
Efficient Model Learning for Dialog Management

Finale Doshi
Nick Roy

Outline

- Motivation and Background
- Our Approach
 - Working with uncertain model parameters
 - Heuristics for learning
- Results
- Future Work

Motivation

Dialog management
allows for natural
human-robot interaction...
...but there are several
challenges:

- noisy speech recognition
- linguistic ambiguities

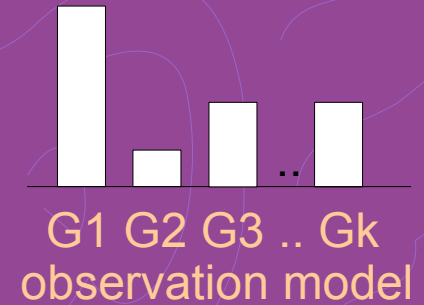
Going to the
cafeteria...

Let's go to
the elevator

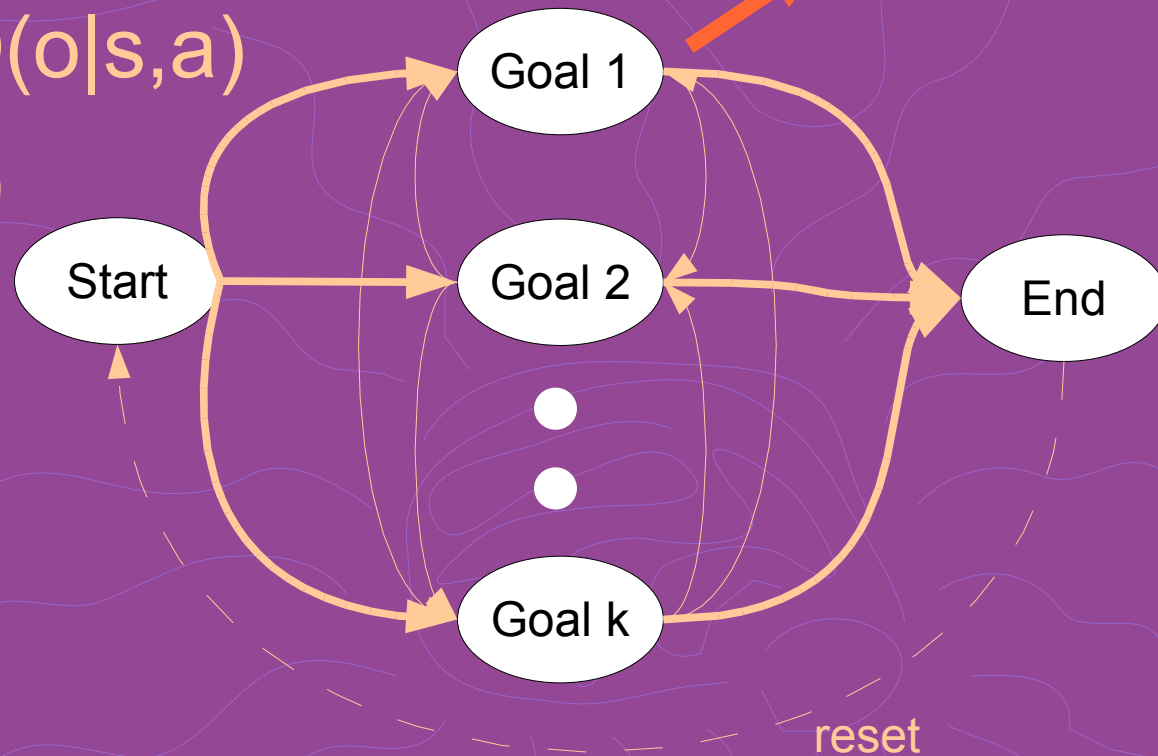


Partially Observable Markov Decision Processes (POMDPs) manage uncertainty

- Set of states (S), actions (A), and observations (O)
- Transition Model $T(s'|s,a)$
- Observation Model $O(o|s,a)$
- Reward Model $R(s,a)$



Large number of parameters are difficult to specify a priori!



