2023 FRR & NRI Principal Investigators' Meeting
May 2-3, 2023

NSF Award #2144804

CAREER: Inferring Minimal but Sufficient Environment Models from Natural Language and Semantic Perception for Collaborative Robots in Dynamic Environments

Thomas Howard (PI), Siddharth Patki, Nikola Raicevic, and Francesca Daszak

UNIVERSITY of ROCHESTER

Motivation

Grounded language communication relies on the robot having a representation of the environment when it is time to interpret the meaning of an instruction, declare a fact, or ask a question. Since it is not practical or efficient to track all objects at arbitrary resolutions in an environment, new approaches for interpreting minimal but sufficient models of the environment are needed for efficient and effective grounded language communication with collaborative robots (Figures 1-3).

Fig. 1. An example of a language instruction that references the current state of the world

Fig. 2. An example of a language instruction that references the past state of the world

Fig. 3. An example of a language instruction that references the future state of the world

An Architecture for Language Guided Temporally Adaptive Perception

Distributed Correspondence Graphs

This project builds upon Distributed Correspondence Graphs (DCG) [1] to infer a symbolic meaning of language in the context of the robot's observed environment. This environment model can be constructed from sensor observations and/or human guidance as demonstrated in [2]. An example DCG is shown in Figure 4.

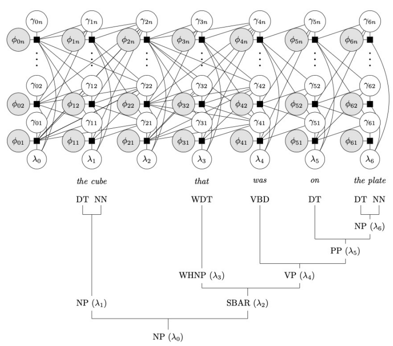

Fig.4. Distributed Correspondence Graph for the instruction "the cube that was on the plate" that references the past state of the world

Language Guided Temporally Adaptive Perception

Previous work under National Science Foundation National Robotics Initiative Awards #1637813 and #1638072 led to the development of Language Guided Adaptive Perception [3-5], which extracts information from language to construct a minimal but sufficient representation of an environment for instructions in observed and partially observed environments. The architecture below in Figure 5 illustrates a different approach to this problem, which attempts to generalize this approach to infer environment models for instructions that reference the current, the past, and the future state of the world.

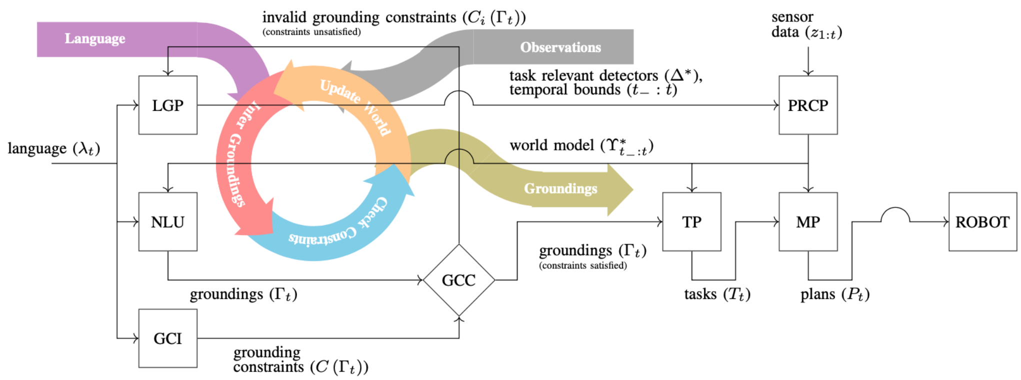

Fig.5. The proposed framework for language guided temporally adaptive perception. This architecture executes a loop involving symbol grounding, environment modeling, and constraint checking by using information from the expression about the type and quantity of symbols that are expected to ground.

Initial Experimental Results

We have implemented a version of our architecture for language-guided adaptive perception that is able to iteratively explore past observations to construct minimal but sufficient environment models for symbol grounding for instructions that refer to the past, such as "point to the box that has two gears" in an environment where the initial location of those gears is not known. The algorithm infers what symbols should be observed and the iterates through past observations using the Efficient Graph Updates [6] until those symbols are found. A detailed description of the problem, approach and experimental results (Figure 6) were submitted to a conference [7] and is currently undergoing peer review.

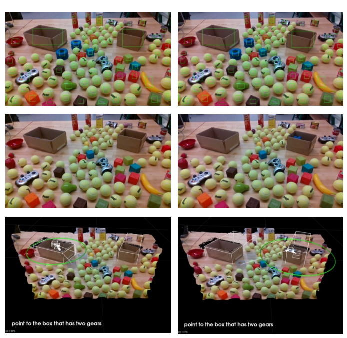

Fig.6. Illustrations of environments where minimal but sufficient models of the environment were extracted for the instruction "point to the box that has two gears". Only the dynamics of specific objects are extracted based on the learned model of language and perception. Further experiments demonstrate the speed gains of using Efficient Graph Updates as described in [6] as part of the closed-loop inference process.

Broader Impacts

The broader impacts of this work focus on four areas: research, course development, student development, and community science engagement. Updated versions of PI Howard's courses on Mechatronics and Embedded Systems and Robot Motion Planning and Manipulation were taught over the past two semesters. Additionally, there was progress on the development of new hardware platforms (Figure 7) that PI Howard is working on with a mechanical engineering senior design team that will be utilized in PI Howard's courses. PI Howard and his students also participated in robot demonstration outreach activities at the Rochester Museum and Science Center's summer camps.

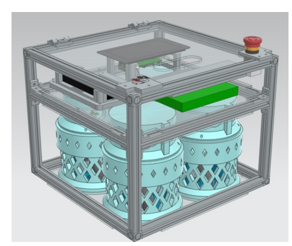

Fig.7. CAD model of a design of an omnidirectional robot chassis with independent steerable wheels from a ME 205 senior design team for teaching mechatronics, control, and autonomous mobile robots

[1] T.M. Howard, S. Tellex, and N. Roy, "A Natural Language Planner Interface for Mobile Manipulators," in Proceedings of the 2014 International Conference on Robotics and Automation, 2014.
[2] J. Arkin, D. Park, S. Roy, M. Walter, N. Roy, T. M. Howard, and R. Paul, "Multimodal estimation and communication of latent semantic knowledge for robust execution of robot instructions," International Journal of Robotics Research, vol. 39, no. 10-11, pp. 1279–1304, Jun. 2020
[3] S. Patki and T. M. Howard, "Language-guided adaptive perception for efficient grounded communication with robotic manipulators in cluttered environments," In: Proceedings of the 19th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 2018.
[4] S. Patki, A. Daniele, M. Walter, and T. M. Howard " Inferring Compact Representations for Efficient Natural Language Understanding of Robot Instructions". In: Proceedings of the 2019 International Conference on Robotics and Automation, 2019
[5] S. Patki, E. Fahnestock, T.M. Howard, and M. Walter, "Language-guided Semantic Mapping and Mobile Manipulation in Partially Observable Environments," In Conference on Robot Learning. Oct. 2019
[6] J. Arkin, S. Patki, J. Rosser, and T.M. Howard, "An efficient algorithm for visualization and interpretation of grounded language models," in IEEE International Conference on Robot and Human Interactive Communication. IEEE, 2022, pp. 266-273.
[7] S. Patki, J. Arkin, N. Raicevic, and T.M. Howard, "Language Guided Temporally Adaptive Perception for Efficient Natural Language Grounding in Cluttered Dynamic Worlds". Submitted to 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems.