CoRe Essay 6
Possible Computational Frameworks for Intentionality

Haochen Zhao*
Yuanpei College
Peking University
2100017417@stu.pku.edu.cn

Abstract

In social interactions between humans, the concept of intention is of great significance. From the environment around us and actions we can observe, we can do intentionality inference naturally. This kind of intentionailty inference helps us understand agents' behaviors, as well as their possible future actions. Also, intentionality is driving our own actions as well. For AI agents capable of cooperating with humans in sophisticated social scenarios, computational frameworks for intentionality is needed. This essay explores diverse methodologies employed to encapsulate goals, ranging from abstract symbolic representations to detailed state-based and hierarchical structures. The advantages and disadvantages of each representation are scrutinized, emphasizing the nuanced trade-offs inherent in the pursuit of effective goal modeling.

1 Introduction

The modern concept of intentionality in 19th-century contemporary philosophy can be attributed to Franz Brentano. As a German philosopher and psychologist hailed as the progenitor of act psychology, also known as intentionalism, Brentano reintroduced this concept in his seminal work, "Psychology from an Empirical Standpoint" (1874)[3]. In delineating intentionality, Brentano posited it as an inherent attribute of all acts of consciousness, endowing them with a distinctly "psychical" or "mental" nature, thereby distinguishing them from the realm of "physical" or "natural" phenomena. More formally, intentionality is the power of minds to be about something: to represent or to stand for things, properties and states of affairs. Intentionality is primarily ascribed to mental states, like perceptions, beliefs or desires.[2]

Intentionality endows machines with a purposeful direction, allowing them to navigate complex environments, make informed decisions, and adapt to evolving circumstances. As we delve into the myriad ways goals can be computationally represented, it becomes evident that the nuanced interplay between intentionality and representation shapes the very fabric of intelligent systems.

As we navigate this intellectual terrain, it becomes increasingly apparent that a deep understanding of intentionality is not only desirable but indispensable for the realization of intelligent systems capable of navigating the complexities of our ever-changing world. The choices made in representing goals within a computational framework not only reflect the objectives of the system but also embody the intentional design decisions that drive its actions.

This essay embarks on an exploration of the multifaceted landscape of computational goal representations, recognizing the pivotal role of intentionality in this endeavor. From abstract symbolic representations to detailed state-based and hierarchical structures, each methodology seeks to capture the intentional essence of goals in its unique way. The ensuing discussion dissects the advantages and disadvantages of these representations, unveiling the intricate dance between intentionality and computational efficiency.

*Thanks to course instructor Yixin Zhu, TA Guangyuan Jiang and Yuyang Li for their helpful suggestions.

2 Symbolic Representation

Symbolic representation involves expressing goals through symbols, abstract language, or conceptual constructs. In the computational realm, these symbols can be variables, tokens, or even high-level linguistic constructs that encapsulate the essence of a goal. This approach allows for the representation of a wide range of goals, from simple and concrete to complex and abstract.

The adaptability of symbolic representation makes it suitable for a diverse array of goals, ensuring that the system can handle a wide range of tasks and objectives. Also, human operators can comprehend and communicate goals effectively, facilitating collaborative decision-making between humans and intelligent systems. In summary, symbolic representation has the advantages of flexibility and readability.

On the other hand, symbolic representations may introduce ambiguity, especially when dealing with abstract or subjective concepts. The interpretation of symbols can vary, leading to potential misunderstandings. What's more, as goals become more detailed and intricate, the symbolic representation may become complex and challenging to manage. Balancing expressiveness with simplicity is a delicate task.

In the early development of this field, symbolic representation often referred rule-based methods, involving logical computations. But nowadays, natural language is another promising way as large language models are developing rapidly.

3 Hierarchical Representation

Hierarchical representation is commonly used in social cognition tasks in this decade, especially the ones with reinforcement learning methods.[1]

The key idea of hierarchical representation is to break down the high-level goal into a hierarchy of subgoals and demonstrable states. Each subgoal can be further decomposed into more specific states if needed. This allows for planning, reasoning, and monitoring of progress towards the top level goal.

The state representations make it clear what conditions need to be satisfied. This enables an agent to monitor the state of the world and track goal achievement. The hierarchical breakdown enables reusable subgoals and clear modularization. The simplicity of representing goals as states makes this approach interpretable but less flexible than richer representations.

The modular nature of hierarchical representation supports ease of design, maintenance, and modification. Changes to one part of the hierarchy don't necessarily impact the entire system. Apart from that, hierarchies provide a clear and intuitive way to represent both high-level objectives and detailed, low-level actions. This abstraction facilitates effective communication and understanding.