Solving the POMDP:

Value of a belief

Value of belief, action pair

$$V(b) = \max_{a \in A} Q(b, a),$$

$$Q(b, a) = R(b, a) + \gamma \sum_{b' \in B} T(b' | b, a) V(b'),$$

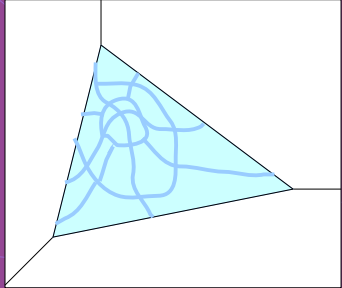
$$Q(b, a) = R(b, a) + \gamma \sum_{o \in O} \Omega(o | b, a) V(b_a^o),$$

Current reward

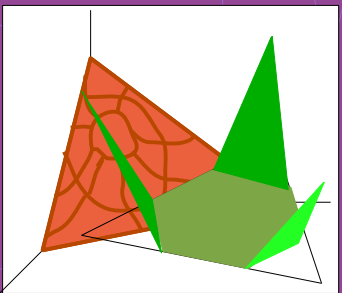
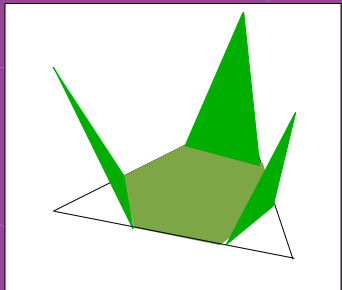
Future Reward

We can apply these equations to solve for $V(b)$ recursively.

Our Approach



- Place priors over parameters
 - “Expert” provides estimate and confidence
 - We convert to Dirichlet/Gaussian prior
- Find a policy that maximizes reward given the uncertainty in the parameters
- After each completed interaction
 - Update priors on the parameters
 - Update POMDP solution with additional backups



Our Approach

If parameters are uncertain, solve POMDP with the expected value of the parameters to optimize reward.

Expectation over states

$$Q(b, a) = \max_i q_a \cdot b,$$

$$q_a(s) = \mathbf{E}[R(s, a)] + \gamma \sum_{o \in O} \sum_{s' \in S} \mathbf{E}[T(s' | s, a) \Omega(o | s', a)] V_{n-1, i}(s)$$

