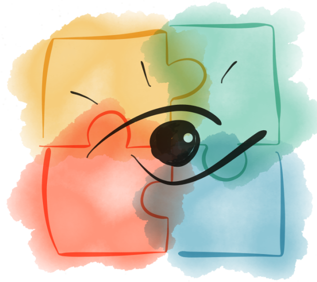
Georgia Tech

CS 7643

Project ID: 11

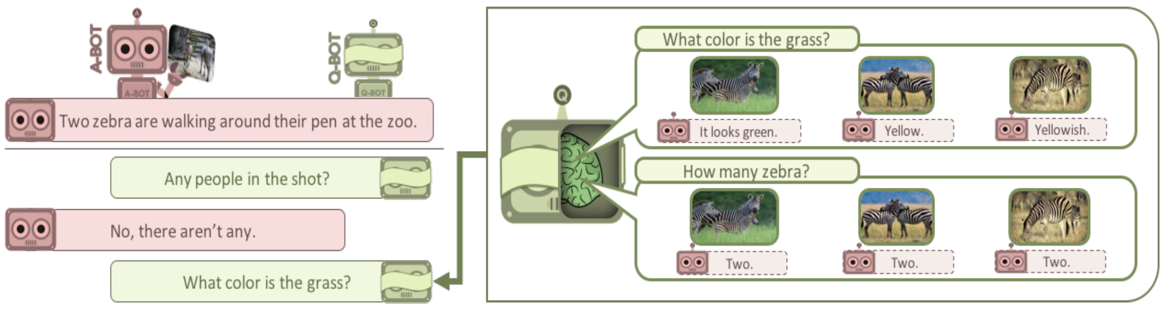
Inner Dialog: Pragmatic Visual Dialog Agents that Rollout a Mental Model of their Interlocutors

Viraj Prabhu



1 Overview

- **Motivation:** "To plan ahead we must simulate the world"
- **GuessWhich:** Cooperative image guessing game (Das & Kottur et al. 2017)
- **Goal**: Select **"pragmatic"** questions at **inference**, via **dialog rollouts** on a **mental model** of teammate (ABot)



2 Approach

- **Modeling AMental:**
  - ACopy: ABot replica (performance upper bound)
  - AMimic: Same architecture, trained on ABot samples
- **Dialog Rollouts:**
  - Reward Estimation via finite-sample approximation
    - Minimize **Bayes Risk** under QBot's beliefs
    - Sample candidate questions, images, answers
    - **Marginalize over beliefs** using AMental

Optimization

Pick question with max expected reward

$$\left\{ q_t^* = \arg\max_{q_t} \underbrace{\mathop{\mathbb{E}}_{\tilde{I},\tilde{a}_t}\left[r_t\left(\hat{I}^t_{|q_t,\tilde{a}_t}, \tilde{I}\right)\right]}_{\text{Estimated reward}} \right.$$

$$\mathop{\mathbb{E}}_{\tilde{I},\tilde{a}_t}\left[r_t\left(\hat{I}^t_{|q_t,\tilde{a}_t}, \tilde{I}\right)\right] \approx \underbrace{\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} r_t\left(\hat{I}^t_{|q_t,a_t^{m,n}}\ ,\ I_m\right)}_{\text{Marginalize over beliefs}}$$

Algorithm 1 Selecting Pragmatic Questions Via Dialog Rollouts

function Pragmatic-Question-Selection(q_t)
  $q_t^1, \ldots, q_t^K \sim \pi_Q(q_t \mid q_0, a_0, \ldots, q_{t-1}, a_{t-1})$ ▷ Decode multiple likely questions
  return argmax RolloutEstimate(q_t^i) ▷ Select q_t with greatest expected reward
end function

