Equip AI with Tool Use Ability: A Brief Survey from Human Evolution View

Jingze Zhang
Department of Automation
Tsinghua University
jz-zhang21@mails.tsinghua.edu.cn

Abstract

One of the key distinctions between humans and animals is proficiency in crafting and utilizing tools. Likewise, acquiring the skill to use or even create tools represents a crucial benchmark and milestone on the path towards *general artificial intelligence*. Nonetheless, teaching an intelligent agent how to employ tools is a notably challenging endeavor, mainly demanding proficiency in three core aspects: perception, manipulation, and cognitive reasoning, which necessitates the agent's clear understanding of physics world and causal relationships[7]. In this essay, I will first offer a definition of *Tool Use* and introduce the background of the problem(Section 1). Subsequently, we will delve into the current advancements and limitations in the three essential capabilities required for an agent to use tools(Section 2). Finally, we will showcase the current abilities of intelligent agents in *Tool Use* before a brief summary, shedding light on both their present capabilities and the future directions of development(Section 3&4).

1 Introduction

Using human tools is a fundamental capability in the path towards *General Artificial Intelligence*[12]. By utilizing human-made tools, robots can accomplish more tasks which are tough without tools. However, tool use is a challenging and open-world problem, which is considered as the benchmark to distinguish human from other animals. Unfortunately, at present, robots still lag behind certain crows and primates in their ability to use tools2, with a notably limited capacity for generalizing reasoning when manipulating objects[1, 7].

Figure 1: Adult New Caledonian crow dexterously extracting a longhorn larva clamped onto the end of its stick tool[8]. Its 0.1W-powered brain possesses the ability to choose the right stick for obtaining food through few-shot learning, a task that currently eludes even the most intelligent artificial agents.

1.1 Definition of robot tool use

Before delving into the discussion of the issue of tool usage, it might be helpful to start by providing a clear definition of tool usage and differentiating it from mere interaction with objects. We adopt the definition of tool usage provided by Qin et al. [7]:

Definition of Robot Tool Use

A robot attaches or secures to its end-effector an external, unanimated, freely available object or an object attached to another object, in order to achieve a goal of altering the state of another object, updating its own state, or other goals, through purposeful manipulations.

The definition above has 2 key points. First, a tool use problem must be motivated by a specific goal, rather than simply manipulating objects with the end effectors. Second, the agent should have a clear understanding of causality between the target and the action itself. Both of these steps are well exemplified in the process of crows using tools to obtain food.

1.2 Classification of tool usage

Tool usage is a complex open-world problem. With an uncountable variety of tools and an endless array of tasks, this issue is exceptionally intricate. Before we delve into the various modules required for further exploration of tool usage, let's begin by analyzing the different categories of tool usage from a human perspective.

Through literature research, we discovered that the issue of the taxonomy of tool usage had been previously explored in the literature review by Qin et al. [7]. The problem of tool usage in machines can be broadly categorized into two main classes: causal tool usage and non-causal tool usage. Furthermore, this article further subdivides tool usage into eight subcategories based on the types of prior knowledge and samples used during tool utilization. In most common scenarios, tool usage requires causality. This classification method provides greater clarity to the problem of tool usage in robotics.

