

Word Games

J Chiu

March 16, 2022

Dialogue and information gathering

- ▶ Resolve ambiguity and coordinate through dialogue
- ▶ OneCommon: Interactive, symmetric reference game
 - ▶ Isolates info gathering (and coordination)
 - ▶ Environment (dots) are completely static
 - ▶ Dynamism comes from dialogue only
- ▶ 20 questions with symmetric information constraints

Previous SotA

- ▶ Purely supervised
 - ▶ Upper-bounded by performance of demonstrators
- ▶ Uncalibrated beliefs: overconfidence
 - ▶ Pushes for to select a dot that will not work
- ▶ Research goal: Improve supervised models via model-based planning

Fixing strategy with planning

- ▶ Prior: Fully supervised neural encoder-decoder
 - ▶ Encode past interactions with a neural net
 - ▶ Generate what to say with a neural net
 - ▶ Brittle strategy, less brittle language
- ▶ Next: Model-based planning
 - ▶ Choose what to say by imagining how partner would respond
 - ▶ Say utterance with best expected outcome
 - ▶ Potentially stronger player than expert demonstrations

Challenges in model-based planning

- ▶ Partner modeling is hard
 - ▶ Variable amount of information in utterances
 - ▶ Long tail of information
 - ▶ Empirical question: what is model capacity used for? Is most of the text used for communicated long tail phenomena? Is the long tail not directly related to what we say?
- ▶ Multi-turn planning
 - ▶ Accuracy of planning depends greatly on the partner model
 - ▶ Errors from the partner model will compound over time¹
- ▶ Single-turn planning
 - ▶ Removes compounding errors
 - ▶ Must optimize a dialogue progress heuristic: reduce uncertainty
 - ▶ Requires belief with uncertainty
 - ▶ Still requires accurate partner model²
- ▶ Research question: Can we improve partner modeling for planning by simplifying partner responses?

¹Errors in planning will be a result of compounding partner model errors on top of search error.

²Belief is function of dialogue history and partner model.

Planning

- ▶ Partner response depends on utterance and conversation history
 - ▶ First history $h_0 = \text{dots you can see}$
 - ▶ History h_t , utterance u , response r
 - ▶ Next history $h_{t+1} = (h_t, u, r)$



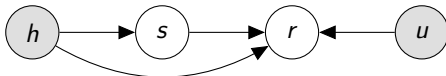
- ▶ Plan by imagining partner response r

$$\min_u \mathbb{E}_{p(r|h,u)} [\text{Cost}(h, u, r)]$$

- ▶ Produce utterances from prior model and rerank
- ▶ Cost should approximate dialogue progress
 - ▶ Goal of dialogue is information gathering and coordination
 - ▶ Focus on information gathering
- ▶ Cost should be a function of belief

Planning with Belief

- ▶ Introduce belief $p(s | h)$
 - ▶ State s is what dots partner can also see
 - ▶ Response model $p(r | h, u, s)$ now has more conditioning



- ▶ Incorporate belief in planning

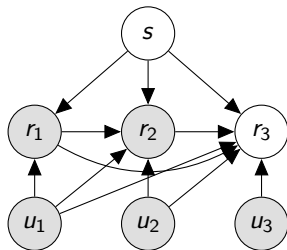
$$\min_u \mathbb{E}_{p(r|h,u,s)p(s|h)} [\text{Uncertainty}(p(s' | h, u, r))]$$

- ▶ Obtain belief posterior via belief update

$$p(s | h, u, r) = \frac{p(r | h, u, s)p(s | h)}{\sum_s p(r | h, u, s)p(s | h)}$$

Partner response model

- ▶ Model their response from their perspective
- ▶ Static latent state s : which dots do they also see
 - ▶ Pick s that is observed during training so we can keep all models supervised
 - ▶ Alternative: actual field of view
 - ▶ Hard to model unshared dots, pick shared dots as compromise
- ▶ Uniform prior $p(s)$ over 7 choose 4 dots partner also sees
- ▶ Partner response model $p(r_t \mid u_{1:t}, r_{1:t-1}, s)$
 - ▶ History $h_t = (u_{1:t-1}, r_{1:t-1})$



