Self-taught AI is best yet at strategy game Go

Artificial-intelligence program AlphaGo Zero trained in just days, without any human input.

Elizabeth Gibney

18 October 2017

AlphaGo Zero came up with Go strategies that human players haven't invented in thousands of years.

An artificial intelligence (AI) program from Google-owned company DeepMind has reached superhuman level at the strategy game Go — without learning from any human moves.

This ability to self-train without human input is a crucial step towards the dream of creating a general AI that can tackle any task. In the nearer-term, though, it could enable programs to take on scientific challenges such as protein folding or materials research, said DeepMind chief executive Demis

Hassabis at a press briefing. "We're quite excited because we think this is now good enough to make some real progress on some real problems."

Previous Go-playing computers developed by DeepMind, which is based in London, began by training on more than 100,000 human games played by experts. The latest program, known as AlphaGo Zero, instead starts from scratch using random moves, and learns by playing against itself. After 40 days of training and 30 million games, the AI was able to beat the world's previous best 'player' — another DeepMind AI known as AlphaGo Master. The results are published today in *Nature*[1], with an accompanying commentary[2].

Getting this technique, known as reinforcement learning, to work well is difficult and resource-intensive, says Oren Etzioni, chief executive of the Allen Institute for Artificial Intelligence in Seattle, Washington. That the team could build such an algorithm that surpassed previous versions using less training time and computer power "is nothing short of amazing", he adds.

Related stories

- Google reveals secret test of AI bot to beat top Go players
- Google's AI reasons its way around the London Underground
- What Google's winning Go algorithm will do next

More related stories

Related stories

- Google reveals secret test of AI bot to beat top Go players
- Google's AI reasons its way around the London Underground
- What Google's winning Go algorithm will do next

More related stories

Strategy supremo

The ancient Chinese game of Go involves placing black and white stones on a board to control territory. Like its predecessors, AlphaGo Zero uses a deep neural network — a type of AI inspired by the structure of the brain — to learn abstract concepts from the boards. Told only the rules of the game, it learns by trial and error, feeding back information on what worked to improve itself after each game.

At first, AlphaGo Zero's learning mirrored that of human players. It started off trying greedily to capture stones, as beginners often do, but after three days it had mastered complex tactics used by human experts. "You see it rediscovering the thousands of years of human knowledge," said Hassabis. After 40 days, the program had found plays unknown to humans (see 'Discovering new knowledge').

Discovering New Knowledge
Deepmind

Approaches using purely reinforcement learning have struggled in AI because ability does not always progress consistently, said David Silver, a scientist at DeepMind who has been leading the development of AlphaGo, at the briefing. Bots often beat their predecessor, but forget how to beat earlier versions of themselves. This is the project's first "really stable, solid version of reinforcement learning, that's able to learn completely from scratch," he said.

AlphaGo Zero's predecessors used two separate neural networks: one to predict the probable best moves, and one to evaluate, out of those moves, which was most likely to win. To do the latter, they used 'roll outs' — playing multiple fast and randomized games to test possible outcomes. AlphaGo Zero, however, uses a single neural network. Instead of exploring possible outcomes from each position, it simply asks the network to predict a winner. This is like asking an expert to make a prediction, rather than relying on the games of 100 weak players, said Silver. "We'd much rather trust the predictions of that one strong expert."

Merging these functions into a single neural network made the algorithm both stronger and much more efficient, said Silver. It still required a huge amount of computing power — four of the specialized chips called tensor processing units, which Hassabis estimated to be US$25 million of hardware. But its predecessors used ten times that number. It also trained itself in days, rather than months. The implication is that "algorithms matter much more than either computing or data available", said Silver.

Think outside the board

Several DeepMind researchers have already moved from working on AlphaGo to applying similar techniques to practical applications, said Hassabis. One promising area, he suggested, is understanding how proteins fold, an essential tool for drug discovery.

Generating examples of protein folding can involve years of painstaking crystallography, so there are few data to learn from, and there are too many possible solutions to predict structures from amino-acid sequences using a brute-force search. The puzzle shares some key features with Go, however. Both involve well-known rules and have a well-described goal. In the longer term, such algorithms might be applied to similar tasks in quantum chemistry, materials design and robotics.

Silver acknowledged that to apply its approach to real-world tasks more generally, the AI will need the ability to learn from smaller amounts of data and experience. Another essential step will be learning the rules of a game for itself, as another DeepMind bot did in 2015 for arcade games. Hassabis reckons this is something AlphaGo Zero could eventually do: "We're pretty sure it would work, it would just extend the learning time a lot," he said.

References

1. Silver, D. *et al*. *Nature* **550**, 354–359 (2017).

2. Singh, S., Okun, A. & Jackson, A. *Nature* **550**, 336–337 (2017).