

Value-Based Policy Teaching with Active Indirect Elicitation

Haoqi Zhang and David Parkes

School of Engineering and Applied Sciences
Harvard University

7.17.08 / AAAI

Many situations arise in which an interested party wishes to affect the behavior of an agent.

Many situations arise in which an interested party wishes to affect the behavior of an agent.

- A teacher wants a student to develop good study habits.
- Parents want their child to come home after school.
- A web 2.0 site wants users to contribute content.
- The government wants people to go green.

Many situations arise in which an interested party wishes to affect the behavior of an agent.

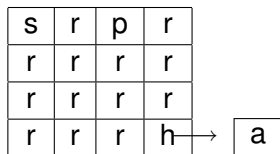
- A teacher wants a student to develop good study habits.
- Parents want their child to come home after school.
- A web 2.0 site wants users to contribute content.
- The government wants people to go green.

Often, the agent does not behave as desired by the interested party.

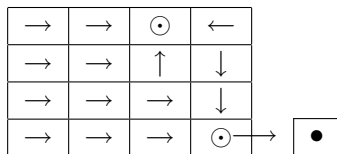
Policy teaching

- Consider agent performing sequential decision task modeled by MDP.
- How to provide limited incentives to induce the agent to follow a policy that is desirable for the interested party?

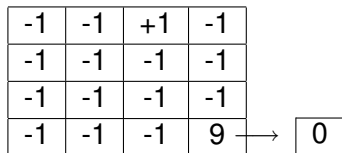
Child walking home from school



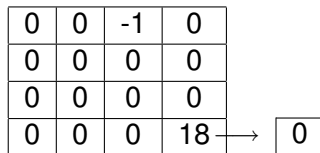
(a) state space



(b) child's policy



(c) child's reward



(d) parent's reward

-1	-1	+1	-1
-1	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	9

(e) child

+

0	0	0	0
0	0	2.07	0
0	0	0.84	0
0	0	0	0.08

(f) motivation

≈

⇒	⇒	⇓	←
→	→	⇓	↓
→	→	⇒	⇓
→	→	→	⊙

(g) induced policy

Problem setup and assumptions

- Consider repeated interaction with agent.
- Interested party knows agent's model except for the reward function.
- Agent policy is observable.
- Can provide incentives, but cannot otherwise impose actions.
- Agent is a planner, but not forward looking.

Understanding preferences

- Direct preference elicitation is costly and intrusive.
- Passive indirect elicitation (revealed preference) is insufficient.
- Look for active, indirect approach.

Our contributions: (1) New paradigm

- First steps to get this to work, so please bare with me.

Our contributions: (1) New paradigm

- First steps to get this to work, so please bare with me.
- Some loosely related work:
 - ▶ From AI: Inverse Reinforcement Learning (Ng and Russell, 2000)
 - ▶ From microeconomics: Principal-Agency and Contract Theory (Bolton and Dewatripont, 2005)

Our contributions: (2) Value-based policy teaching

A step beyond: how to provide limited incentives to induce a policy that maximizes value to the interested party?

- Best inducible policy is not known a priori.
- Affect agent's reward to induce agent policy that is then valued by the interested party.

Our contributions: (3) Active, indirect elicitation

- Idea: provide incentives, observe, repeat.

Value-Based Policy Teaching

- Agent plans with respect to $M = \{S, A, R, P, \gamma\}$.

Value-Based Policy Teaching

- Agent plans with respect to $M = \{S, A, R, P, \gamma\}$.
- Interested party judges the agent's policy w.r.t. $T = \{S, A, \mathbf{G}, P, \gamma\}$.

Value-Based Policy Teaching

- Agent plans with respect to $M = \{S, A, R, P, \gamma\}$.
- Interested party judges the agent's policy w.r.t. $T = \{S, A, \mathbf{G}, P, \gamma\}$.
- Interested party provides **admissible** incentive $\Delta : S \rightarrow \mathbb{R}$.
(e.g., limited budget)

Value-Based Policy Teaching

- Agent plans with respect to $M = \{S, A, R, P, \gamma\}$.
- Interested party judges the agent's policy w.r.t. $T = \{S, A, \mathbf{G}, P, \gamma\}$.
- Interested party provides **admissible** incentive $\Delta : S \rightarrow \mathbb{R}$.
(e.g., limited budget)
- Agent performs π' w.r.t. $R + \Delta$.