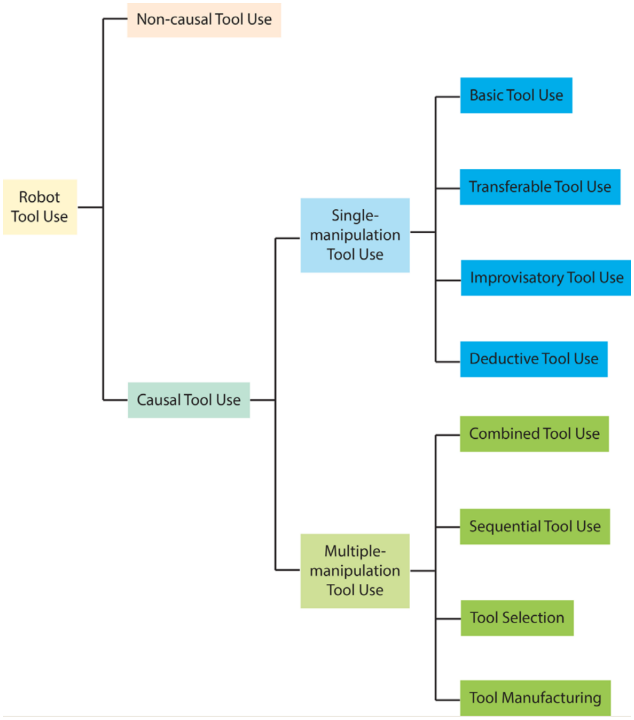

Figure 2: Classifation of tool use problem. [7]

In other words, the problem of tool use can be classified into 4 categories or stages:

- Solving old tasks with old tools (a simple problem, very strong prior).
- Solving novel tasks with old tools (transfer learning, strong prior).
- Solving novel tasks with novel tools (complex causality problem, weak prior)
- Manufacture novel tools to solve novel tasks, a sign of AGI.

2 Abilities behind tool use problem

The utilization of tools by robots primarily requires capabilities in perception, manipulation, and reasoning [2, 7, 10]. In the following two subsections, our focus will be on the analysis of perception and reasoning abilities. Regarding the manipulation abilities of robots, there are already numerous algorithms that have made substantial progress, and we will not delve into that discussion here.

2.1 Perception: object and scene level understanding

Tool use is an *open-world task* involving complex relationships between object, actions and effects[2]. As for human, the manipulation of a tool can only take place after our understanding the functionaliy or affordance of the object. Similarly, the first step of tool usage is to understand the affordance and nature of an object or scene. For instance, in the tools transfer learning task demostrated in 3, the agent needs to understand hard objects are required to crack open the walnut.

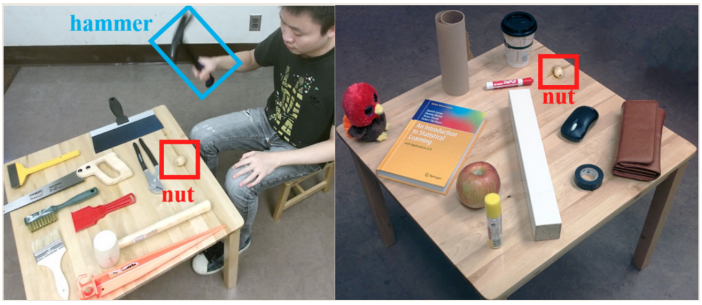

Figure 3: Tools transfer learning task. [11]

A good news is that affordance learning and scene understanding has been reserched from different perspectives[3, 4, 9].

2.2 Causality: pave the path towards target

The use of tools is a task-driven challenge. In the process of transitioning from an initial state to a goal state with the assistance of tools, an intelligent agent needs to comprehend the causal relationships involved, infer the effects of tool usage through these relationships, and thereby achieve the task with a relatively low error rate in scenarios with limited sample learning. The inference and deductive process requires the information from the perception process.

Furthermore, causality is also required when it comes to ethical. There is a strong need for interpretability when it comes to the process of intelligent agents solving problems with tools. When machines have access to tools, they cease to be the *brains in a vat*. It can be highly perilous if we do not have a clear understanding of the intent behind their actions. The intention cannot be revealed without causality.

3 Discussion

In the field of robotic tool usage, I am particularly interested in two aspects: tool deductive learning, and tool manufacturing. From my aspect of view, without these 2 tasks, We were not much different from animals. Here I will give a brief discussion these 2 fields.

Deductive tool use is the task that a robot can solve a novel task with a novel tool with little prior knowledge about the tool and the task. Solving *Chinese Ring Puzzle* is a process of deductive learning.

