Exploring Large Language Models as a Source of Common-Sense Knowledge for Robots

Felix Ocker[1,*], Jörg Deigmöller[1] and Julian Eggert[1]

[1]Honda Research Institute Europe, Carl-Legien-Str. 30, 63073 Offenbach am Main, Germany

Abstract

Service robots need common-sense knowledge to help humans in everyday situations as it enables them to understand the context of their actions. However, approaches that use ontologies face a challenge because common-sense knowledge is often implicit, i.e., it is obvious to humans but not explicitly stated. This paper investigates if Large Language Models (LLMs) can fill this gap. Our experiments reveal limited effectiveness in the selective extraction of contextual action knowledge, suggesting that LLMs may not be sufficient on their own. However, the large-scale extraction of general, actionable knowledge shows potential, indicating that LLMs can be a suitable tool for efficiently creating ontologies for robots. This paper shows that the technique used for knowledge extraction can be applied to populate a minimalist ontology, showcasing the potential of LLMs in synergy with formal knowledge representation.

Keywords

Knowledge Extraction, Large Language Models, Common-Sense Knowledge, Robotics

1. Introduction

To assist humans effectively in everyday life, service robots need a sound foundation of common-sense knowledge to guide their actions. While there has been significant research on Knowledge-Based Systems (KBSs) for robotics [1], creating the underlying knowledge bases remains a considerable challenge [2]. This is due to the fact that common-sense knowledge is inherently implicit, i.e., it is naturally understood by humans but remains mostly unexpressed. However, the recent advances in LLMs have opened up a potential source of such knowledge. This paper explores state-of-the-art LLMs as a source of common-sense knowledge for robots, and investigates if they could even replace KBSs.

2. Related Work

Common-sense knowledge for robots involves actions necessary to reach a desired state, defined by an action verb, an actor, an object, and a tool, forming an action pattern [3]. Existing community efforts, e.g., ConceptNet [4], offer valuable information but face scalability challenges and heterogeneity due to the variety of contributors. LLMs such as OpenAI's ChatGPT and

ISWC 2023 Posters and Demos: 22nd International Semantic Web Conference, November 6–10, 2023, Athens, Greece

*Corresponding author.

felix.ocker@honda-ri.de (F. Ocker); joerg.deigmoeller@honda-ri.de (J. Deigmöller); julian.eggert@honda-ri.de (J. Eggert)

BigScience's BLOOMZ seem like a promising alternative due to the large amounts of common-sense knowledge they incorporate. As the usability of LLMs increases, knowledge extraction methods have evolved from demasking [5] to zero-shot prompts [6].

The application of LLMs to robotics leads to the emergence of embodied Artificial Intelligence (AI). However, success rates of such systems are still limited, for instance, up to 85 % for tidying up [7] and 48 % for task planning [8], posing challenges for robotics applications. Until the performance of LLMs improves further, a synergistic use with KBSs seems reasonable.

3. Extracting Action Patterns from Large Language Models

3.1. Formalizing Action Patterns

This work focuses on common-sense knowledge in the form of action patterns. An action pattern AP is represented by an action a, which is executed by a set of *Agents* $\mathscr{A}$ using a set of *Tools* $\mathscr{T}$ to modify a set of *Objects* $\mathscr{O}$, cp. Equation (1). These patterns capture various scenarios a robot might encounter and how it can act to change these situations. This knowledge allows the robot to infer appropriate actions to fulfill its goals, such as serving bread. Note that the knowledge should be grounded in real-world settings and continuously tested, possibly resulting in the robot confirming, discarding, or refining its knowledge. Note that the states of all elements before and after the action can be described via a set of attributes $\mathscr{S}_{before}$ and $\mathscr{S}_{after}$, respectively. In addition, the spatial relations $\mathscr{R}_{spatial}$ between the elements involved may change.

$$AP = (a, \mathscr{A}, \mathscr{O}, \mathscr{T}, \mathscr{S}_{before}, \mathscr{S}_{after}, R_{spatial}) \tag{1}$$

The set $\mathscr{AP}$ comprises all valid action patterns in the sense that they are meaningful in real-world applications. For instance, when tasked with baking bread, a robot has to identify suitable tools, as illustrated by the incomplete action pattern in Equation (2).

$$AP_{bake} = (bake, \{robot\}, \{bread\}, \{\textbf{\$tool}\}, \{cold\}, \{hot\}, \{outside\}) \tag{2}$$

