The student becomes the master: what can recent advances in artificial intelligence teach us about our own learning?

Jake Topping

When learning by trial-and-error, the way our AI views its environment evolves over the course of its training but in a different way to humans.

Experiments suggest that replacing some of an AI's memories with real human experiences can make this evolution more similar to that of humans, without unacceptable costs in final performance and training time.

This may mean we can use an AI trained in this way to assist or accelerate human learning on a given task by isolating the most useful situations to learn.

"Learning"

In decision making – artificial or human – we can consider notions of policy and value.

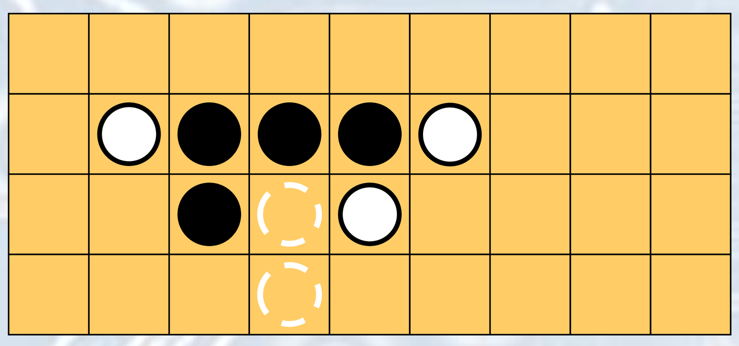
- **Policy** is how good you consider each choice you have in this state to be. For instance at a given point while driving, you might consider the merit of braking, accelerating or turning the wheel.
- **Value** is how good you consider a position or environment itself to be. For example while driving, being on the right side of the road should be more highly valued than on the wrong side!

How you value choices and positions evolves as you learn a task.

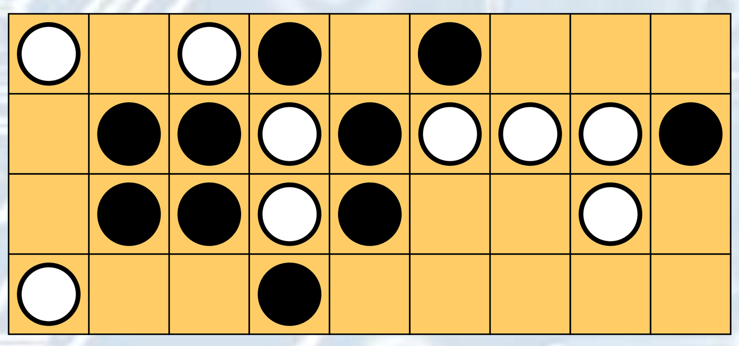
The task: 4-in-a-row

In 2017, van Ophesuden et al at the Wei Ji Ma lab, NYU, conducted a study on humans learning a game called 4-in-a-row to infer how people plan ahead in sequential decision making (example positions can be seen below). Participants came in over a few weeks where they learned the game by playing each other, and were also trialled in the following experiments:

2AFC (Policy):
which of the highlighted moves should white make next?



Evaluation (Value):
on a scale of 1 to 7, how good is this position for white given it is their move?

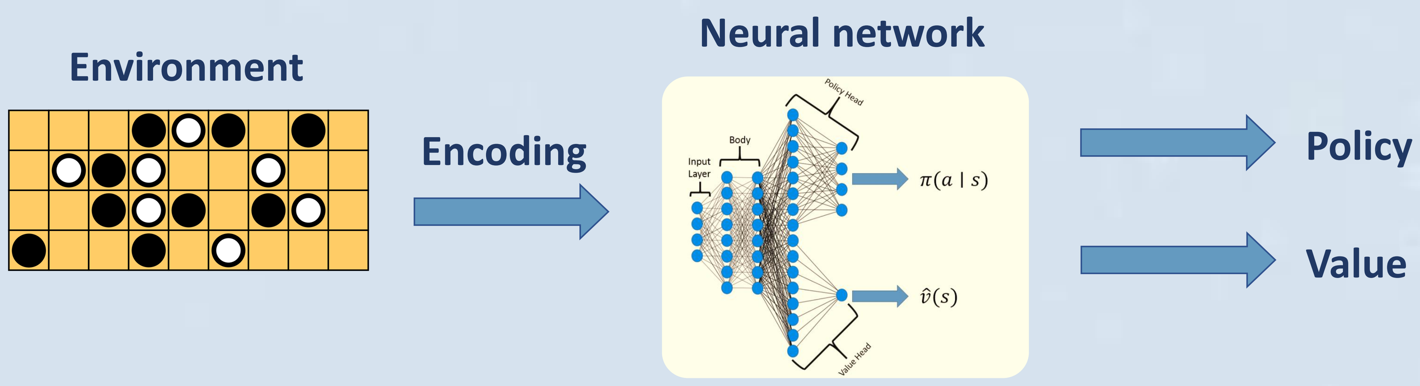


Reinforcement learning

The 'artificial intelligences' we considered were agents trained to play 4-in-a-row using reinforcement learning. This meant they:

- Started with no knowledge of the game apart from the rules
- Only learned through playing the game against itself
- Played a batch of games, then looked back at a selection of its 'memories' from these games to update its policy and value

The human equivalent of this is trial and error, and this resemblance to a style of human behaviour is part of the reason for our hypothesis that the way such an agent learns may resemble that of a human.



Methods

The first stage of the project involved implementing a neural-network based reinforcement learning agent (AI) capable of learning 4-in-a-row, and tuning it to reliably train to at least human standard. We achieved this, and further refinements are still being made.

We then repeated the experiments from the '17 study on each successful AI agent. Then the agent's behaviour was modelled using the same computational model (a linear model of value shown below) as for humans, such that the parameters could be compared. This allowed us to quantitatively measure how the AI's notion of value changes over learning, which can be compared to same results from the human data.

$$V(s) = c_{\text{self}} \sum_{i=0}^{4} w_i f_i(s, \text{self}) - c_{\text{opponent}} \sum_{i=0}^{4} w_i f_i(s, \text{opponent})$$

Later in the project we repeated this process with an adjusted AI that at each stage of learning had some of its memories replaced with real human data, with the hypothesis that its parameters over learning could be interpreted as closer to those from human play.

Results and Conclusions

The experiments described, both with the original AI and 'human-handicapped' AI, were performed several times. With the original AI, we found that the parameters at each stage had high variance between agents, and could not be said to resemble the human-fitted parameters. With the human-handicapped AI the parameters had a lower variance around a mean that was visibly closer to the human-fitted parameters.

This suggests that while a neural network's interpretation of a task is not the same as a human's when left to learn on its own, cooperating with it and providing it some human experiences as its own can push this interpretation to one we can compare to humans.

Applications in Assisted Learning

While interpretation of model parameters and learning is a big problem in the field of artificial intelligence, the mission statement of this project was to consider how studying an AI can benefit a human learning a task.

As part of a more ongoing collaboration with the Wei Ji Ma lab, we will be running further experiments on the human-handicapped AI to use the mathematical features of the model to find positions of the task where the AI's choices differ most between skill levels. If we can find such positions, that suggests that between these points of learning the AI has learned something. Once these have been isolated, they can be given to a human learning the task to accelerate moving up through skill levels. While this task is just a moderately-challenging game, it acts as a proof of concept for the many other domains that reinforcement learning has recently had success in.

I would like to thank the Wei Ji Ma lab at New York University for hosting me and collaborating with me for this project, Prof. Wei Ji Ma for his supervision and friendship over the summer, and the Laidlaw Scholars Programme at the University of Oxford for their support in making this project possible.