Progress Report: Enhancing Robotic Construction Intelligence through Tree-of-Thoughts Reasoning

Student: Yizhi Liu
Email: yl4993@columbia.edu

Project Overview:

In recent years, the research to endow robots with higher-level perception abilities has become a main trend in robotics field. For construction robots specifically, one challenge lies in enabling these robotic systems to understand entire task operation processes rather than just isolated actions (e.g., pick-and-place). Construction tasks typically involve multiple connected steps; for instance, when laying bricks, a robot should not only pick and place each brick but also align it, apply mortar, and move to the next layer. This multi-stage workflow requires the robot to recognize the state of the project and integrate knowledge from building codes or real-time sensor data – requirements that remain hard to achieve with current development of robotic perception. To address this gap, the proposed project will explore how advanced reasoning capabilities of LLMs can enable robots to gain a high-level understanding of complex, multi-step tasks in construction.

Building on progress made in this semester, I will focus on the Tree-of-Thoughts (ToT) framework, one of the core reasoning structures from LLM, to enhance robotic perception of complex, multi-step tasks in construction. As introduced in the mid-term project, ToT, which originates from cognitive science and has recently been adapted for use in large language models, provides a structured paradigm for multi-step reasoning. ToT decomposes problems into branching sequences of "thoughts," which serve as nodes in a tree-like structure. For construction robotics, this decomposition strategy closely mirrors how human experts break down tasks; for instance, construction engineers and project managers usually divide complex workflows into discrete yet interconnected steps. Moreover, ToT can enable robots to operate systematically: a robot might begin laying bricks in a certain pattern, discover that a subsequent step is impeded by environmental constraints, and then revert to a previous "thought" in the tree to re-evaluate its approach.

Moving forward, this project aims to (1) explore the feasibility of implementing ToT approach to interpret construction tasks (e.g., material handling, structural assembly, site surveying), (2) integrate the ToT approach into robotic operation system, and (3) test the approach in simulated environments. Overall, By combining the structured, iterative reasoning of ToT with robotic perception system, we plan to enhance the efficiency of robotic-assisted construction.

Research Questions:

This project aims to explore how large language models (LLMs) can enhance robotic understanding and execution of multi-step construction tasks by addressing three research questions: First, this project plans to explore whether existing LLMs are capable of accurately interpreting and reasoning through multi-step construction tasks, which often require understanding procedural sequences, task dependencies, and contextual information. Second, the project explores how LLMs can be integrated into robotic systems to provide human-like reasoning capabilities, enabling robots not only to understand assigned tasks but also to make accurate inferences and decisions similar to those made by human workers. Third, the project considers how a perception-and-action framework can be developed to enable robots to fully understand, decompose, and perform the construction tasks in real-world environments.

Please note that while these three questions collectively define the broader scope of this research topic, this particular project - given its constraints in terms of time and computational resources - will primarily focus on the first two questions.

Value to User Community:

The potential users of this project include construction professionals, robotics engineers, human-robot interaction researchers focused on construction automation. This LLM-based interface has the potential to offer value by enabling non-expert users – particularly construction workers without prior robotics training – to directly assign tasks to construction robotics (robotic arm). In other words, to enable robots to perform the construction tasks, there is no need to train workers to control and collaborate with robots. This training process will be time-consuming and may be impossible due to the lack of enough knowledge. By leveraging large language models (LLMs) and integrating techniques like Tree-of-Thought (ToT) prompting, the robot may have the potential to have human-like perception capability, making robotic systems to understand and perform the task by itself. For researchers and developers, the system serves as a valuable testbed for evaluating LLMs in multi-step physical task settings, contributing new insights into prompt engineering, task decomposition, and adaptive motion planning.

In sum, the value of this project (1) will open a door for robotic technology in the traditional construction industry, and (2) researchers and developers with a novel opportunity to explore the application of large language models (LLMs) in construction robotics—an area that has remained largely underexplored to date.

Demo:

If possible, I plan to prepare a poster that illustrates the full narrative of the project – from the development of the LLM component (as presented in the midterm) to the integration of LLM-driven robotic perception (as targeted in final). The poster will provide an overview of the project's background, research questions, methodology, and key findings, helping the audience clearly understand the overall scope and significance of the work. If the poster plan is not suitable, for the 5-minute demo, I will show the experimental results, demonstrating how the LLM decomposes high-level construction instructions into a sequence of interrelated sub-tasks. I will also present how these sub-tasks are translated into robot actions within a simulated construction environment.

If the second step (robot execution in simulation) is not yet fully implemented, I will focus on presenting the first step in detail.

Delivery:

I will upload the code to GitHub. The code will have two main parts, one for LLM to decompose construction tasks into sub-level tasks; and the second part will show the code of transferring the LLM outputs into the robotic operating system to control robots.

Is there anything else I need to include?

To evaluate the performance of my project, I also plan to include a public benchmark for robotic manipulation. I plan to use RoboBench (https://project-roco.github.io/), which provides standardized tasks and evaluation protocols for assessing robotic manipulation systems.