Action patterns can be expressed using Web Ontology Language (OWL) 2 Description Logic (DL), a subset of predicate logic. Here, *Actions* are defined by their relations to the notions *Object*, *State*, *Location*, and *Time*, forming action patterns. These notions, rooted in previous work, cp. Section 2, can be aligned with top-level ontologies such as Basic Formal Ontology. An *Object* can also take the role of an *Agent* or a *Tool*, and spatial relations can be expressed using existing ontologies. Equation (3) shows the DL representation of the notions *Action* and *Object*.

$$\begin{aligned} Action \sqsubseteq\ & \exists \geqslant 1.has\_agent.Object \sqcap \exists \geqslant 1.has\_object.Object \sqcap \exists \geqslant 0.has\_tool.Object \sqcap \\ & \exists = 1.has\_location.Location \sqcap \exists = 1.has\_time.Time \\ Object \sqsubseteq\ & \exists \geqslant 1.has\_state.State \end{aligned} \tag{3}$$

3.2. Extracting Parts of Action Patterns

In order to replace KBSs, LLMs would need to be capable of consistently answering queries such as "Which tool can I use to bake bread?". This task can be posed generically using Prompt 1. Note that specifying *candidates* corresponds to a simple grounding in reality. Further, we modified Prompt 1 to extract the state of an object before and after a given action has been applied to it. The prompt can also be adapted to extract details about the spatial relationships between the object and the tool used.

Prompt 1. *In the following, I will ask you a question. In your response, I want you to answer with nothing but a list of suitable comma-separated words sorted by relevance. Which tool can I use to* ***$action $object****? Choose only from the following candidates:* ***$candidates****.*

Example: ... Which tool can I use to bake bread? Choose only from the following candidates: bowl, oven, knife, ...

3.3. Extracting Entire Action Patterns

The large-scale extraction of action patterns to incorporate into a knowledge base can be achieved using Prompt 2. Compared to Prompt 1, this prompt specifies the domain of interest and permits constraints within the action pattern, offering additional flexibility. The information extracted by Prompt 2 may not be directly applicable, but it is well-suited for populating an ontology with broad common-sense knowledge on a large scale.

Prompt 2. *Please respond with nothing but lists of the form '(action, agent, object, tool)'. An action pattern is defined by an action, i.e., a verb, an agent who executes the action, an object, which is modified, and optionally a tool. Generate* ***$number*** *action patterns for the domain of interest '****$domain_of_interest****'.*

Example: ... Generate 100 action patterns for the domain of interest 'kitchen'.

4. Experiments

To evaluate the LLMs' suitability as a source of common-sense knowledge, we used a ground truth consisting of 97 action patterns. The data set was created in a study with 20 participants, who completed the action patterns by determining the states of each object before and after the action, identifying the tools used, and outlining the spatial relationships between the objects and tools. An excerpt of this ground truth for the action "cut bread" is shown in Listing 1.

Listing 1: Excerpt from the ground truth for action patterns.

```
{
    "action": "cut",
    "object": "bread",
    "tools": ["knife", "fork", "key"],
    "object_states_before": ["fresh", "uncut", "whole", "unsliced", ...],
    "object_states_after": ["fresh", "cut", "crusty", "sliced", ...],
    "spatial_relations": ["bread near knife", ...]
},
```

4.1. Architecture and Models

We conducted preliminary experiments for choosing suitable LLMs, including ones with Vicuña's 7B and 13B models and OpenAssistant. Despite performing well in chat-like scenarios, these models had difficulty returning structured results. OpenAI's models appeared particularly promising, so we included them along with BLOOMZ as an open-source alternative and BERT as a baseline. For BERT, we rephrased the prompts as demasking tasks. We accessed gpt-3.5-turbo and gpt-4 via the OpenAI API and BLOOMZ via the Hugging Face API, and deployed BERT locally using a Hugging Face transformers pipeline. While we experimented with various temperatures, we ended up using a temperature of 0 to ensure reproducability. Providing candidates leads to significantly better results when comparing the responses to the ground truth. However, it is valid to assume that candidates are available, as this corresponds to a grounding in the robot's environment, which can be achieved using an appropriate perception module. The architecture we developed for the evaluation is shown in Figure 1. We also adapted this architecture to populate an ontology revolving around the notions described in Section 3.1 with the action patterns extracted to demonstrate the use of the information extracted, see Figure 1 bottom right.

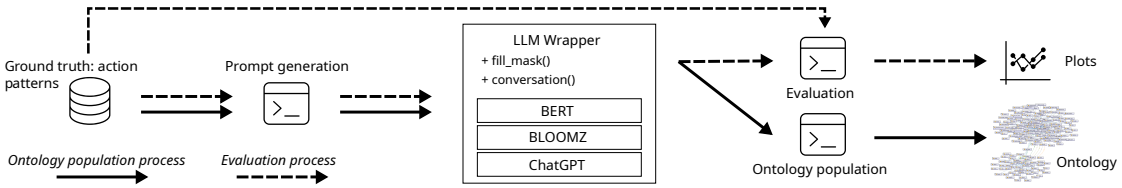

Figure 1: Architecture for assessing the LLMs and populating an ontology.

