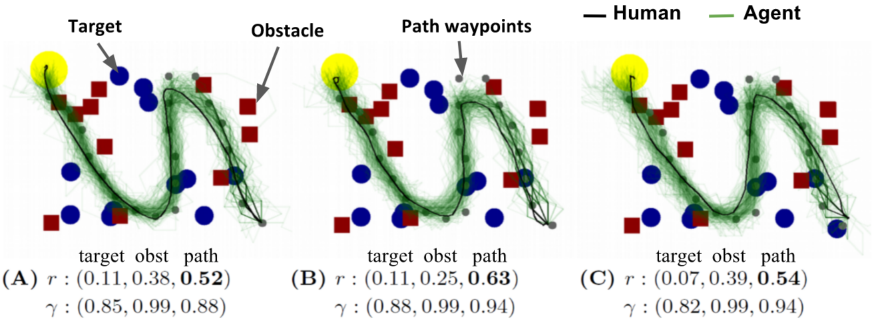
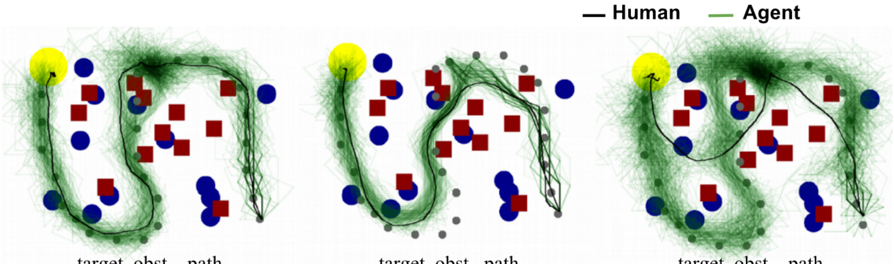


Figure: Top-down view of generated trajectory clouds for 3 subjects performing **Task 1: follow the path only.**

Imitation: Avoiding Obstacles and Following the Path



target obst path
(A) $r : (0.00, \mathbf{0.57}, \mathbf{0.43})$
$\gamma : (0.00, 0.69, 0.91)$

target obst path
(B) $r : (0.02, \mathbf{0.57}, \mathbf{0.41})$
$\gamma : (0.99, 0.59, 0.97)$

target obst path
(C) $r : (0.06, \mathbf{0.75}, \mathbf{0.19})$
$\gamma : (0.95, 0.60, 0.88)$

Figure: Top-down view of generated trajectory clouds for 3 subjects performing **Task 2: ignore targets, avoid obstacles, and follow the path.**

Imitation: Collecting, Avoiding, and Following

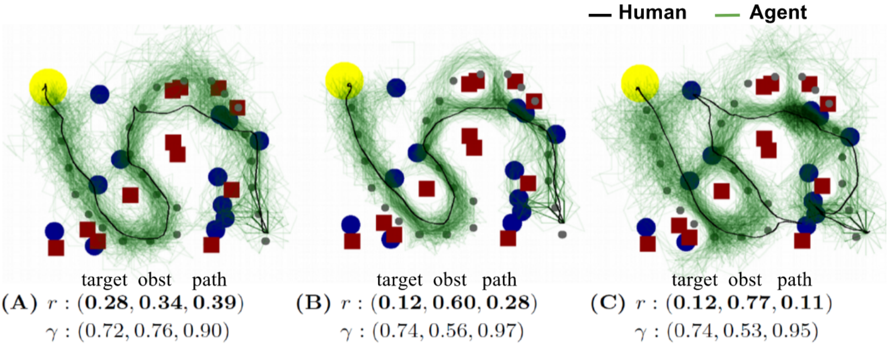


Figure: Top-down view of generated trajectory clouds for 3 subjects performing **Task 4: collect, avoid, and follow together.**

Conclusions

- A variety of multitask navigation behaviors in our experiments can be compactly captured by two decision variables per task: the reward and the discount factor.
- The modular reinforcement learning + modular inverse reinforcement learning approach can be used to reproduce human behaviors.

References

Andrew Y Ng and Stuart J Russell.
Algorithms for inverse reinforcement learning.
In *Proceedings of the Seventeenth International Conference on Machine Learning*, pages 663–670. Morgan Kaufmann Publishers Inc., 2000.

Constantin A Rothkopf and Dana H Ballard.
Modular inverse reinforcement learning for visuomotor behavior.
Biological cybernetics, 107(4):477–490, 2013.

Thank You! Questions?