Computational Models of Humans for Studying and Improving Human-AI Interaction

Nicholas Bindela

Kalman Research Symposium

Department of Computer Science, Bucknell University, Lewisburg, Pa.

Abstract

My research is centered around Microsoft's "Project Malmo" and their "Collaborative AI Challenge", which was a competition that had contestants aim to develop and train AI to work with humans to accomplish a common goal. My research aims to develop cognitive agents that can effectively simulate human cognition when interacting with the AI agents from the Project Malmo competition. Project Malmo uses the Minecraft client to build the environments in which the AI simulations run. Our research team uses Python to program the environments where the AI and human/cognitive agents will interact. The Python environments behave as a mediator between the Minecraft client and the software we use to program our cognitive agents. My portion of the research project is aimed at developing the cognitive agents to interact with the AI. We use the Adaptive Control of Thought-Rational (ACT-R) cognitive architectural model as a basis for creating our cognitive agents. ACT-R is a psychological theory that allows us to break down human actions and behaviors into discrete operations so we can effectively simulate human cognition. The ACT-R model has a software available online that we use to program our models so they can interact with the AI solutions. I work on creating the cognitive models using the ACT-R software, which is programmed using Common-Lisp. I also work on building the Python environments, sending observations made by the ACT-R agent to Python and making sure the information is properly received in Python.

Introduction

We plan to develop simulations that will give rise to specific stimuli in humans when collaborating with artificial intelligence agents to understand human-AI interaction. Our goal is to analyze how AI algorithms can be advanced to respond to variations in behavior of humans when put under different levels of arousal. We hope to develop simulations that will help us determine how stimulus from the environment affects human-AI interaction, as well as determining how to increase cooperation between human and AI agents through more advanced algorithms. We will use Adaptive Control of Thought-Rational (ACT-R) architecture to develop cognitive agents to interact with the environment and AI. Using the cognitive agents we can see how stimulus from the environment and behavior of AI will change performance of the agent. [2]

Adaptive Control of Thought-Rational

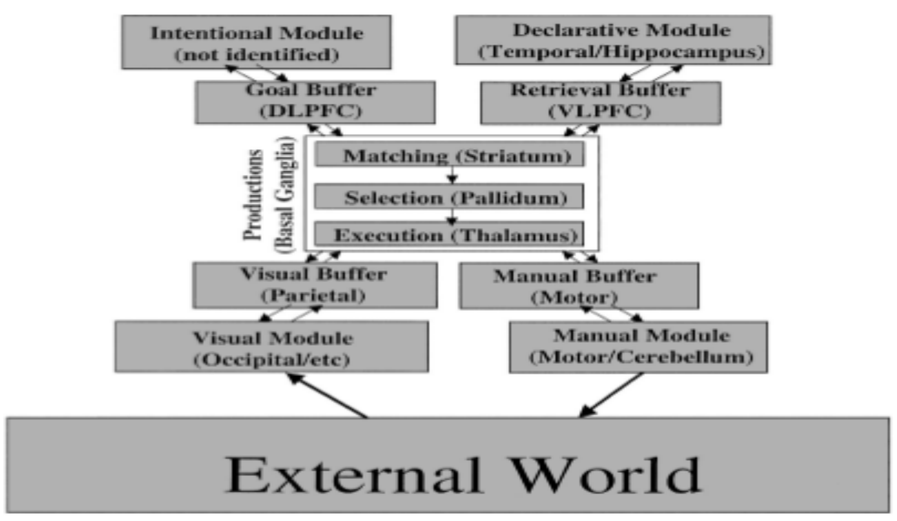
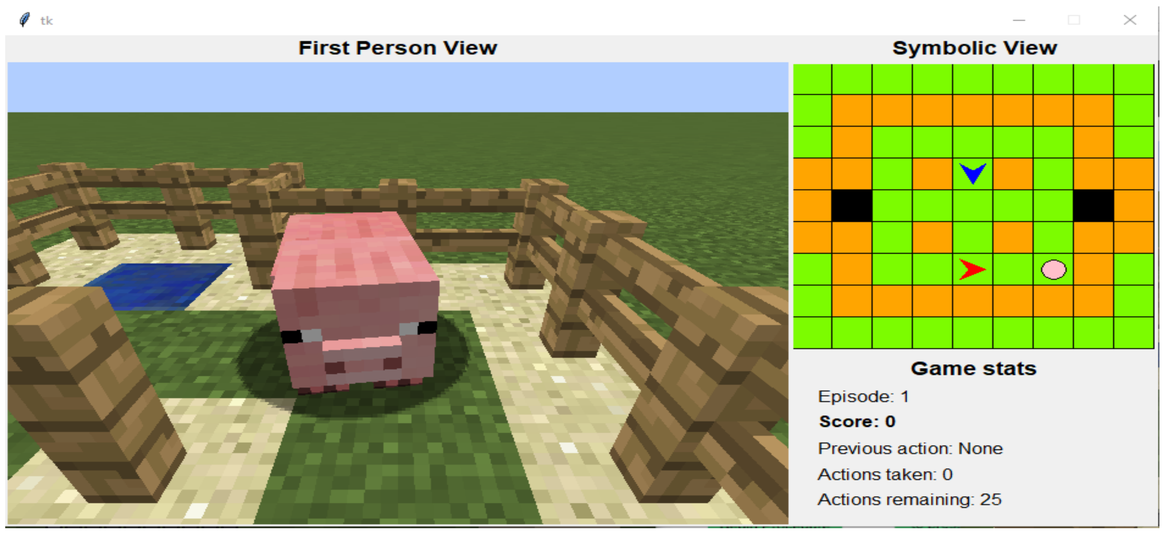


