LLAMA-ACT-R, a Neuro-Symbolic Architecture (ACT-R) for LLM Decision Making

Siyu Wu, C Lee Giles, Frank E. Ritter

17mar24

Objectives

Background

Currently, LLMs learn from very large amounts of data to make predictions and decisions. But they tend to use shortcuts and cannot remember past interactions, which limits their ability to improve over time unlike humans. They also have a difficulty with tasks that require deeper thought or learning from experience.

Introduction

Create LLMs that can make decisions and learn from experiences much like people do. Right now, they are good at quick thinking but not so much at the careful, thoughtful decision-making that humans do, especially when facing new problems. We want to improve this by blending the way LLMs make decisions with methods inspired by computational modelling for unified theory of mind. To do this, we design Llama-ACT-R. We choose Llama for this research because it provide researchers with complete access to the network architecture including its pre-trained weights.

AI Goal

This work relates to a big goal in AI, which is to make LLMs that can understand and interact with the world in a more human way. The state of the art in AI has given us programs that can output in ways that seem very human, but these programs often don't have complex reasoning or learn from their actions. By combining the current capabilities of these AI programs with models that emulate human thought processes such as ACT-R (Ritter et al. 2019), our model should enable LLMs to better understand, represent, and learn from the world.

Related Work and Technical Approach

Related Work

This work builds upon the disparity between LLMs and human decision-making, noting LLMs' focus on rapid, intuitive processes and their limitations in complex reasoning and continuous learning. To bridge this gap, we propose integrating LLMs with the ACT-R cognitive architecture, a framework that models human cognitive processes.

This approach has been pioneered by Binz & Schulz (2023). They fine-tuned the last layer of Llama embeddings using data from past human psychological studies. Their work confirmed that fine-tuned language models exhibit more human-like behavior. However, this approach faces limitations in data collection, expansion, and exploitation costs, making it difficult to scale.

Technical Approach Overview

We overcome the bottleneck of data scarcity by using a well-established neural-symbolic AI representation, the ACT-R cognitive architecture to make fine tuned hybrid agents that replicate a human decision making task. This integration will demonstrate how to enhance LLMs with human-like decision-making and learning patterns by aligning ACT-R's decision-making processes with LLM's internal representations. Our architecture has the potential to create large-scale studies that enable LLMs to make decisions and learn in ways that more closely mirror human cognition, addressing the critical challenge of aligning machine reasoning with human processes.

Technical Approach

The Task

Observations revealed behavior disparities between ChatGPT 3.5 and humans in the Building Sticks Task, highlighting two discrepancies: strategy and learning curve. While initially adopting the optimal overshoot strategy, ChatGPT lacked the human-like blend of exploratory and exploitative strategies and did not show a learning curve, quickly reverting to less efficient strategies without improvement.

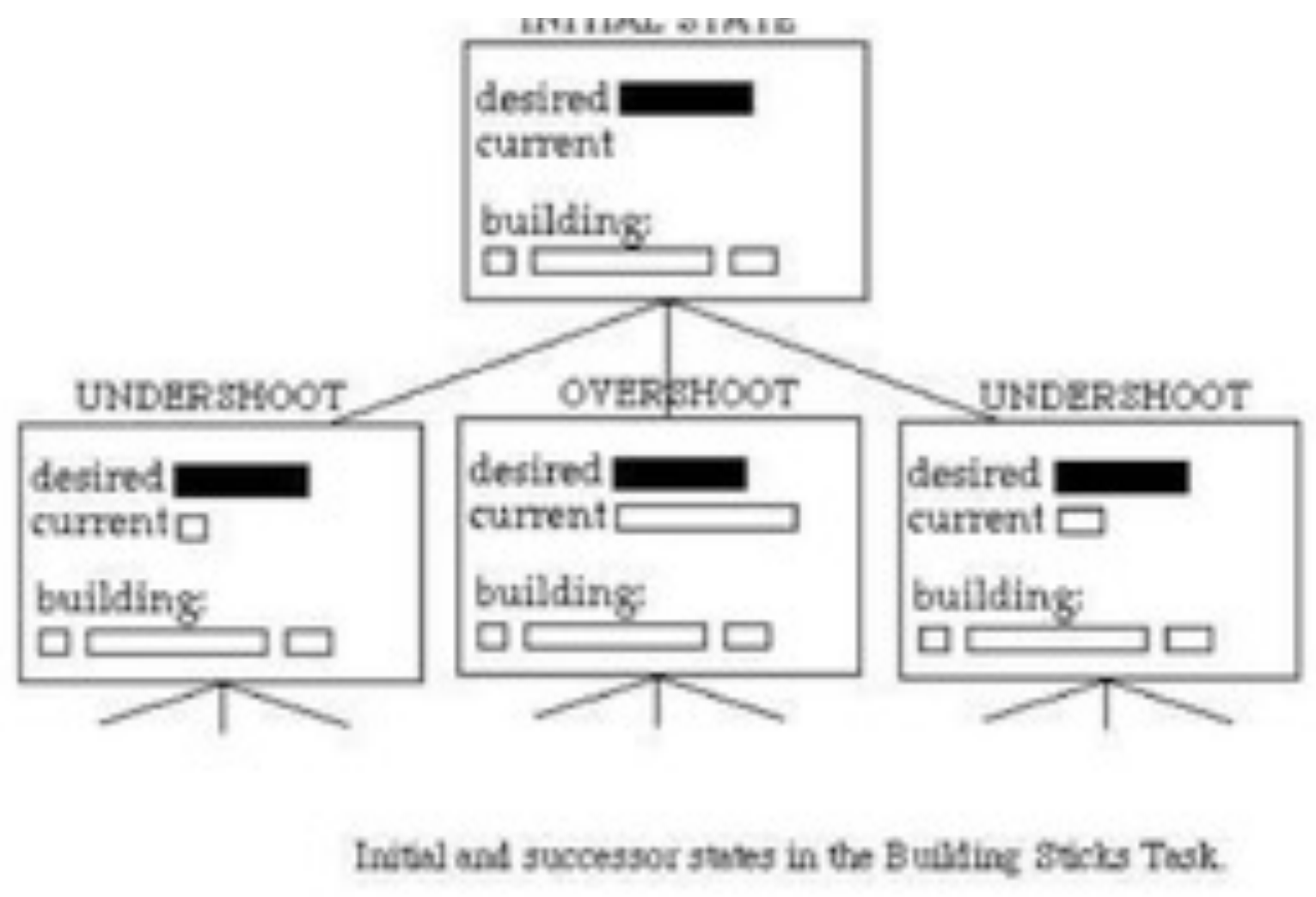

Fig. 1. The sticks choosing task, taken from Lovett (1998).

The Model

A fine-tuned ACT-R model of this task is more closely aligned with human decision-making and learning patterns, and demonstrated its effectiveness in the Building Sticks Task. Through trials, the model showcased a balanced use of explorative and exploitative strategies, improving its use of the overshoot strategy from 20% to 53%.

Experimental Setup

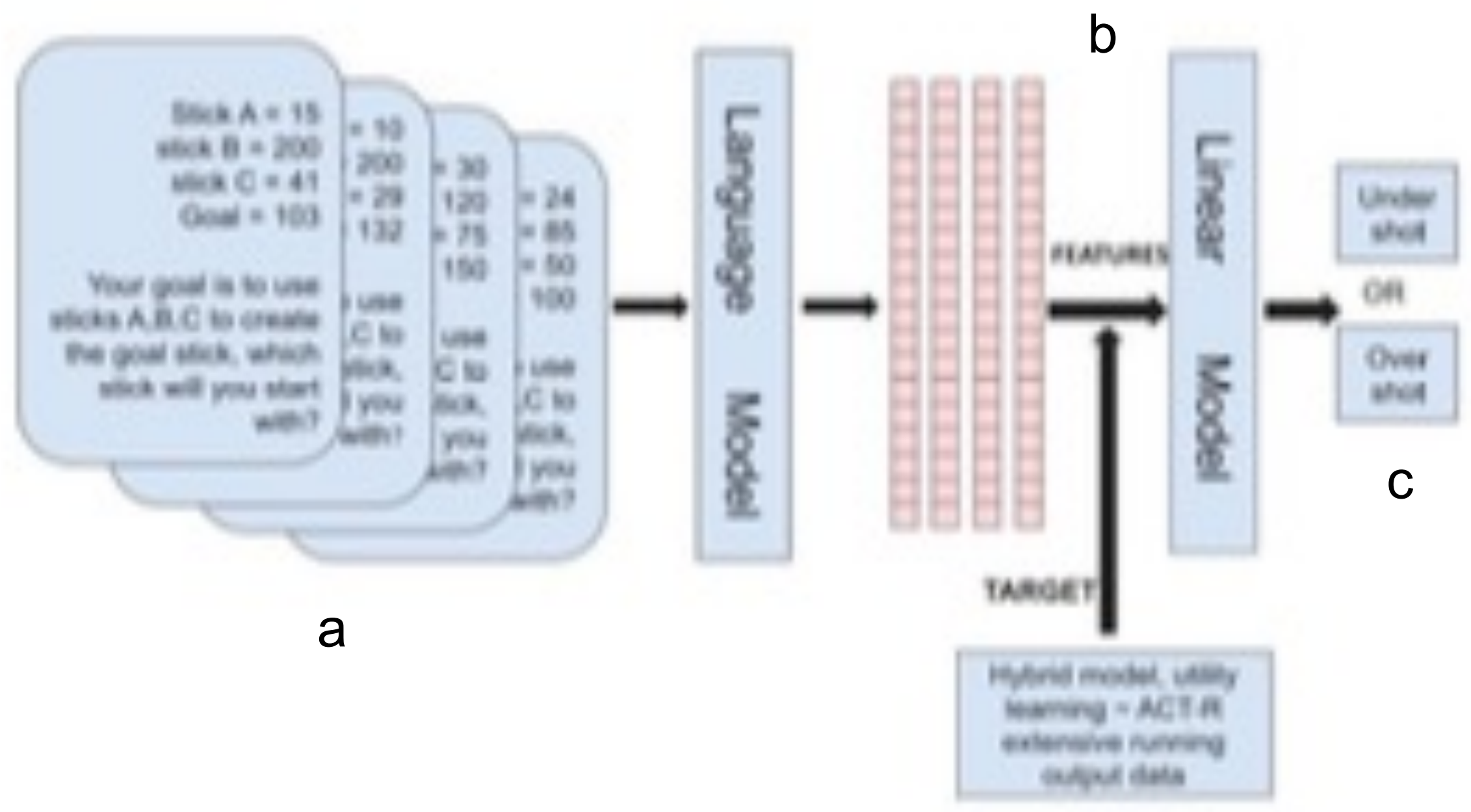
b
Stick A = 15
stick B = 200
stick C = 41
Goal = 103
Your goal is to use sticks A,B,C to create the goal stick, which stick will you start with?
Language Model
FEATURES
Linear Model
Under shot
OR
Over shot
c
a
TARGET
Hybrid model, utility learning - ACT-R extensive running output data

Figure 2. Schematic of LLAMA-ACT-R.

ACT-R Model Setup

1. Use ACT-R to symbolize domain knowledge for decision-making, create, and finetue ACT-R synthetic agent in stick task.
2. Create a dataset from 10,000 problem-solving iterations by running the ACT-R model.

LLAMA-ACT-R System

As shown in Figure 2, (a) provide a text-based description of the building sticks experiment to Llama and extracted the last layer of resulting embeddings. (b) fine-tune a linear layer (a logistic regression model) using embeddings (variables and weights) and targets (ACT-R data) (c) on top of these embeddings predict ACT-R decision-making choices.

Results and Comparison to the State of the Art

Comparison to the State of the Art

Cognitive architectures (CAs) require extensive pre-defined rules and reliance on manual rule input, reducing their usability. Conversely, Large Language Models (LLMs) excel in adaptability, learning rules from vast text and other data, bypassing the need for manual input. To exploit their advantages, integration of CAs with LLMs is being explored to form a more cohesive computational model, harnessing LLMs' implicit knowledge and CAs' symbolic methods. This research proposal is unique in reversing the integration direction, applying ACT-R's cognitive framework within LLMs to provide structure and clarity to LLM reasoning, enhancing their decision-making transparency and explainability, a unique approach not yet explored in current studies.

Results

The representations that result in integrating ACT-R will help both transparency and intelligence within the language model's reasoning chain.

LLAMA-ACT-R System will support generating and explaining LLMs procedural knowledge. It will also help model and understand human cognition. It may also provide a way to insert values and goals into such hybrid systems.

Key Accomplishments, Lessons Learned, and Next Steps

Key Accomplishments

Developed a framework that integrates the ACT-R cognitive architecture with LLMs to enhance machine decision-making, allowing LLMs to exhibit more human-like reasoning patterns and learning curves.

Demonstrated through the Building Sticks Task that the fine-tuned ACT-R model aligned closely with human decision-making, has the potential to improve the LLM's strategy selection and adaptability in learning.

A demo with the architecture of Building Sticks Task showing the Ilama-ACT-R hybrid is in progress.

Lesson Learned

A critical lesson learned is the importance of human-like learning curves and balanced strategy application in AI decision-making, which can be achieved by integrating cognitive architectures like ACT-R.

Address the limitations of LLMs, such as their issues with concepts of state and their tendency to favor exploitative strategies over exploratory ones.

Next Steps

Next, refine the Llama-ACT-R hybrid model further, potentially expanding the scope of tasks it can learn from and adapt to, as well as enhancing its ability to generalize this learning to new, unseen problems. Additionally, address any computational inefficiencies or integrating feedback to improve the model's functionality.

Acknowledgements: We thank Dan Bothell for assistance with the ACT-R model and Marcel Binz for encouraging this idea, based on his previous work on fine-tuning language models to mend behavior disparity, and META for making language models open accessible.

References: Binz, M., & Schulz, E. (2023). Turning large language models into cognitive models. arXiv preprint arXiv:2306.03917.

Lovett, M. C. (1998). Choice. In J. R. Anderson & C. Lebiere (Eds.), The atomic components of thought (pp. 255-296). Mahwah, NJ: Erlbaum.

Wu, S., Ferreira, R., Ritter, F. E., Walter., L. (2024) Comparing LLMs for prompt-enhanced ACT-R and Soar model development: A case study in cognitive simulation. *Proceedings of 38th Annual Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence Fall Symposium Series on Integrating Cognitive Architecture and Generative Models* at Arlington, Virginia, USA. DOI: https://doi.org/10.1609/aaaiss.v2i1.27710

Ritter, F. E., Tehranchi, F., & Oury, J. D. (2019). ACT-R: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 10*, 833-838

PennState

AI Week | April 2024

AI hub