AlphaGo's Journey to Becoming Arguably the Best Player In History

By: Hien Le

Go is an ancient Chinese game that has players alternate making single moves at a time in order to try and establish territory. The game is typically played on a square grid of lines, typically 19 intersections to a side. Other common smaller sizes like 13 by 13 board and the 9 by 9 board. The board starts out empty and the two players have a set of either black or white pieces that are exclusive to that player for the rest of the game. Black makes the first move. After that, white and black alternate placing one piece on the board at a time. Any player can pass on their turn at any time, but two consecutive passes ends the game. A player is considered to have control of a certain area of the board when they have completely surrounded it. For example, in figure 1, white has completely surrounded that area so it belongs to white. At the end of the game, the player with the most area of the board captured is the winner. In figure 2, white has more area occupied so it wins the game. The paper "Mastering the game of Go without human knowledge" by Silver et al, describes the new techniques that the AlphaGo team implemented to create AlphaGo Zero. Which is a version that improves and dominates the previous iteration of AlphaGo that beat the eighteen time champion Lee Sedol.

In ancient times, this game was often a popular game amongst generals who would try to sharpen their strategic thinking and prove themselves worthy to the emperor. and people who want a challenge in thinking several moves ahead. In more recent history, it is enjoyed by many people in Asia as chess is enjoyed in more western parts of the world. You might ask yourself, how does this compare to chess in its complexity? The rules are basic enough, every piece serves the exact same function and behaves in the exact same way. But then you sit down and will probably start thinking to yourself: "well I know what I am *technically allowed* to do, but I have

no clue what I *should* do. As far as computers are concerned, go is much more difficult to tackle than chess. Compared to say chess, the number of possible moves in a position is much larger. In chess, it is approximately 20. In go, it is about 200, as a matter of fact, the number of possible configurations of the board is more than the number of atoms in the universe. This means that even if you took all the computers in the world and ran them for a million years, that wouldn't be enough compute power to calculate all the possible variations. Oftentimes, if you ask a great go player why they made a certain move, they will just say that it is their intuition that led them to placing that piece in that position.

This is the problem that the AlphaGo team had to try to solve. The first real indication that they were on the right path was when they invited a professional go player, Fan Hui, to play against their program. During the first game, he made some mistakes and lost to AlphaGo. He noted that it was one of the strangest games he had ever played. In the second game, he tried to change up his play style. Unfortunately, he also lost that game…and all the game after that. AlphaGo had beaten him five to zero. This made national headlines as the first time that a machine had beaten a professional go player. Fan Hui said that while he was, of course, sad to lose to a computer, he was happy that he is a part of history.

The media caught wind of this story and ran with it. Sensational headlines such as "Scientists have SOLVED the game of go a decade earlier than expected" went flying throughout the entire world. They correctly noted that there is a big difference in what happened with AlphaGo and Fan Hui and the way that the IBM computer beat Kasparov. The program that beat Kasparov had been programmed by expert chess players. Whereas AlphaGo more or less learned itself. Lee Sedol, an 18 time world champion, was asked in a press conference how he thought he would fare against AlphaGo, to which he responded by saying that he saw many flaws in the

version that played against Fan Hui and that he would be able to beat it without much difficulty. Of course, the AlphaGo team jumped at this opportunity to prove themselves against such a formidable appointment and set up a series of 5 games against Lee Sedol. Because the game was televised. AlphaGo would have a human interpreter, so to speak. As Lee Sedol would make a move, someone would enter that move into the computer. Then AlphaGo would select the next move, which is to be physically played on the real board by a member of the AlphaGo team. As game by game went AlphaGo's way, there was a growing sense of dread that much of the world watching felt. This was a televised moment, when one of the best human players in all of history seemed helpless against a machine. Although he did steal the fourth game away from AlphaGo, Lee Sedol resoundingly lost four out of five games. In his later interviews, he was very respectful in his loss and even noted that he thinks of AlphaGo as a tool to further humanity's collective understanding of the ancient game. One of the things that he mentioned that was different about AlphaGo was that it correctly realizes that it does not really matter how much a game is won by, all that matters is the win. So it would avoid making aggressive moves to capture a large area that could later turn out to be poor decisions. He notes that he will definitely work this new information into his game play.