Figure 1. The organization of information in ACT–R 5.0. Information in the buffers associated with modules is responded to and changed by production rules. DLPFC = dorsolateral prefrontal cortex; VLPFC = ventrolateral prefrontal cortex.

Adaptive Control of Thought-Rational(ACT-R) is a psychological model of the human brain's cognitive processes. ACT-R is a cognitive architecture that defines basic operations that define the human mind. The model breaks down human actions and decisions into discrete atomic cognitive operations and perceptions. We use an ACT-R software to test the models that we design. The software is made of different modules(see image above) that represent different cognitive functions. Models we design with the ACT-R software can make decisions and perform actions based on observations it makes about the outside world. The figure above shows the different modules of the ACT-R architecture and how they interact.[1]

Project Malmo



Project Malmo is an artificial intelligence experimentation platform developed by Microsoft. The platform is built on top of Minecraft. We use the Minecraft client to run our experiments. The 'Malmo Collaborative AI Challenge' was held by Microsoft and had contestants build AI to interact with humans to complete a common goal. The goal was to have the two agents(human and AI) try to trap a pig in as few moves as possible. The environment we use can be seen above.

We program our environments using Python and XML documents. We use the Malmo module in Python and run experiments using the pig chase environment. Our goal is to use the winners from the Collaborative AI Challenge and agents we design using ACT-R to interact in the pig chase environment without the intervention of humans.

Implementation

We have done a lot of work that has moved us closer to our goal of developing agents that will react to their environment and the behavior of other agents. A large part of our implementation is getting the environments and different software we are using to interact properly. We must ensure that the ACT-R software is able to interact with the Python environments we create, and they can both send and receive information from the other. Our experiment requires that our Python program is able to perceive the Minecraft world around it and send observations about the environment to the ACT-R software, where the agent can use the information to make a decision and return information about its decision back to Python so that the environment can react accordingly. I currently have Python and ACT-R interacting efficiently and our next step is to develop algorithms that will have our ACT-R agent make reactionary decisions based on the state of the environment.

References

[1] Anderson, John R., et al. "An Integrated Theory of the Mind." *Psychological Review*, vol. 111, no. 4, 2004, pp. 1036–1060., doi:10.1037/0033-295x.111.4.1036.

[2] Glinert, Ephraim. "Computational Models of Humans for Studying and Improving Human-AI Interaction." *NSF Award Search: Award#1849869 - CRII: CHS: RUI: Computational Models of Humans for Studying and Improving Human-AI Interaction*, 15 May 2019, www.nsf.gov/awardsearch/showAward?AWD_ID=1849869.