Response model: Informativity

- ▶ Example exchange
 - ▶ Action: Do you see a red dot?
 - ▶ Observation: No, but I see a blue one.
- ▶ Utterances are multifaceted
 - ▶ Responses contain more information than asked
 - ▶ New information injected by partner, not constrained
 - ▶ Difficult to model new information
- ▶ Introduce informativity as another variable
- ▶ Make worst-case assumption about informativity **during planning**
 - ▶ Assume a limit on number of bits transmitted
 - ▶ Conservative lower bound on uncertainty reduction / dialogue progress

Incorporating informativity

- ▶ Add partner-level informativity i to response model $p(r \mid h, u, s, i)$
 - ▶ Determines partner willingness to give more information
 - ▶ More information improves game-play
- ▶ Introduce uncertainty set over $i \in [\text{low}, \text{high}]$

$$\min_u \min_{i \in [\text{low}, \text{high}]} \mathbb{E}_{p(r|h, u, s, i) p(s|h)} [\text{Uncertainty}(p(s \mid h, u, r))]$$

- ▶ Optimize worst-case informativity $i = \text{low}$
- ▶ Constrain informativity in response via discrete coding

Limiting informativity with discrete coding

Hypothesis: Limiting informativity results in both more accurate models and conservative policies

- ▶ Compress response into discrete code $z \in [K]$
- ▶ Use variational information bottleneck (VIB)³
- ▶ Optimize mutual information objective⁴
 - ▶ Focus code z on response to new information introduced by utterance u
 - ▶ But limit the information shared with response r and history h

$$\max_{\theta} I(z, u; \theta) - \beta I(z, r; \theta) - \gamma I(z, h; \theta)$$

- ▶ Lagrange multipliers β, γ
- ▶ Plan with response model $p(z \mid h, u, s)$ instead of r

³Alemi et al. (2016)

⁴Closely related to the approach of Li and Eisner (2019)

Belief update

- ▶ Performing the belief update with z instead of r would lose information
- ▶ Use original model $p(r \mid h, u, s)$ to update belief?
 - ▶ May run into original possible issue of poor calibration
- ▶ Learn another discrete representation of r that does not aim to disentangle
 - ▶ Need as baseline for model with z
 - ▶ Likely adequate when combined with mentions

Summary

- ▶ Goal: Show (partner) model-based planning works for dialogue with purely supervised components
 - ▶ Single-turn planning for limiting compounding model errors
 - ▶ Belief-based heuristic for measuring dialogue progress
 - ▶ Response coding for conservative planning
- ▶ Hypothesis: Limiting informativity results in both more accurate models and more conservative policies

Belief unit tests

Belief calibration: check belief dynamics on static conversations.

- ▶ Diminishing returns
 - ▶ Ask the same question twice (rephrased) should not change belief
 - ▶ Asking about the same dot should have diminishing uncertainty reduction
- ▶ Updates are conservative
 - ▶ Require multiple positive answers to diff questions before being certain about a dot
- ▶ High probability after confirming all neighbouring dots

Extrinsic evaluation: Selfplay

Compare to prior work (without belief)

- ▶ Success rate
 - ▶ Should be higher
- ▶ Efficiency (success / num turns)
 - ▶ Possibly higher because more success, but more turns
- ▶ Number of repeated mentions
 - ▶ Should be higher if policy more conservative

Questions

1. How do we measure belief calibration?
 - ▶ Unit tests examining conservativity of beliefs
2. How calibrated are the belief updates using various response representations?
 - ▶ Too much information (ie full word response) results in low probability to responses, which results in a large belief update = optimism
 - ▶ Not enough information = conservative
3. How well can we produce a range of utterances to search over?

Hypothesis (Info hypothesis)

- ▶ High information response representation
 - ▶ Calibrated policy (assuming accurate model)
 - ▶ Hard to model responses
- ▶ Low information response representation
 - ▶ Conservative policy (assuming accurate model)
 - ▶ Easy to model responses
- ▶ Prefer conservative + accurate response model over inaccurate response model

Response representations

- ▶ Words in response
 - ▶ Full response
 - ▶ First k words of response
- ▶ Dot mentions
 - ▶ All dot groups mentioned in response
 - ▶ First k dot groups mentioned in response
- ▶ Discrete encoding
 - ▶ K-means cluster of sentence rep
 - ▶ Specialized cluster (information bottleneck)
- ▶ Continuous encoding
 - ▶ Sentence rep
 - ▶ Specialized encoding (information bottleneck)

