Can robots use tools in a super system? No, but maybe in the future

Yao Yuanhang
Peking University
2000017813@stu.pku.edu.cn

Abstract

The use of tools was originally considered to be the ability that distinguishes humans from other animals. We aim to enable robots to use tools like humans, and even surpass human abilities in a reasonable manner. This is a highly challenging task. Robot tool use requires three skills: perception, manipulation, and high-level cognitive skills. According to reference literature, robot tool use can be categorized into two major classes: causal tool use and non-causal tool use, with several subcategories including single-tool manipulation and multi-tool manipulation. Each subcategory requires different skills and is highly complex. As for whether all the required skills can be integrated into a single super system, enabling our robots to have the flexibility to use various tools in complex real-world environments, this is the question we will explore next.

1 Introduction

Before delving deeper into the skills required for robots to use tools, it is essential to establish a precise definition of "robot tool use."[1] Drawing upon previous research in the academic community regarding "animal tool use" and considering the specific application environment for robots, the following definition can be derived: Robot tool use involves attaching or affixing an external, inanimate object that is freely manipulable on the robot's end effector or attaching it to another object with the purpose of achieving a change in the state of another object, altering the robot's own state, or accomplishing other goals through purposeful manipulation. This definition is highly accurate and rigorous.

First, we restrict the scope of tool use to external, manufactured, and inanimate objects. Unlike organisms, robots typically do not generate materials from their own bodies (such as feces or spiderwebs). We require that the tools are inanimate, as the most likely animate objects robots would manipulate are other robots. We believe this definition is more suitable for multi-agent systems rather than the field of tool use, as it involves synchronization and communication among robots. Second, we relax the interaction constraints between the tool and the manipulated object, including dynamic and static interactions. Therefore, repositioning another object using a container will be considered as tool use. Third, we loosen the constraints on the goals of tool use.

The causal tool use of robots that we mainly deal with can be categorized into two types: single-tool manipulation and multi-tool manipulation. Single-tool manipulation is further subdivided based on tool characteristics and prior tool use experience. This type of tool use can be considered as the building blocks of causal tool use. Multi-tool manipulation combines these building blocks in different ways and classifies different combinations as multi-tool manipulation. The contexts for causal tool use are diverse. General task complexity is used to categorize causal tool use into single-tool manipulation and multi-tool manipulation. The skills required for each sub-type are different.

2 three skills required for robot tool use

As mentioned earlier, robot tool use requires three fundamental skills (which are also essential functions that a potential super-system needs): perception, manipulation, and high-level cognition.

2.1 perception

When robots use tools, they need to possess various perception skills, including vision, touch, and hearing. Visual perception helps robots identify tools and the objects they are manipulating, determine their positions and orientations, and provide feedback for the robot's motion trajectory. Tactile perception provides information about contact forces, micro-vibrations, and heat flux resulting from contact with external objects, which is crucial for dexterous manipulation. Auditory perception assists robots in recognizing sound signals, such as alarm sounds or command voices, as well as detecting vibrations and noise from the tools or manipulated objects. Additionally, robots also need spatial perception skills to accurately locate themselves and the objects they are manipulating when using tools. These perception skills are implemented through sensors such as cameras, pressure sensors, accelerometers, etc. Furthermore, robots require algorithms and models for processing and analyzing this perceptual information.

2.2 manipulation

The second skill involved in robot tool use is manipulation. Manipulation skills focus on how to achieve the required tool manipulation actions in terms of kinematics and dynamics to ensure that the robot can perform the desired tool manipulation actions accurately and efficiently.

Kinematic Control: Manipulation skills involve how the robot plans and controls its motion to position the tool at the desired location and orientation within its workspace. Kinematic control considers the geometric relationships between the robot's joints and end effector to ensure the precision and accuracy of the actions. This includes inverse kinematics, which calculates joint trajectories from the target position and orientation.

Dynamic Control: Manipulation skills also encompass how to manage the forces and torques applied by the robot to perform the required actions. This includes considering the robot's load, inertia, and dynamic responses to ensure smooth, safe, and efficient movements. Dynamic control is critical for handling different workloads and tool characteristics.

Trajectory Planning: To achieve specific tool actions, manipulation skills involve planning the robot's trajectory to ensure that the tool moves along the desired path. Trajectory planning can be achieved through interpolation, spline curves, or other motion planning techniques to ensure smooth movement.

Force Control: In some cases, manipulation skills may require the robot to adjust its actions based on the contact forces with the tool. This may involve precise control of the forces applied by the tool to avoid damaging objects or to adapt to different workloads.

In summary, manipulation skills encompass how the robot performs actual tool manipulation actions, including aspects of kinematics, dynamics, trajectory planning, and force control. These skills are crucial for ensuring that the robot can interact with tools accurately and efficiently to perform the required tasks.

Additionally, there is a supplementary point to consider about object affordance in the scene[3]. Object affordance in the scene often depends on its geometric and physical properties. For instance, a tool with a handle part with higher friction is easier to grip, while the tool's front end is harder for exerting force. Robot manipulation processes need to take into account the affordance of objects in the scene to perform rational actions.

