Jacob Devasier

Review of Dr Park's Presentation on Human-Level Artificial Intelligence

Dr Park's presentation on HLAI summarizes most of what we have learned throughout the semester followed by an explanation of his work and his plans for the future of his work. The presentation begins by laying out the goal of creating a robot that acts as a butler; the creation of which requires 3 main parts, the theory, the test, and the primary part, the model. Following the theory section, Dr Park explains the need for a sufficient but easy test for human-level AI and points to the existence of many easy but insufficient and sufficient but hard tasks as a primary reason for such demand. Dr Park proposes a language acquisition test for HLAI dependent on the agent's ability to learn language in some environment.

As mentioned in the presentation, humans cannot be used for training due to time and monetary constraints in addition to being hard to reproduce. Following this is an introduction to SEDRo which aims to solve problems introduced in the grounded language acquisition paper, primarily the reward-based learning system. The reasoning for this is well thought out as training an agent to focus on rewards only creates a greedy agent.

Following the task section, the presentation moved into the model section, the meat and most important section of the presentation. As mentioned in the presentation, it is imperative that the model works autoregressively. The important part of this fact is that the available model architectures are quite limited and requires sequential models. The architecture of the model proposed begins with an encoder which pools inputs 4:1, essentially

acting as a convolution of the inputs. This idea is promising as, much like in computer vision, individual pixel-level information is less important than pixel-group information, e.g., seeing a patch of 9 pixels at once will be more useful than seeing 9 individual pixels.

The model also uses a layered hierarchical structure which feeds the output of the aforementioned encoded values to a higher-level encoder, thus looking at groups of groups of input signals much like convolutional layers in a convolutional neural network. In addition to this, the higher layers append the outputs to lower-level inputs. Effectively allowing lower-level agents to see higher-level information and converting the model to be heterarchical. Overall, I think the model section looks exciting and I look forward to seeing more from it in particular.