Expectations over model parameters

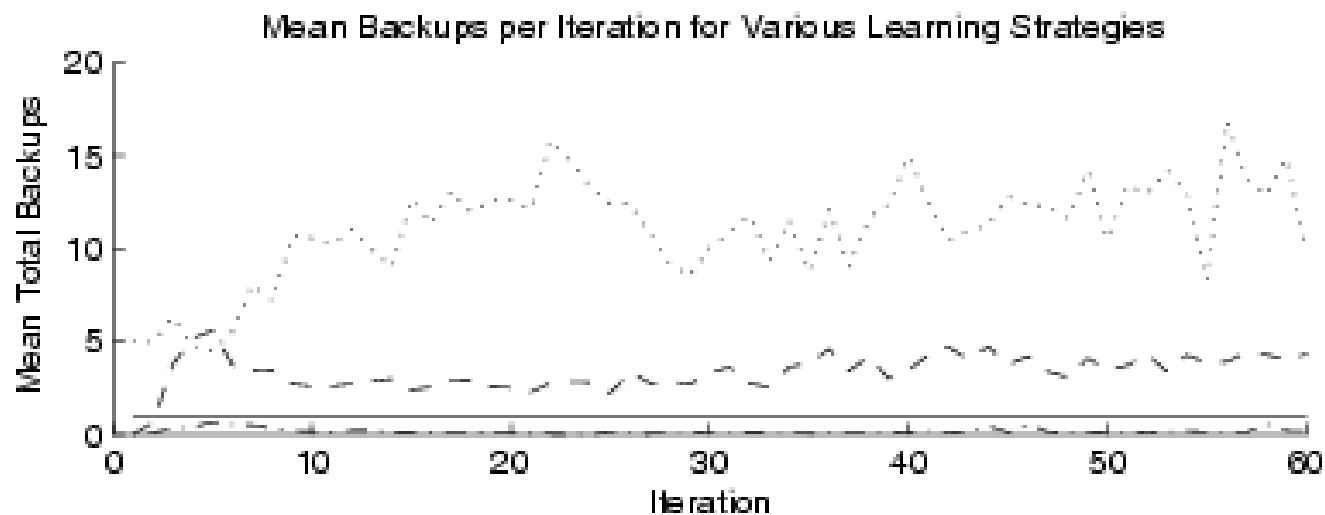
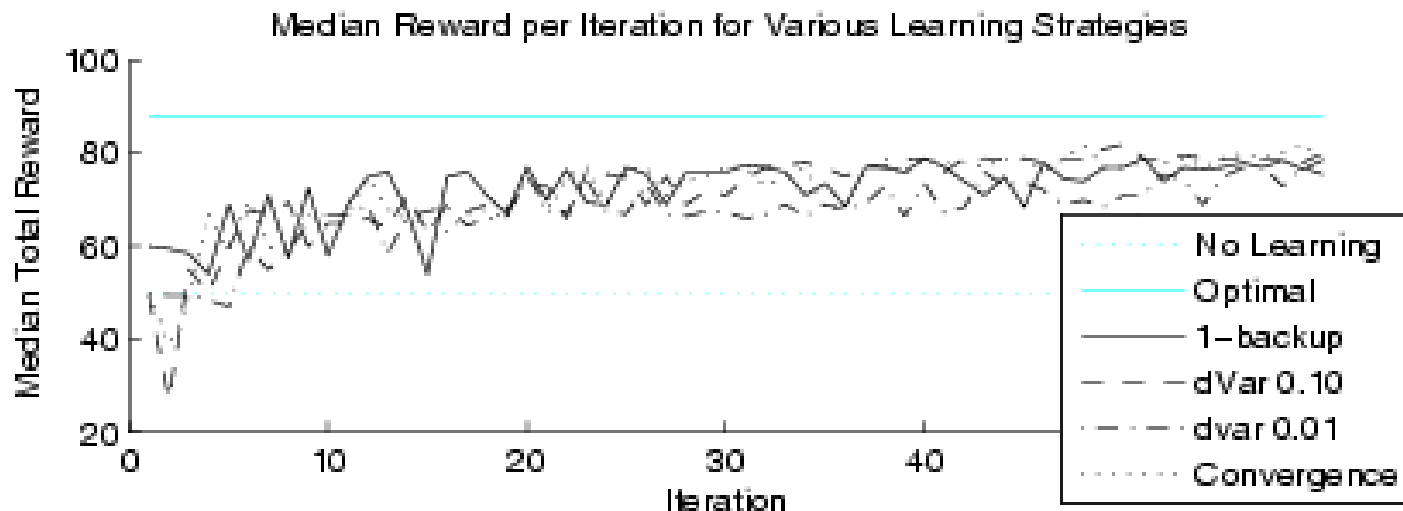
Expectation over model stochasticity

Our Approach

Update the policy after each dialog:

- How? Compute additional backups
 - POMDP solution will converge given new expected values for the parameters.
- How many backups?
 - Backup to convergence
 - Backup a fixed number of times
 - Backup **proportionally to variance reduction** in the parameters.

Simulation Results



Wheelchair Results

Non-learner

User: Take me to the elevator.

Robot: Where did you want to go?

User: The Gates elevator please.

Robot: Do you want to go to the Gates Elevator?

User: Yes.

Robot: Going to Gates.