Behind this incredible achievement is a deep neural network and reinforcement learning. This works due to the fact that go is a game with such clearly defined goals that AlphaGo can perform reinforcement learning. In previous iterations, specifically the one that defeated Lee Sedol, the AlphaGo team would start off by showing the neural network one hundred thousand games played by amateur players. Then, it would play games against itself, using reinforcement learning, many millions of times correcting and learning from its errors. When trying to decide its next move, it performs a tree search on the next possible positions, how it would affect the

position after that, and how it would affect the game. Because of the number of possibilities mentioned before, the researchers had to shrink the breadth and depth of moves that are searched. “The depth of the search may be reduced by position evaluation: truncating the search tree at state s and replacing the subtree below s by an approximate value function $v(s) \approx v^*(s)$ that predicts the outcome from state s”; additionally “the breadth of the search may be reduced by sampling actions from a policy $p(a|s)$ that is a probability distribution over possible moves a in position s” said Silver et al. This iteration would have two separate neural networks: the policy network and the value network. The policy network is used to output the next move that the player should make. The value network is used to predict the winner of the game from each position. These are the central ideas that powers the machine that beat Lee Sedol four games to one. Still, the team still saw improvements that could be made.

In the most recent iteration of AlphaGo, AlphaGo Zero. Researchers tried to move away from any possible previous notions and skipped the step of showing the neural network one hundred thousand human-played games. They thought that there was a possibility that the neural network may learn some intrinsically bad habits that humans made in their game play, so they seeked to approach this problem with ***tabula rasa,*** *an absence of preconceived ideas or predetermined goals; a clean slate.* The researchers had AlphaGo start by simply playing itself, starting with completely random play. Additionally, this version would only consider the position of black and white pieces as inputs to the model, abandoning the handful of hand engineered features that previous iterations used. This iteration also combines the two previous neural networks into one that could perform the duties of both its predecessors. These were the main changes that led to the new iteration of AlphaGo Zero. It very quickly learned some of the most fundamental strategies of the game, such as life and death, that took centuries to emerge in

humans. Within 3 days, it has surpassed the capabilities of the version that beat Lee Sedol. After a short period of additional, it was then able to defeat the AlphaGo Master, a smaller iteration that beat 60 professionals online and the then champion Ke Jie three out of three times, one hundred out of one hundred times. It can be argues that AlphaGo Zero is the greatest player in history with an ELO score higher than 5000, where the highest human score was lower than 4000.

Sources

Silver, D., Schrittwieser, J., Simonyan, K. et al. Mastering the game of Go without human knowledge. Nature 550, 354–359 (2017). https://doi.org/10.1038/nature24270

Silver, D., Huang, A., Maddison, C. et al. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489 (2016). https://doi.org/10.1038/nature16961

Figure 1

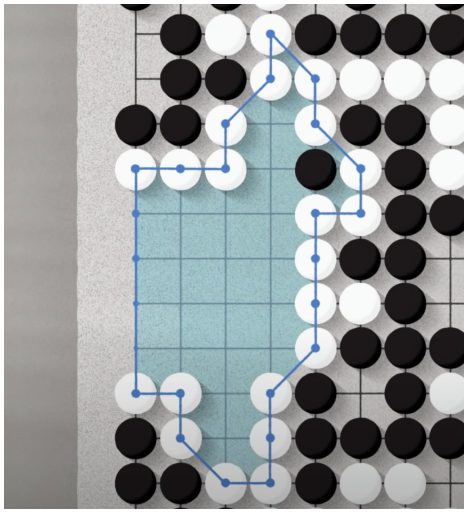

Figure 2