Mastering the game of Go without human knowledge

Paper by D.Silver et al. (2017)
Seminar by Jack Kleinsman
The game of Go

- $\sim 2 \times 10^{170}$ legal game states
- Thousands of years of history
- Over 20 million active players
# Machine Learning Methods

<table>
<thead>
<tr>
<th>Supervised Learning</th>
<th>Reinforcement Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>AlphaGo Fan, AlphaGo Lee, Alpha Go Master</em></td>
<td><em>AlphaGo Zero</em></td>
</tr>
<tr>
<td>- Learns from labelled data</td>
<td>- Learns from repeated self-play</td>
</tr>
<tr>
<td>- Requires a large dataset</td>
<td>- Only requires knowledge of the games rules</td>
</tr>
<tr>
<td>- Susceptible to overfitting</td>
<td></td>
</tr>
</tbody>
</table>
AlphaGo Zero’s performance

![Graph showing AlphaGo Zero's performance over time with Elo Rating on the y-axis and time in blocks on the x-axis. The graph compares AlphaGo Zero 40 blocks, AlphaGo Lee, and AlphaGo Master.]
Knowledge learnt by AlphaGo Zero

- The fundamental human knowledge of Go
- Discovered new techniques

AlphaGo Zero surpassed thousands of years of Go knowledge in just 40 days.
AlphaGo Zero’s architecture

- Input is the current game state
- One convolutional layer
- 40 residual layers
- Value and policy heads
What AlphaGo Zero sees...

Current position of black's stones
...and for the previous 7 time periods
All 1 if black to play
All 0 if white to play

19 x 19 x 17 stack

1 if black stone here
0 if black stone not here

Current position of white's stones
...and for the previous 7 time periods

jkke533 113166519
Convolutional layer

- The input game state has 6,137 dimensions.
- It suffers from ‘curse of dimensionality’.
- The convolutional layer reduces the dimensions of the data while preserving the important information.
Convolutional layer

gif from www.towardsdatascience.com

jkle533 113166519
After passing a threshold, the performance of traditional CNN’s decrease as the number of layers increase.
Residual layers

Figure 2. Residual learning: a building block.
Limitations

- AlphaGo Zero requires full knowledge of the rules
- Simulating problems is computationally expensive
- There must be a way to calculate how correct a solution is
Conclusion

- Computers have achieved superhuman performance in Go
- A reinforcement learning based algorithm surpassed the supervised learning algorithms
- Reinforcement learning was able to learn strategies both existing and new
- A generalized version of the AlphaGo Zero algorithm, AlphaZero, has achieved similar results in chess and shogi
References


References

