16. Context Modeling for Dialog Systems

Kata Naszádi
Supervisors: Christof Monz (UvA), Frans A. Olihoek (TU Delft)

Hybrid Intelligence

TU Delft

Universiteit van Amsterdam

Research Goals

Human communication is inherently guided by communicative goals found in real-world contexts. In this project we aim to move beyond simple sequence-to-sequence supervised learning approaches to dialog. We build models that include explicit accounts of how words affect the environment and other agents. We concern ourselves with communication taking place in grounded environments where agents have a shared goal.

Aligning Predictive Uncertainty with Clarification Questions in Grounded Dialog

⟨**A**⟩: start with blue

⟨**B**⟩: Excellent

⟨**A**⟩: put a line of three down

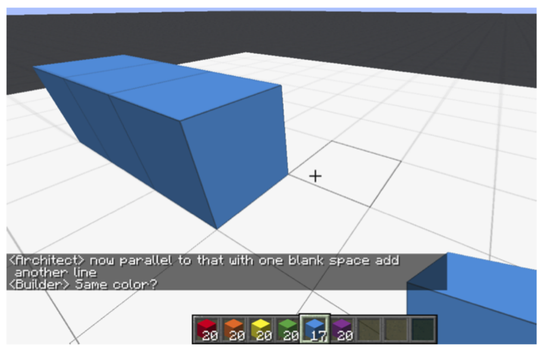
⟨**B**⟩ *puts down a blue blocks at (1,1,1), (1,1,0,), (1,1,-1)*

⟨**A**⟩: now parallel to that with one blank space add another line

⟨**B**⟩: Same color?

⟨**A**⟩: yep

⟨**B**⟩ *puts down blue blocks at (-1,1,1), (-1,1,0), (-1,1,-1)*



Asking for clarification is fundamental to effective collaboration. An interactive artificial agent must know when to ask a human instructor for more information in order to ascertain their goals.
In this work, we investigate if ambiguous linguistic instructions can be detected by uncertainty in neural models. We train an agent using the T5 encoder-decoder architecture to solve the Minecraft Collaborative Building Task.

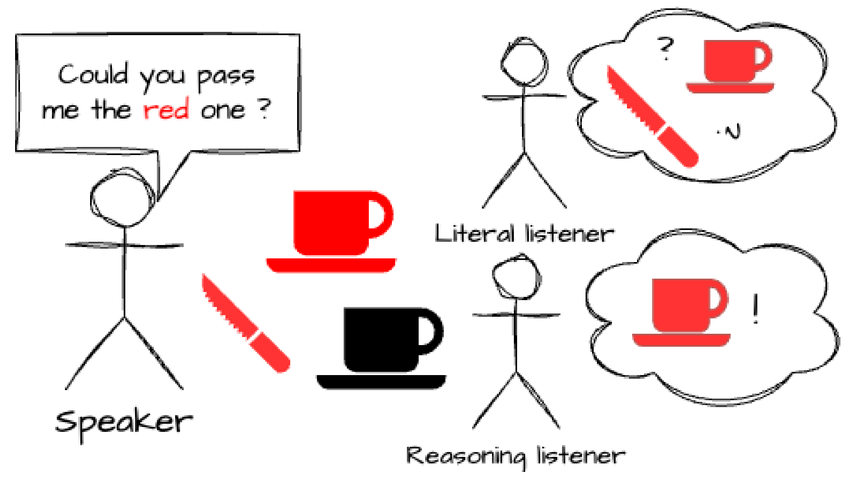
1. We identify uncertainty metrics that can detect ambiguous instructions.
2. We show that calibration can help increase the statistical separation between the uncertainty metrics of ambiguous and non-ambiguous instructions.
3. We provide a novel empirical analysis on the relationship between uncertainty and dialog history length. We show uncertainty increases as length increases, confounding the statistical separation that is important for detecting the need for questions. This poses an inherent difficulty as in collaborative dialogues, the history grows over the course of interaction.

	Measure	Level	Aggr.	N-best	ENS	Single
0	ENT	SEQ	-	5-best	0.54	0.49
1	ENT	TOK	avg	5-best	0.65**	0.59**
2	KL	TOK	avg	5-best	0.59**	-
3	LL	SEQ	-	1-best	0.62**	0.55*
4	LL	SEQ	diff	2-best	0.59	0.57

Area under the ROC curve for separating clear from unclear dialog turns. The better calibrated ensemble model (ENS) achieves better separation across all uncertainty measures, but both models struggle to separate cases where the human participants asked a question and where they decided to carry out the instruction.

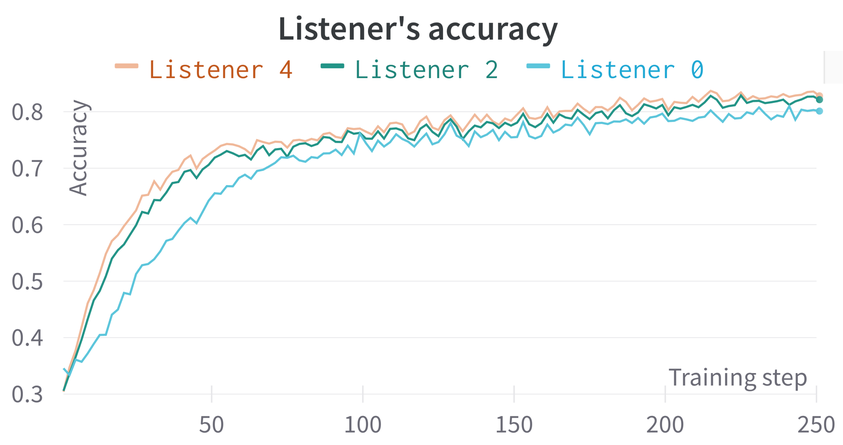
Using Recursive Reasoning While Learning in a Visual Reference Game

The meaning of a natural language utterance can greatly vary depending on the context of the communication. An artificial agent interpreting natural language needs to be able to integrate models of the human speaker and the communicative goal in order to arrive at the correct interpretation. Previous work has demonstrated that applying explicit reasoning about the situational factors during deployment leads to better comprehension. In this work, we propose to include contextual reasoning about such factors already during agent training, not only evaluation.



By focusing on a listener that has an internal model of a speaker, we show that a listener equipped with pragmatic reasoning during training:

- can successfully learn to interpret messages in a communicative task.
- learn faster in the initial stages as opposed to models that do not integrate explicit models of the speaker.
- achieve higher task accuracy at evaluation time compared to models that only apply reasoning during evaluation.
- learn interpretation functions that are more robust to changes in the environment.



Listeners applying higher levels of recursive reasoning about their opponents learn quicker. In this comparison all other parameters such as speaker level, number of distractors, correlation between observations are left constant.