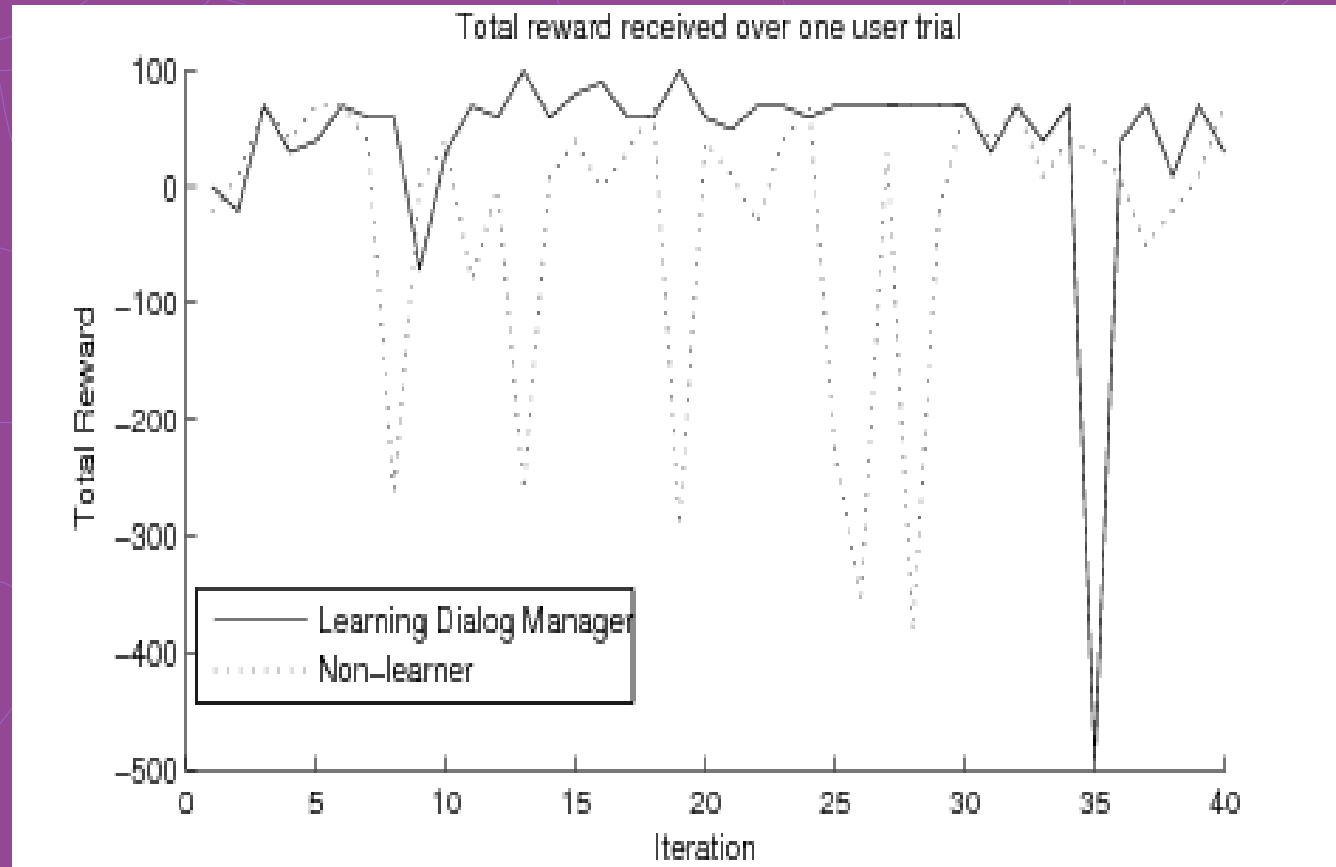
Learner

User: Take me to the elevator.

Robot: Do you want to go to the Gates Tower?

User: Yes

Robot: Going to Gates



Future work: Improve the Heuristic

- Three sources of variation in solutions
 - Stochasticity inherent in model
 - Uncertainty in parameters
 - **Incomplete solution convergence**
- Refine solution until variation due to incomplete convergence is not significant compared to other sources.