Value-Based Policy Teaching

- Agent plans with respect to $M = \{S, A, R, P, \gamma\}$.
- Interested party judges the agent's policy w.r.t. $T = \{S, A, \mathbf{G}, P, \gamma\}$.
- Interested party provides **admissible** incentive $\Delta : S \rightarrow \mathbb{R}$.
(e.g., limited budget)
- Agent performs π' w.r.t. $R + \Delta$.
- **Goal**: provide Δ that induces a policy that maximizes the value to the interested party, subject to admissibility constraints.

Known agent reward

- Optimization problem: We provide a mixed integer program formulation.
- Problem is NP-hard (reduction from KNAPSACK).

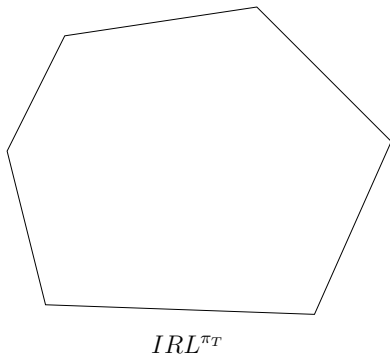
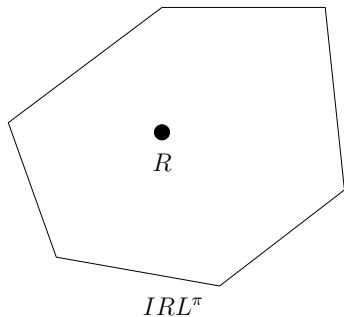
Unknown agent reward

- How to locate agent's reward?
- Inverse Reinforcement Learning (IRL)
 - ▶ Space of rewards consistent with observed policy is characterized by linear constraints. (Ng and Russell, 2000)

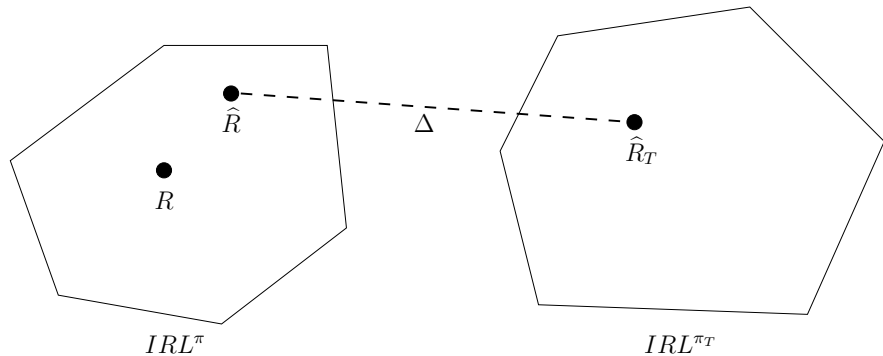
Two problems to tackle

- How to induce a particular target policy?
- What is the best inducible policy with respect to the agent's *unknown* reward? (when are we done?)

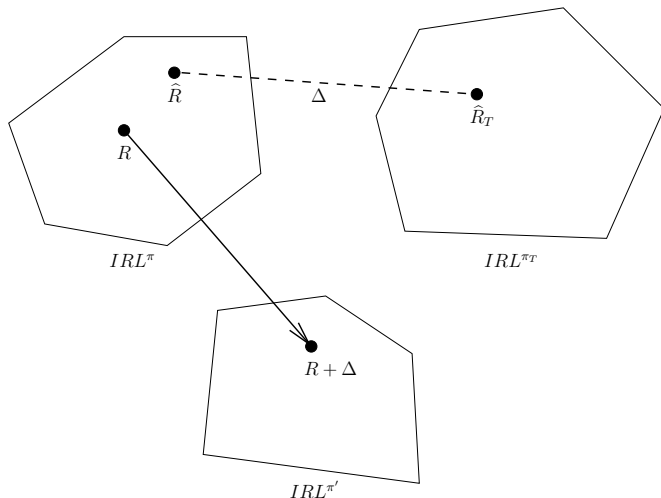
Active, indirect elicitation (fixed target policy)