Experiments

- ▶ Evaluate response representations on unit tests and selfplay
- ▶ Binary matrix of representation x unit test property
- ▶ Main unit tests
 - ▶ Diminishing returns
 - ▶ Conservativity
 - ▶ High probability
- ▶ Representations
 - ▶ Words
 - ▶ Dots
 - ▶ Learned (discrete, continuous)
- ▶ Consider highest and lowest information representations, work our way in
 - ▶ Hypothesize that best point is in between first dot and all dots
- ▶ Evaluate selfplay in parallel

Experiment 1

Evaluate word response models on belief unit tests

- ▶ Belief representations
 - ▶ Full response
 - ▶ First k words
- ▶ Hypothesis
 - ▶ Full response will fail all unit tests
 - ▶ First k will fail high probability test. Too conservative due to missing information
- ▶ Outcomes
 - ▶ Full is too optimistic, motivating different representations for exploring more conservative belief updates
 - ▶ Full is conservative, which is a success. Method is general, just need to try it on other datasets
 - ▶ First k is too optimistic. This would indicate information hypothesis is wrong, and potential issues with any other subsequent representation approach.

Experiment 2

Evaluate dot response models on belief unit tests

- ▶ Belief representations
 - ▶ All dot mentions
 - ▶ First k dot mentions
- ▶ Hypothesis
 - ▶ These might do okay on unit tests
 - ▶ First $k = 1$ often captures responses
- ▶ Outcomes
 - ▶ First k fails due to optimism, implies info hypothesis is wrong.
 - ▶ First k succeed, which gives an indication that a learned structured generalization has promise. If all also succeeds, same.

Experiment 3

Evaluate learned discrete response models on belief unit tests

- ▶ We can tune the amount of information via number of clusters
- ▶ If training goes well, should be able to get a fine-grained view of information - conservativity tradeoff
- ▶ Belief representations
 - ▶ K-means cluster on pretrained sentence rep
 - ▶ Specialized cluster⁵
- ▶ Hypothesis
 - ▶ K-means might be pretty competitive, depending on sentence representations
 - ▶ K-means should be conservative, since untuned sentence reps may not pick up relevant information
 - ▶ Specialized clusters should do better than k-means with naive sentence rep
- ▶ Outcomes
 - ▶ If any of these succeed, selfplay and try on another dataset
 - ▶ If fail, end of line

⁵Li and Eisner (2019)

End

Full planning details

- ▶ Given history h , we need to choose an action a by optimizing heuristic utility

$$\min_a C(h, a)$$

- ▶ Cost $c = -\text{information gain} + \text{utterance} + \text{pragmatic cost}$
 - ▶ IG: Reduce uncertainty
 - ▶ Utterance cost: Can't send a full paragraph
 - ▶ Pragmatic cost: Want utterance to be accurate
- ▶ Ideally would estimate and optimize future reward directly
 - ▶ Heuristic approximation of future reward U
 - ▶ Limited-horizon planning to minimize impact of model error

Expected information gain

- ▶ Maximizing expected information gain equivalent to minimizing uncertainty

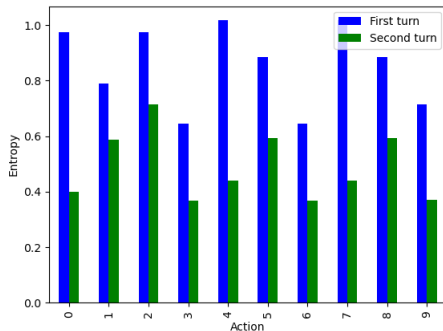
$$\min_u \sum_r \sum_s \underbrace{p(r | h, u, s)}_{\text{response model}} \underbrace{p(s | h)}_{\text{belief}} \text{Uncertainty}(\underbrace{p(s | h, u, r)}_{\text{new belief}})$$

Experiments

- ▶ Mutual Friends
 - ▶ Augment rule-based (prior work) to optimize info gain
 - ▶ After OneCommon: Add neural on top
- ▶ OneCommon
 - ▶ Use attributes = raw mention configurations
 - ▶ Need belief / info gain / LR weights
 - ▶ How to deal with redundancy? (i.e. correlation between features)
 - ▶ Learn latent refinement on top of mention configurations