It requires the agent to infer the entire affordance model and have a world model in alignment human. This is a very challenging task. And I didn't find any studies showcase their agents can conduct deductive learning.

Tool manufacturing is the final target of robot tool use problem. It's also an open problem and have a long way to explore. In Nair et al. [5] research, they put forward a method to construct tools via geometry learning and reasoning. A similar method was used in [6]. However, these abilities fall behind from human and have a weak capability of generalization. And I think a super model which is a complex combination of perception and reasoning is necessary to solve the task.

4 Summary

During this essay, we have delved into the definition of tool usage by intelligent agents, its significance, the necessary dimensions of capability, and the current state of academic progress. Through literature review, it becomes apparent that intelligent agents currently lack the ability to learn to use simple tools through few-shot or zero-shot learning, let alone the capacity to fabricate tools. We have explored that, for an intelligent agent, acquiring tool usage necessitates a complex array of skills such as comprehensive perception, causal reasoning, object manipulation, and a clear understanding of objects and the environment. It seems to require a complex system containing all the module mentioned above.

References

[1] AMP von Bayern, Samara Danel, AMI Auersperg, Berenika Mioduszewska, and A Kacelnik. Compound tool construction by new caledonian crows. *Scientific reports*, 8(1):15676, 2018. 1

[2] Teresa McCormack, Christoph Hoerl, and Stephen Butterfill. *Tool use and causal cognition*. Oxford University Press, 2011. 3

[3] Huaqing Min, Ronghua Luo, Jinhui Zhu, Sheng Bi, et al. Affordance research in developmental robotics: A survey. *IEEE Transactions on Cognitive and Developmental Systems*, 8(4):237–255, 2016. 3

[4] Luis Montesano, Manuel Lopes, Alexandre Bernardino, and José Santos-Victor. Learning object affordances: from sensory–motor coordination to imitation. *IEEE Transactions on Robotics*, 24 (1):15–26, 2008. 3

[5] Lakshmi Nair, Jonathan Balloch, and Sonia Chernova. Tool macgyvering: Tool construction using geometric reasoning. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 5837–5843. IEEE, 2019. 4

[6] Lakshmi Nair, Nithin Shrivatsav Srikanth, Zackory M Erickson, and Sonia Chernova. Autonomous tool construction using part shape and attachment prediction. In *Robotics: Science and Systems*, 2019. 4

[7] Meiying Qin, Jake Brawer, and Brian Scassellati. Robot tool use: A survey. *Frontiers in Robotics and AI*, 9:1009488, 2023. 1, 2, 3

[8] Elisabetta Visalberghi, Gloria Sabbatini, Alex H Taylor, and Gavin R Hunt. Cognitive insights from tool use in nonhuman animals. 2017. 1

[9] Natsuki Yamanobe, Weiwei Wan, Ixchel G Ramirez-Alpizar, Damien Petit, Tokuo Tsuji, Shuichi Akizuki, Manabu Hashimoto, Kazuyuki Nagata, and Kensuke Harada. A brief review of affordance in robotic manipulation research. *Advanced Robotics*, 31(19-20):1086–1101, 2017. 3

[10] Zeyu Zhang, Ziyuan Jiao, Weiqi Wang, Yixin Zhu, Song-Chun Zhu, and Hangxin Liu. Understanding physical effects for effective tool-use. *IEEE Robotics and Automation Letters*, 7(4): 9469–9476, 2022. 3

[11] Yixin Zhu, Yibiao Zhao, and Song Chun Zhu. Understanding tools: Task-oriented object modeling, learning and recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2855–2864, 2015. 3

[12] Yixin Zhu, Tao Gao, Lifeng Fan, Siyuan Huang, Mark Edmonds, Hangxin Liu, Feng Gao, Chi Zhang, Siyuan Qi, Ying Nian Wu, et al. Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense. *Engineering*, 6(3):310–345, 2020. 1