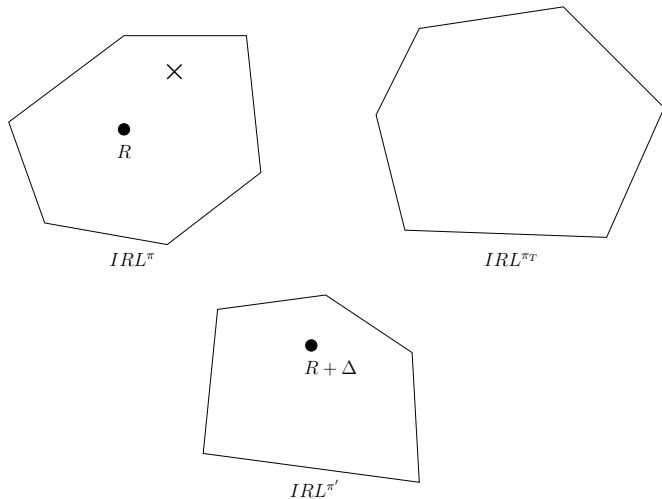
Active, indirect elicitation (fixed target policy)



Active, indirect elicitation (fixed target policy)



Active, indirect elicitation (fixed target policy)



Maximization w.r.t. unknown agent reward

- Keep track of value of best induced policy so far.
- Only consider rewards that may reach policies with higher value to interested party.

Convergence

Theorem

The elicitation method terminates in finite steps with an admissible mapping Δ to a policy that maximizes the value of the interested party (with respect to the true agent reward).

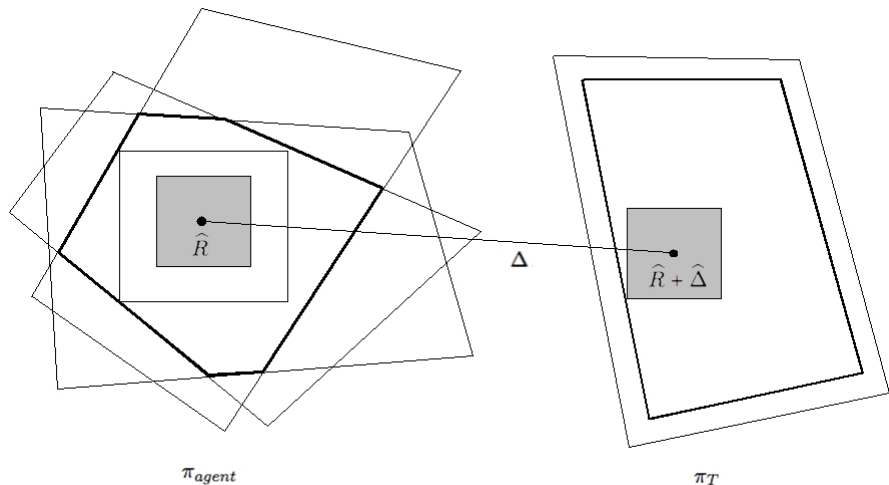
Intuition:

Pigeonhole argument on number of hypercubes that can fit in the polyhedron □

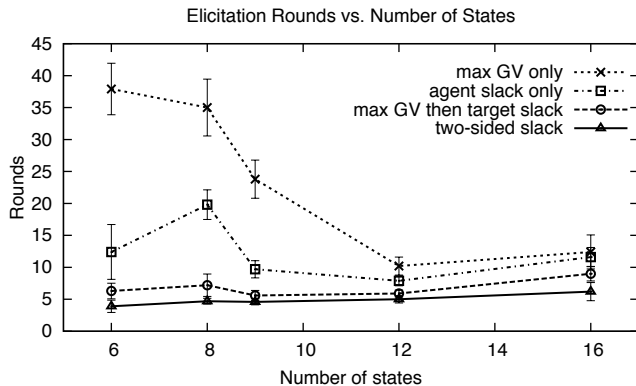
Choosing an objective function

- Tractable
- Few elicitation rounds

Two-sided slack maximization heuristic



Simulations



Future work

Addressing limitations:

- Handling strategic agents
- Relaxing fully observable
- Tighter formulations

Extending out:

- Multiple agents, multiple interested parties
- Alternate design levers
- Learning agents

If I only had one slide, I'd say this:

- Policy Teaching is a new paradigm for studying the relations among preferences, behavior, and environment.
- We introduce a general active, indirect elicitation method.
- Plenty of exciting open questions!

Thank you

Please send comments and suggestions to hq@eecs.harvard.edu.