4.2. Results

Figure 2 summarizes the results. It shows F1@n-scores, i.e., the harmonic mean of precision and recall considering the first n elements from the ground truth. The left plot shows that the OpenAI models outperform others, including a ConceptNet-based baseline, in extracting tools $\mathcal{T}$. For extracting spatial relations $\mathcal{R}_{spatial}$, demasking was particularly effective. Extracting object states before $\mathcal{S}_{before}$ and after $\mathcal{S}_{after}$ the action yielded worse results. Changes in model temperature had little impact. These results are in line with success rates in LLM applications [7, 8]. To also evaluate the general usefulness of LLMs for common-sense knowledge extraction, we extracted 100 action patterns for the "kitchen" domain. Upon manual review, all 100 action patterns retrieved were found to be valid.

5. Conclusion and Outlook

Our experiments indicate that LLMs are a scalable source for general common-sense knowledge in the form of action patterns, which is a valuable basis for an ontology usable by robots. However, LLMs are still insufficiently reliable for providing actionable knowledge for robotics applications by themselves. Thus, we suggest pursuing the combination of LLMs and knowledge

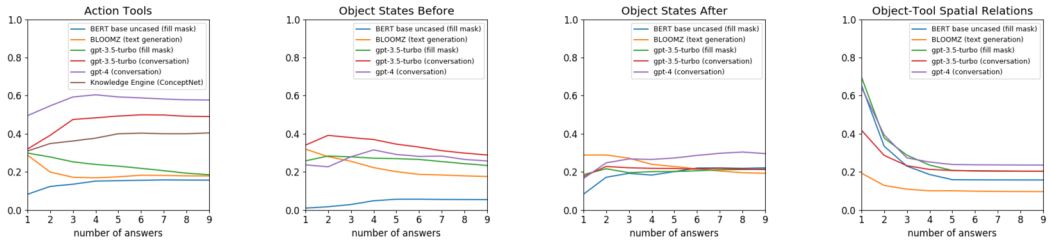

Figure 2: F1@n-scores, with n on the x-axis, for the extraction tasks compared to the ground truth ($temperature = 0$, candidates provided).

graphs for robotics, especially in the form of an integrated pipeline for ontology population from LLMs, and validation and reasoning using symbolic AI. Future work should also address the validation of newly extracted information using the existing knowledge graph in the spirit of Pan et al. [9], and the continuous evolution of the knowledge base using all sources available, ranging from databases created in community efforts to LLMs.

Supplementary Material: The ground truth dataset, the prompts, and a Python script for creating an OWL ontology from the action patterns extracted are available via GitHub[1].

References

[1] A. Olivares-Alarcos, D. Beßler, A. Khamis, et al., A review and comparison of ontology-based approaches to robot autonomy, The Knowledge Engineering Review 34 (2019).

[2] R. Gupta, M. J. Kochenderfer, D. Mcguinness, G. Ferguson, Common sense data acquisition for indoor mobile robots, in: AAAI, 2004.

[3] J. Eggert, J. Deigmöller, L. Fischer, A. Richter, Action representation for intelligent agents using Memory Nets, in: IC3K, 2020.

[4] R. Speer, J. Chin, C. Havasi, Conceptnet 5.5: An open multilingual graph of general knowledge, in: AAAI, 2017.

[5] V. Losing, L. Fischer, J. Deigmöller, Extraction of common-sense relations from procedural task instructions using BERT, in: Global WordNet Conference, 2021.

[6] J. H. Caufield, H. Hegde, V. Emonet, et al., Structured prompt interrogation and recursive extraction of semantics (SPIRES): A method for populating knowledge bases using zero-shot learning, arXiv preprint arXiv:2304.02711 (2023).

[7] J. Wu, R. Antonova, A. Kan, et al., Tidybot: Personalized robot assistance with large language models, arXiv preprint arXiv:2305.05658 (2023).

[8] I. Singh, V. Blukis, A. Mousavian, et al., Progprompt: Generating situated robot task plans using large language models, in: ICRA, 2023.

[9] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, X. Wu, Unifying large language models and knowledge graphs: A roadmap, arXiv preprint arXiv:2306.08302 (2023).

[1] https://github.com/HRI-EU/common_sense_for_robots