April 2022

Learning to Ask Timely Questions in a Collaborative Grounded Language Understanding Task

Kata NASZADI [a,1], Michiel VAN DER MEER [1, b], Taewoon KIM [1, c] and Putra MANGGALA [1, a]

[a] *University of Amsterdam*
[b] *Leiden University*
[c] *Vrije Universiteit Amsterdam*

Keywords. grounded language understanding, predictive uncertainty, hybrid intelligence, dialog management

Introduction Asking questions is essential to effective collaboration. Being able to follow instructions is a fundamental way of demonstrating natural language understanding, but realizing when an instruction is unclear is an equally important element of successful grounded language interpretation. If an agent does not ask for clarification to ascertain their goal in a timely manner, it may fail to infer the most optimal path to achieving the best outcome, or worse, diverge to a poor quality local optimum.

Behavioral cloning is a common practical approach to learning from human-human demonstrations in a supervised manner. We focus our exploration on agents learned in this manner. Learning when to ask questions in a supervised manner may be ineffective due to data sparsity. Instead, we wish to employ predictive uncertainty over the agent's action space to estimate the agent's need to ask for further information. In this work, we evaluate how well the agent's uncertainty aligns with expected human clarification questions. We present this as a first step toward grounding the dialog turn-taking for general collaborative human-machine tasks.

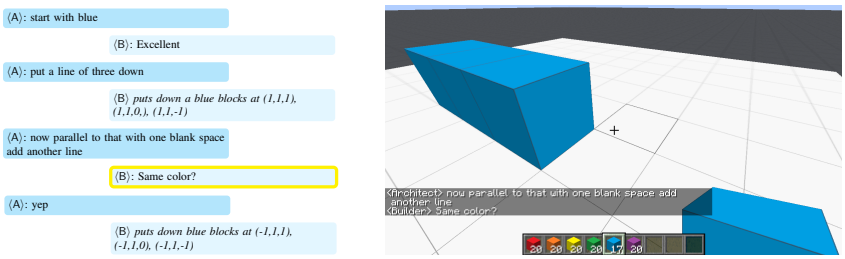

Figure 1. Example chat interaction between human instructor Architect (A) and agent Builder (B) from the Minecraft Dialog Corpus. The yellow highlighted Builder message indicates a clarification question asked by the Builder.

[1] Equal contributors.

Interactive Collaborative Task Setup In order to situate our agent in a collaborative task setting, we use the Malmo Minecraft engine [1] and a set of human-to-human conversations collected while the participants had to solve a variety of building tasks [2]. Our setup involves a human architect agent who has access to a target structure and can direct an artificial builder agent with natural language.

The builder is presented with multimodal input, as well as a history of preceding actions. For our experiments, we use the model presented in previous work [3] for solving the Builder Action Prediction (BAP) task. In the BAP task, the builder is silent: it is only trained to place and remove colored blocks. Training is done using a supervised signal from action sequences recorded from human-human interactions. We use beam-decoding to obtain action sequences from the model.

Using uncertainty for asking questions Consider the example in Figure 1 where the instruction is underspecified in terms of which color should be used (”now parallel to ...”). A model that attempts to predict color based on underspecified instructions may have its predictions lie on the decision boundary between color classes. We formalize this intuition using the model’s predictive uncertainty.

As the model was trained to output a sequence of actions, we define uncertainty over this structured prediction space. While developing such uncertainty estimates is an active field of research, as a first attempt we consider two simple baselines: length-normalized log-likelihood of the predicted sequence and entropy of the 5-best hypotheses. For the latter, we normalize the sequence likelihoods for the 5-best hypotheses to sum up to one.

At inference time, we evaluate predictive uncertainty in two cases: when the recorded data indicates that the human builder acted without asking any question (*no-question*) and when the builder decided to ask a question (*question*). An instruction following model with uncertainty that is calibrated to when questions should be asked will have relatively higher uncertainty in the *question* case.

In this ongoing work, our initial result for one trained model and two uncertainty metrics show that there is no clear difference in uncertainty metric magnitudes (Figure 2 and Figure 3) between *no-question* (blue) and *question* (red) cases. This indicates that these uncertainty values cannot be used at face value for this model to determine when to ask clarification questions. We intend to make this analysis more rigorous such that we can make recommendations on how to calibrate the uncertainties of probabilistic models in language understanding tasks.

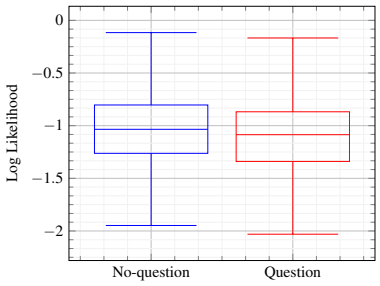

Figure 2.: Log likelihood over action predictions for gold-labeled question actions.

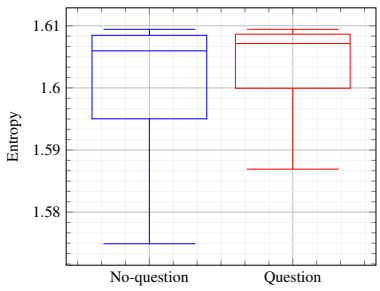

Figure 3.: Entropy scores of action probabilities for gold-labeled question actions.

References

[1] Johnson M, Hofmann K, Hutton T, Bignell D. The Malmo Platform for Artificial Intelligence Experimentation. In: IJCAI. Citeseer; 2016. p. 4246-7.

[2] Narayan-Chen A, Jayannavar P, Hockenmaier J. Collaborative Dialogue in Minecraft. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics; 2019. p. 5405-15. Available from: `https://aclanthology.org/P19-1537`.

[3] Jayannavar P, Narayan-Chen A, Hockenmaier J. Learning to execute instructions in a Minecraft dialogue. In: Proceedings of the 58th annual meeting of the association for computational linguistics; 2020. p. 2589-602.