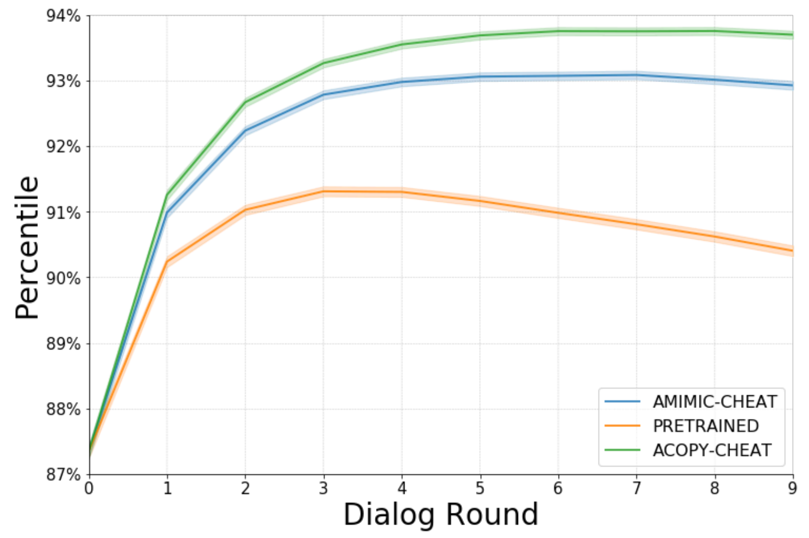
function RolloutEstimate(q_t)
  $\tilde{r} \leftarrow 0$
  $I_1, \ldots, I_M \sim G_{t-1}(\hat{I}^{t-1} | q_0, a_0, \ldots, q_{t-1}, a_{t-1})$ ▷ Sample likely source images
  for m $\in$ {1, …, M} **do**
    for n $\in$ {1, …, N} **do**
      $a_t^{m,n} \sim \pi_M(a_t \mid I_m, q_0, a_0, \ldots, q_t)$ ▷ Sample answer given q_t and I_m
      $\hat{I}^t_{|q_t,a_t^{m,n}} \leftarrow \text{argmax } G_t(\hat{I}^t | q_0, a_0, \ldots, q_t, a_t^{m,n})$ ▷ Update Q-Bot's prediction
      $\tilde{r} \leftarrow \tilde{r} + r_t\left(\hat{I}^t_{|q_t,a_t^{m,n}}, I_m\right)$ ▷ Aggregate the reward
    end for
  end for
  return $\tilde{r}/MN$ ▷ Return approximate expected reward
end function

3 Preliminary Results

Metric: Rank percentile over dialog rounds

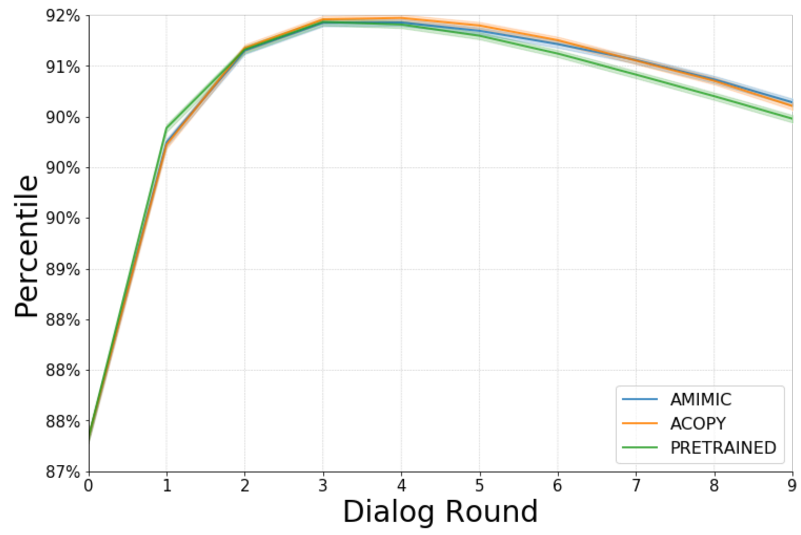
Baseline: Pretrained (no mental modeling)

Cheat Setting: "Cheat" with GT target image



Mental modeling can help ..

Real World: Expectation under QBot's beliefs



.. but real-world gains are not observed yet

Challenges

- Estimating rewards is a bottleneck
- Scaling up approximation is expensive

Takeaways

- Pragmatic inference can in theory provide an alternative to fine-tuning with RL
- But in presence of information assymetry, accurately estimating reward in hard

4 References

Das, A., Kottur, S., Moura, J.M., Lee, S. and Batra, D., (2017). Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning. ICCV 2017

Das, A., Kottur, S., Gupta, K., Singh, A., Yadav, D., Moura, J., Parikh, D., and Batra, D., Visual dialog. CVPR 2017.