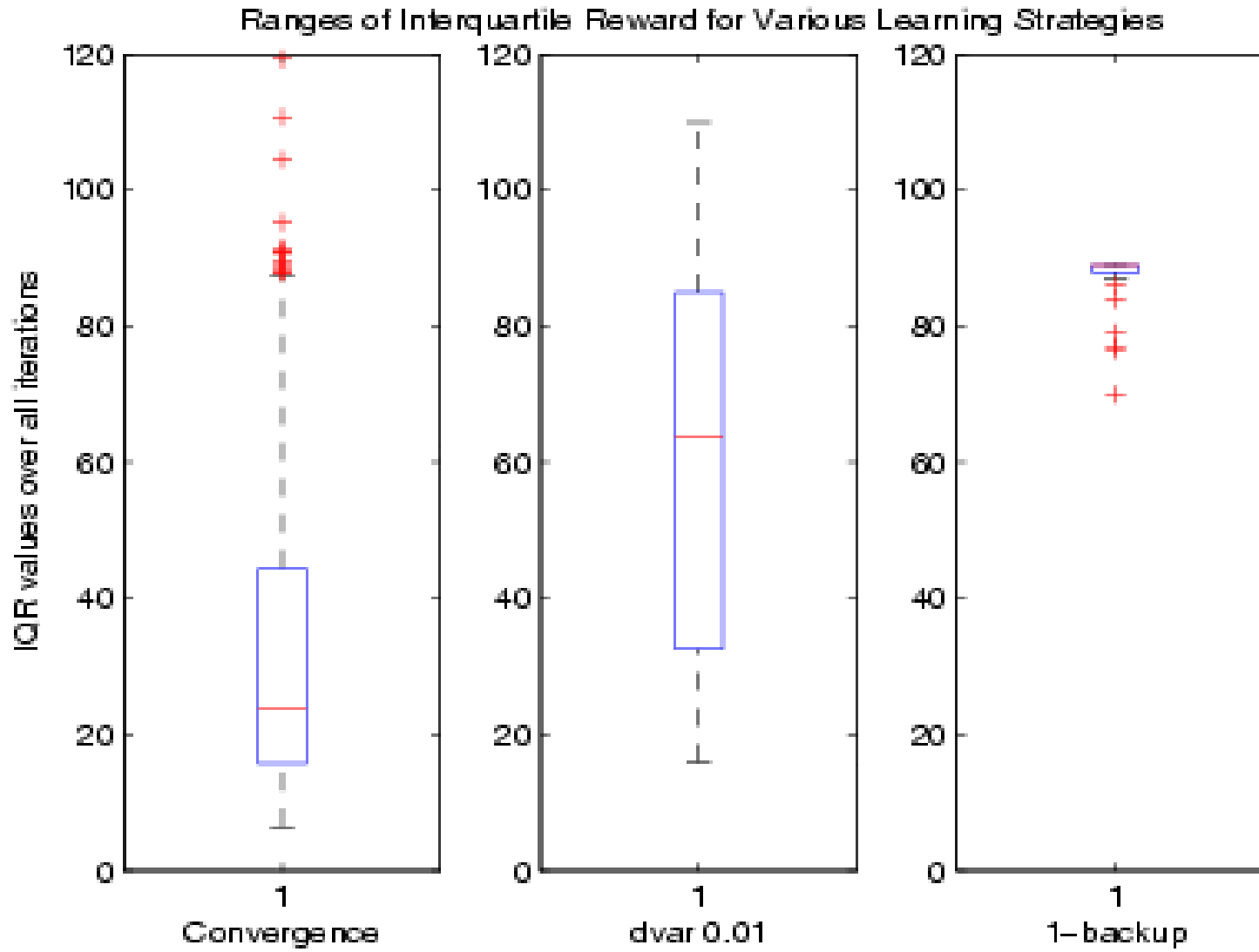
Future Work: Larger Questions

- How can we guarantee convergence of the important parameters?
- How do we balance the exploration-exploitation trade off? (the POMDP is unaware of parameter uncertainty)
 - Should take exploratory actions *and*
 - Should act robustly



Thank-you

Simulation Results



Estimating Problem Uncertainty

$$\text{var}(V(b)) = f(b) + \sum_{a \in \mathcal{O}} \eta_a(b) \text{var}(V(b_a^*))$$