2.3 high-level cognition

The third skill we mentioned is high-level cognitive skills. This includes reasoning and planning tool use operations in a given task and available tools. For example, the robot may need to reason how to grasp a tool for convenient use, determine tool manipulation, contact posture, trajectory, and the required forces for successful tool use. When the learned tools are not suitable, the robot should also consider using new tools. Additionally, the robot should plan the use of multiple tools to achieve a specific goal. Neurological evidence also supports that when humans use tools, different cognitive processes are involved compared to separate hand and tool actions. It's worth noting that this part is the most crucial and challenging aspect.

3 Evaluate the robot's performance in using tools

To evaluate the performance of an intelligent agent in tool use, various methods and metrics can be employed, which can be customized based on specific task and application requirements. Here are some common evaluation methods and metrics:

Task Completion Time: An obvious measure is the time it takes for the agent to complete a task. Shorter times typically indicate more efficient tool use.

Task Success Rate: The ratio of successful task completions to the total number of attempts. This can be used to evaluate the agent's accuracy and reliability.

Accuracy: When the agent uses tools to perform tasks, the accuracy of its actions can be evaluated, i.e., how closely they align with the ideal outcome. This can be measured using distance, angles, or other appropriate metrics.

Obstacle Avoidance Capability: If the agent needs to navigate around obstacles while operating tools, its obstacle avoidance ability can be assessed, i.e., whether the agent can safely maneuver around obstacles during task execution.

Resource Utilization: Evaluate the resources used by the agent, such as energy, materials, or tool wastage. An economically efficient agent minimizes resource waste while completing tasks.

Learning Capability: The agent's ability to learn is a critical metric. It assesses the agent's adaptability in tool use, whether it can learn from mistakes, adapt to new tasks or tools, and continuously improve its performance.

Economical Efficiency: Economical efficiency evaluates the cost-effectiveness of the agent's tool use, including its operational and maintenance costs, as well as the benefits associated with task completion.

Scalability: If the agent needs to perform various types of tasks or use different types of tools, assess its performance and adaptability in different contexts.

Human Assessment: Finally, performance can be evaluated through expert assessments or user feedback. This can help identify potential issues and improvement points in the agent's tool use.

These evaluation methods and metrics help determine how well the intelligent agent performs in using tools and provide insights into its efficiency, effectiveness, and adaptability in different scenarios.

4 Can we build a super system?

Integrating perception, manipulation, and high-level cognitive skills into a super-system is one of the significant goals in robotics and artificial intelligence research. This integration can make robots more intelligent and autonomous, enabling them to better handle complex tasks and diverse environments. The super-system that combines these skills is often referred to as an "intelligent control system" or "cognitive architecture," and it comprises the following three main components[2]:

Perception Module: This module is responsible for gathering information from the environment. It includes various sensors such as cameras, LiDAR, tactile sensors, etc., to perceive object positions, shapes, colors, textures, and more. The perception module transforms this information into data that the robot can understand, providing the robot with environmental awareness.

Manipulation Module: The manipulation module processes the perceived information and executes actions based on task requirements. This includes techniques such as dynamic control, trajectory planning, and force control to ensure that the robot can accurately use tools, move objects, and perform actions. The manipulation module translates high-level commands into specific actions.

High-Level Cognitive Module: The high-level cognitive module serves as the robot's brain, handling decision-making, planning, and task execution. This module includes machine learning, planning algorithms, and artificial intelligence technologies to help the robot understand task requirements, make decisions, plan paths, and solve problems. The high-level cognitive module is responsible for formulating the robot's long-term and short-term strategies to accomplish tasks.

The coordination and communication between these modules are critical to integrating perception, manipulation, and high-level cognitive skills. By integrating these skills into one system, robots can better understand and adapt to complex tasks and environments, enhancing their autonomy and intelligence.

However, this is a vision for a future super-system that is not fully achievable with current technology. There are a series of challenges that need to be addressed. In practice, robots may need to collaborate with human partners to solve tasks. Considerations about the safe handling of tools and ways to facilitate effective cooperation are essential. Additionally, the real world is much more complex than controlled experimental environments, with various sources of noise, adding significant external complexity to the system.

In the video provided, the crow population exhibits the ability to learn (often through parent-child transmission). Currently, our robots do not possess the ability to learn and transmit tool use understanding among intelligent agents like crows do. It's clear that there is still some way to go before achieving a true super-system, but it's believed that these challenges will be addressed in the near future.

References

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

[2] Y. Wang and X. Li. A survey of artificial intelligence super-integration system. *Journal of Intelligent & Fuzzy Systems*, 42(4):1–12, 2022. 3

[3] Chao Xu, Yixin Chen, He Wang, Song-Chun Zhu, Yixin Zhu, and Siyuan Huang. Partafford: Part-level affordance discovery from 3d objects. *arXiv preprint arXiv:2202.13519*, 2022. 2