To make it easier to understand, a simple example is provided as followed.

Top-Level Goal

- State: Prepare a healthy dinner

Subgoals

- State: Kitchen is clean
  - State: Counters and sink are cleared
  - State: Dishes are loaded in dishwasher
  - State: Surfaces are wiped down
- State: Ingredients purchased
  - State: Check ingredients available
  - State: Make grocery list
  - State: Go to grocery store
  - State: Purchase needed ingredients
- State: Meal is cooked

  - State: Look up recipe
  - State: Prep ingredients
  - State: Follow recipe instructions
  - State: Set cooking timers
  - State: Food is fully cooked
- State: Kitchen is cleaned up
  - State: Leftovers are stored
  - State: Serving dishes are washed
  - State: Counters are wiped down
  - State: Trash is taken out

4 Possible Implementation

The representations mentioned above have their own drawbacks. Thus, a carefully designed combination of these representations may alleviate this problem and achieve better performance. Actually, some important steps are already been progressed.

A popular open-source agent, Auto-GPT, which takes a goal in natural language and breaks it into sub-tasks and tries to solve them using internet and other tools, made a sensation when it first came out. Though its limited capability for real-world engagement and the absence of benchmarks contribute to uncertainties of performance[4], the idea behind it is inspiring. Recent large language models shows a sign of generalization, which can be exploited in complicated intentionality tasks as well.

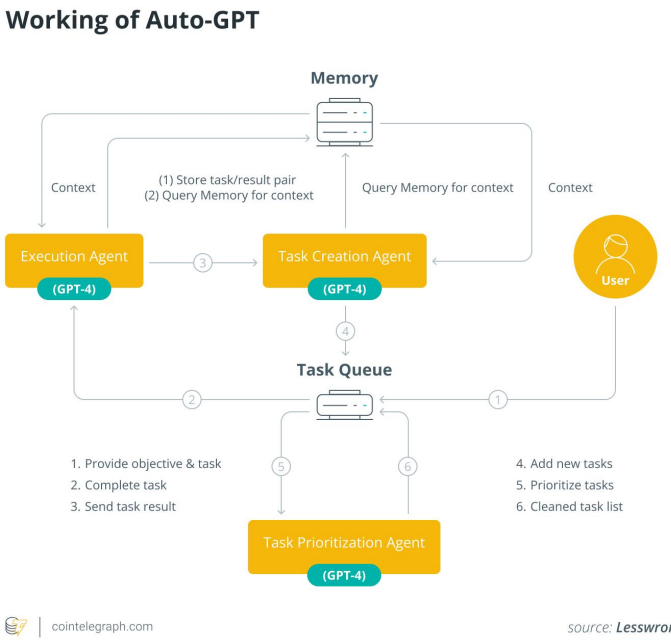

Figure 1: Overview of Auto-GPT

Many people have found that advanced large language models like ChatGPT or GPT-4 are good at abstract reasoning, but perform less satisfying when comes to low-level subtask implementations. It is reasonable because these large language models are trained on large natural language corpus, without much low-level control information.

Thus, adding some downstream models in the hierarchy to carry out the real steps on the ground may be beneficial. For example, embodied agents can be equipped with low-level motion planner trained by mature reinforcement learning methods after the llm-driven subgoal generator.

5 Conclusion

In the intricate dance between computational systems and the representation of goals, the significance of intentionality becomes evident. The exploration of diverse methodologies, from symbolic representations to hierarchical structures, underscores the dynamic interplay between computational efficiency and the intentional design of systems. As we conclude this exploration, it is crucial to reflect on the broader implications of our understanding. This essay gives a brief introduction of two different representation and a hybrid implementation. The hybridization of these representations may offer a middle ground, leveraging the strengths of each while mitigating their respective weaknesses.

References

[1] Thommen George Karimpanal. Intentionality in reinforcement learning. 2021. 2

[2] Wikipedia contributors. Intentionality — Wikipedia, the free encyclopedia, 2023. URL https://en.wikipedia.org/w/index.php?title=Intentionality&oldid=1169027981. [Online; accessed 12-November-2023]. 1

[3] Wikipedia contributors. Psychology from an empirical standpoint — Wikipedia, the free encyclopedia, 2023. URL https://en.wikipedia.org/w/index.php?title=Psychology_from_an_Empirical_Standpoint&oldid=1180857947. [Online; accessed 12-November-2023]. 1

[4] Hui Yang, Sifu Yue, and Yunzhong He. Auto-gpt for online decision making: Benchmarks and additional opinions, 2023. 3