We need a task that provides both, (a) flexibility to facilitate learning, and (b) feedback to improve performance. A question is a linguistic expression used to make an observation, ask a question, or express an opinion. We propose a new computational framework that explains how people construct rich questions, treating question asking as program synthesis.

Related work
Goal-directed dialog systems typically only choose between a small set of canned questions (“What type of food are you looking for?”). Recent deep learning systems have shown interesting results but require large datasets of images paired with human questions. However, human data can, with virtually no practice, ask intelligent questions in novel scenarios, and can flexibly adapt to changes in task or goals.

We represent questions as programs that, when executed on the state of the world output an answer.

$$answer([size Red]) = \text{"3"}$$

We generate novel questions by synthesizing question components.

- What is the size of the red ship?
- What is the orientation of the red ship?
- Are the red ship and the blue ship parallel?
- What is the total size of all the ships?

PROBABILISTIC GENERATIVE MODEL

We fit a log-linear model over semantic expressions, in order to estimate the latent probabilities of asking different questions given the current context.

This model can be used to ask novel questions (i.e., plausible questions that no human asked) and to predict what questions people will ask in novel (untried) contexts.

- The space of questions $X$ is defined by a grammar.

**Grammar for Questions**

**Rules (Subset)**

- $A \rightarrow B$  
- $A \rightarrow C$  
- $B \rightarrow \text{TRUR}$  
- $B \rightarrow \text{ALU}$  
- $B \rightarrow \text{AU}$  
- $N \rightarrow \text{size C}$  
- $C \rightarrow \text{Blue}$  
- $C \rightarrow \text{Red}$  
- $C \rightarrow \text{Yellow}$  
- $C \rightarrow \text{Purple}$

**Question Features**

- $f_1$: Informativeness
- $f_2$: Expected Information Gain
- $f_3$: Complexity
- $f_4$: Probability under the probabilistic grammar
- $f_5$: Answer type
- $f_6$: Boolean, Number, Location
- $f_7$: Relevance

**Auxiliary Features** to filter out questions that do not address the game board.

- $\mathcal{E}(x) = \sum_{z \in \mathcal{Z}} \theta_f(z|x)$
- $\mathcal{M}(x) = \sum_{z \in \mathcal{Z}} \log(\mathcal{E}(x))$

**Maximum Likelihood Estimation**

- With question $x$ from the human-question data set.
- Optimization: We had to approximate the gradient (via importance sampling) since the set of all questions $X$ is intractably large.

RESULTS

**1. Ask Novel Questions**

Models that had one key feature lesioned achieved a lower log-likelihood than the full model using all features.

**2. Predicting What People Will Ask**

- The predictions of the full model showed strong alignment with the question frequencies in the data set for some contexts and more modest alignment for others (average correlation $r = 0.65$).

**3. Future Directions**

- Generalization to more domains beyond Battleship
- Tuning behavior of human vs machine questions
- Automatic translation between question programs and natural language

**Take Home Points**

- By treating questions as symbolic programs, our model can produce interesting and informative, “human-like” semantic questions.
- It can predict what questions people will ask in a given game context.
- It can learn from provided answers.

This also represents a new approach to query synthesis in active learning.