Information gain issues

- ▶ Best info gain could be to ask the same question twice
- ▶ Usual fix: Limit to asking once only
- ▶ Would be nice to have a principled way to deal with correlated features though



- ▶ Second turn after taking action with lowest entropy

Related work: 20 questions

- ▶ Padmakumar and Mooney (2020)
 - ▶ Attribute-based classification (string heuristic to map to description) + activate learning about attributes
 - ▶ Info gain (on top of binary unweighted logistic regression) as feature for RL policy
- ▶ Yu et al. (2019)
 - ▶ Question-based classification (attributes)
 - ▶ Learn weights of features
 - ▶ Do not consider feature correlations
- ▶ More interesting language, symmetric setting
- ▶ Learn weights, account for correlation
- ▶ Symmetry, deal with unexpected features

End

Concerns

- ▶ Would a large LM solve all of this?
 - ▶ Fine tune on small onecommon dataset, are there still repeats?
 - ▶ Unlikely to solve strategy / over optimism

Expected Information Gain

$$IG(h, a) = H(i | h) - \mathbb{E}_{p(o|h,a)} [H(i | h, a, o)]$$
$$\mathbb{E}_{p(o|h,a)} [H(i | h, a)] = \sum_o \sum_{i'} p(o | h, a, i) p(i | h) H(i | h, a, o)$$

- ▶ Equivalent to minimizing expected uncertainty after receiving a response
- ▶ Cite Yu et al, White et al

Citations I

- Alemi, A. A., Fischer, I., Dillon, J. V., and Murphy, K. (2016). Deep variational information bottleneck. *CoRR*, abs/1612.00410.
- Li, X. L. and Eisner, J. (2019). Specializing word embeddings (for parsing) by information bottleneck. *CoRR*, abs/1910.00163.
- Padmakumar, A. and Mooney, R. J. (2020). Dialog policy learning for joint clarification and active learning queries. *CoRR*, abs/2006.05456.
- Yu, L., Chen, H., Wang, S. I., Artzi, Y., and Lei, T. (2019). Interactive classification by asking informative questions. *CoRR*, abs/1911.03598.

Games

Friends of agent A:

Name	School	Major	Company
Jessica	Columbia	Computer Science	Google
Josh	Columbia	Linguistics	Google
...

A: Hi! Most of my friends work for Google

B: do you have anyone who went to columbia?

A: *Hello?*

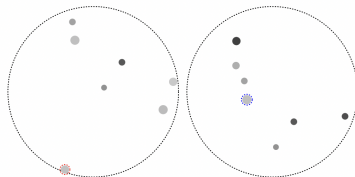
A: I have Jessica a friend of mine

A: and Josh, both went to columbia

B: *or anyone working at apple?*

B: SELECT (Jessica, Columbia, Computer Science, Google)

A: SELECT (Jessica, Columbia, Computer Science, Google)



Human A's view Human B's view

Human B: three light grey dots in a diagonal line

Human A: i dont have that but i have a black dot neer the top to the right, the only black dot in the circle

Human B: i have two black dots. find something else

Human A: ok i have a light grey dot by itself at the bottom to the left. right on the line

Human B: how big is it

Human A: its one of the bigger ones

Human B: okay just pick it then

Human A: ok

Human B: SELECT blue

Human A: SELECT red

Mutual Friends and OneCommon

Issue: Poor neural reasoning

From Mutual Friends: Neural + Human

- ▶ A: Know anyone who likes chess?
- ▶ B: None of my friends like chess.
- ▶ (conversation continues)
- ▶ A: Crocheting?
- ▶ B: None like crocheting.
- ▶ A: Chess?
- ▶ B: None like chess either, haha.

Sample of prior work in model-based planning

- ▶ 20 questions (Yu et al., 2019; Padmakumar and Mooney, 2020)
 - ▶ Sym: Asymmetric questioner + answerer
 - ▶ Turns: Multi-turn game
 - ▶ Lang: Closed class answers (observations)
 - ▶ Heur: Expected info gain heuristic
- ▶ EVPI (??)
 - ▶ Sym: Asymmetric questioner + answerer
 - ▶ Turns: No interaction, single turn game
 - ▶ Lang: Open
 - ▶ Heur: Expected utility heuristic
- ▶ RSA reference game (?)
 - ▶ Sym: Symmetric
 - ▶ Turns: Multi-turn game
 - ▶ Lang: Symbolic language
 - ▶ Heur: Bounded depth search

Conditioning in partner modeling

- ▶ Assuming conditional independence $p(o \mid h, a, y) = p(o \mid a, y)$ is harmful
- ▶ If you ask the same question twice, your belief changes both times!
 - ▶ $p(\text{yes} \mid h = \emptyset, a = \text{red dot?}, y)$ can vary depending on the latent y
 - ▶ $p(\text{yes} \mid h = (\text{red dot?}, \text{yes}), a = \text{red dot?}, y) = 1$, since we just asked!
- ▶ ‘Questions with correlated answers’ and deficient observation model lead to uncalibrated beliefs, and therefore poor strategy
- ▶ Contribution: relax independence assumption
 - ▶ Let past obs vote on current one (weighted by action similarity)
 - ▶ Probably solved by Transformers⁶

⁶Copy attention, depends on amount of data

Example dialogue 1: Overconfidence

0 i have a large black dot , it 's not the same size , but it is not the darkest

1 Mine is the darkest

0 i see it . i think it is the one .

1 I have multiple dark dots so I am not sure

0 i have a large black dot , it is in the center

1 I have a large dot in the lower part with a dark smaller dot at seven o'clock

0 is the black dot the darkest ? if so click it

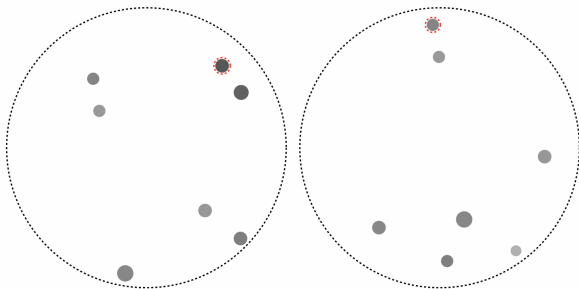
1 I have three the same darkness

0 pick the lightest one

1 Ok but our odds are not good.

0 pick the light one

Example dialogue 2: Overconfidence

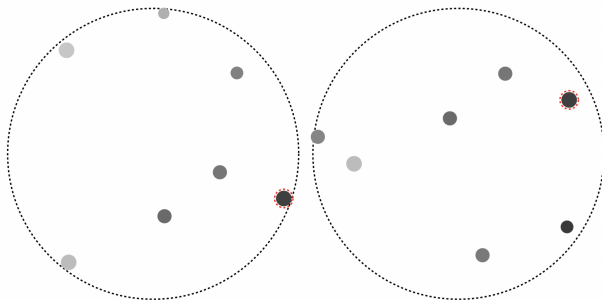


0 i have two dark dots , one on top and slightly smaller than the other

1 i see it. pick the top one?

0 ok

Example dialogue 3: Good humans



1 I have a large black dot by itself

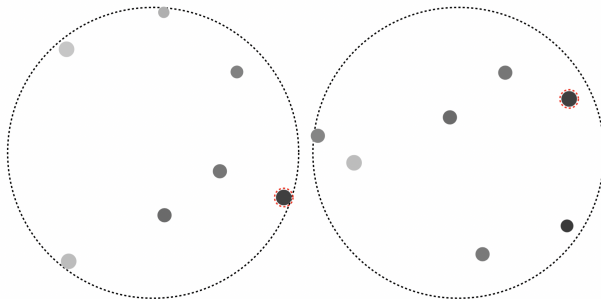
0 I see a large, very dark dot on the edge of my screen (so I won't be able to see anything to its right). Can you see anything on the left of your large black dot?

1 Yes, my large dark dot is on the edge of the right side

0 Ok, to the left of the dark dot, and slightly above it, do you see a slightly-smaller, slightly lighter dot?

1 yes

Example dialogue 3: Good humans



- | | |
|---|--|
| 0 | and then far above (and a bit to the right of) that lighter one, do you see a slightly smaller, identically colored dot? |
| 1 | No, the first lighter dot is the closest dot to the top of mine |
| 0 | Okay. what do you see to the left of that lighter dot? |
| 1 | A slightly darker dot that is below it just a bit |
| 0 | ok, I think we're in the same place. Let's click that original, blackest dot |
| 1 | Okay sounds good |