PLDAC - Spring 2023
Are language models able to generate instructions for robots?

Laure Soulier, Nicolas Thome

Information

Supervisors: laure.soulier@isir.upmc.fr, nicolas.thome@isir.upmc.fr
Localization: Sorbonne University, France

Context

Autonomous agents require reasoning and planning strategies for performing tasks. We, therefore, believe that the semantics captured by large language models can enhance the decision process at different levels.

First, it can allow grounding object representations with common sense to identify their intrinsic and actionable properties. Large language models and also common sense knowledge bases, such as ConceptNet[1], can be used as complementary information sources, implying to design representation model leveraging multi-modal information. The difficulty would be to identify which properties are relevant for objects and how to fuse them into a single representation. Another strategy can be to encode objects differently according to each modality and then use self-attention to learn the possible interactions that are relevant for the task solving. Object grounding has been addressed in previous work [1, 6, 7], but we believe that a larger point of view related to the object scene is crucial to better model the context and object properties.

Second, natural language can serve for building and clarifying the planning strategy, and therefore the actions done by a robot. Several works have addressed instruction identification as abstract representation [2, 4, 5] or natural language expression, but the limited data supervision is often a challenge [3, 5]. To tackle this issue, we propose to develop interactive training processes, which imply asking humans to label situations with sentences, with strong care on limiting interactions to a few relevant situations, to reduce human effort. The underlying assumption is that the compositionality of language is correlated to compositionality in the agent's world.

In this project, we aim at evaluating the ability of neural language models to generate instructions for robots. We would like to identify if they have all the required knowledge for instruction generation. Typically, we know as humans that it is impossible to cut a bowl with a knife. However, it is not obvious that language models are able this knowledge. Indeed, early experiments in the domain highlight that when we train a model with realistic use cases, such as "generating instructions for cutting a tomato", the model is able at inference step to generate the same instruction when we ask the robot to cut a bowl. However, some knowledge resources includes those knowledge (from a direct or indirect way: a bowl is in ceramic, and a knife cannot cut ceramic, therefore we can deduce that a knofe cannot cut a bowl).

[1] `https://conceptnet.io/`

Objectives

The workplan proposed to the students are as follows :

- Given an instruction dataset, identifying salient knowledge required to generate instructions
- Match these knowledge with knowledge resources
- Evaluate whether language models integrate these resources. One inspiring paper is entitled: "language models as knowledge bases?" https://arxiv.org/abs/1909.01066

References

[1] Michael Ahn et al. "Do As I Can and Not As I Say: Grounding Language in Robotic Affordances". In: *arXiv preprint arXiv:2204.01691*. 2022.

[2] Jacob Andreas et al. "Learning with Latent Language". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, June 2018, pp. 2166–2179. DOI: 10.18653/v1/N18-1197. URL: https://aclanthology.org/N18-1197.

[3] Haonan Chen et al. "Enabling Robots to Understand Incomplete Natural Language Instructions Using Commonsense Reasoning". In: *2020 IEEE International Conference on Robotics and Automation, ICRA 2020, Paris, France, May 31 - August 31, 2020*. IEEE, 2020, pp. 1963–1969. DOI: 10.1109/ICRA40945.2020.9197315. URL: https://doi.org/10.1109/ICRA40945.2020.9197315.

[4] Athul Paul Jacob et al. "Multitasking Inhibits Semantic Drift". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, June 2021, pp. 5351–5366. DOI: 10.18653/v1/2021.naacl-main.421. URL: https://aclanthology.org/2021.naacl-main.421.

[5] Pratyusha Sharma et al. "Skill Induction and Planning with Latent Language". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*. Ed. by Smaranda Muresan et al. Association for Computational Linguistics, 2022, pp. 1713–1726. DOI: 10.18653/v1/2022.acl-long.120. URL: https://doi.org/10.18653/v1/2022.acl-long.120.

[6] Mohan Sridharan et al. "Combining Commonsense Reasoning and Knowledge Acquisition to Guide Deep Learning in Robotics". In: *CoRR* abs/2201.10266 (2022). arXiv: 2201.10266. URL: https://arxiv.org/abs/2201.10266.

[7] Antigoni Tsiami et al. "Multi3: Multi-Sensory Perception System for Multi-Modal Child Interaction with Multiple Robots". In: *2018 IEEE International Conference on Robotics and Automation, ICRA 2018, Brisbane, Australia, May 21-25, 2018*. IEEE, 2018, pp. 1–8. DOI: 10.1109/ICRA.2018.8461210. URL: https://doi.org/10.1109/ICRA.2018.8461210.