Mastering the Game of Go without Human Knowledge

06/15/18

Presented by: Henry Chen

CS885 Reinforcement Learning
Introduction
Introduction

The Game of Go

- ancient board game
- 19 x 19 grid
- complexity: $\sim 10^{170}$

Challenging AI problem

- How to search through an intractable search space?
- Breakthrough: AlphaGo

Image source: https://medium.com/@karpathy/alphago-in-context-c47718cb95a5
Background

AlphaGo

- March 2016: defeated 18-time world champion Lee Sedol 4-1

Image source: https://www.tastehit.com/blog/google-deepmind-alphago-how-it-works/
Background

AlphaGo - Architecture

1. Policy Network
   - Purpose: decide next best move
   - *Convolution Neural Network* (13 hidden layers)
   - Stage 1: *Supervised Learning* to predict human expert moves (57%)
   - Stage 2: Improve network by *Policy Gradient Reinforcement Learning* through self-play using roll-out policy (80% > stage 1)

Source: Google DeepMind, Mastering the Game of Go with Deep Neural Networks and Tree Search
Background

AlphaGo - Architecture

2. Value Network
   - Purpose: evaluate chances of winning
   - Convolution Neural Network (14 hidden layers)
   - Train network by regression on state-outcome pair sampled from self-play data using policy network

Source: Google DeepMind, Mastering the Game of Go with Deep Neural Networks and Tree Search
Background

Policy Network (stage 1):
- 30 millions position from 160,000 human games
- 50 GPUs
- 3 weeks

Policy Network (stage 2):
- 10,000 mini-batches of 128 self-play games
- 50 GPUs
- 1 day

Value Network
- 30 millions unique positions
- 50 GPUs
- 1 week

Source: Google DeepMind, Mastering the Game of Go with Deep Neural Networks and Tree Search
Background

3. Monte-Carlo Tree Search (MCTS)
   - Purpose: Combining policy and value networks to select actions by lookahead search
   - Asynchronous multi-threaded search (distributed ~50 GPUs)

Source:
Google DeepMind, Mastering the Game of Go with Deep Neural Networks and Tree Search
Background

Limitations

- Require large data-set of expert games
- Use of handcraft features
- Asynchronous training and computation intensive

<table>
<thead>
<tr>
<th>Feature</th>
<th># of patterns</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>1</td>
<td>Whether move matches one or more response features</td>
</tr>
<tr>
<td>Save atari</td>
<td>1</td>
<td>Move saves stone(s) from capture</td>
</tr>
<tr>
<td>Neighbour</td>
<td>8</td>
<td>Move is 8-connected to previous move</td>
</tr>
<tr>
<td>Nakade</td>
<td>8192</td>
<td>Move matches a nakade pattern at captured stone</td>
</tr>
<tr>
<td>Response pattern</td>
<td>32207</td>
<td>Move matches 12-point diamond pattern near previous move</td>
</tr>
<tr>
<td>Non-response pattern</td>
<td>69338</td>
<td>Move matches 3 x 3 pattern around move</td>
</tr>
<tr>
<td>Self-atari</td>
<td>1</td>
<td>Move allows stones to be captured</td>
</tr>
<tr>
<td>Last move distance</td>
<td>34</td>
<td>Manhattan distance to previous two moves</td>
</tr>
<tr>
<td>Non-response pattern</td>
<td>32207</td>
<td>Move matches 12-point diamond pattern centred around move</td>
</tr>
</tbody>
</table>

Source: Google DeepMind, Mastering the Game of Go with Deep Neural Networks and Tree Search
Content of paper

Mastering the Game of Go without Human Knowledge


DeepMind, 5 New Street Square, London EC4A 3TW.

*These authors contributed equally to this work.
Content of paper

AlphaGo Zero

1. uses no Human Knowledge and learn only by Self-Play

Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge
Content of paper

AlphaGo Zero

Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge
Content of paper

AlphaGo Zero

2. Single Neural Network with ResNets Structure
   ▪ Dual purpose: decide next best move and evaluate chances of winning

Source: Google DeepMind, Mastering the Game of Go without Human Knowledge

Source: http://neural.vision/blog/article-reviews/deep-learning/he-resnet-2015/
Content of paper

AlphaGo Zero

3. Simpler Tree Search

Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge
Content of paper

AlphaGo Zero

4. Requires no handcraft features
   - Only requires raw board representations and its history, plus some basic game rules as neural network input

5. Improved computation efficiency
   - Single machine on Google Cloud with 4 TPUs

Source:
Google DeeMind, Mastering the Game of Go without Human Knowledge
Empirical Evaluation

- Training for 3 days

Source: Google DeepMind, Mastering the Game of Go without Human Knowledge
Empirical Evaluation

- Comparison of neural network architectures

Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge
Empirical Evaluation

- Discovering existing strategies and some unknown by human
Empirical Evaluation

- Training for 40 days

Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge
Conclusion

▪ Pure reinforcement learning is fully feasible, even in the most challenging domain

▪ It is possible to achieve superhuman performance, without human knowledge

▪ In the matter of days, AlphaGo Zero rediscover Go knowledge accumulated by human over thousands of year; it also discover new insights and strategies for the game
Discussion

- Some critics suggest AlphaGo is a very narrow AI and it rely on many properties of Go. Do you think the algorithm can be generalized for another domain?
- Did this paper inspire you in any way? Any suggestions for improvement?
- Do you think we should use AI to discover more knowledge?
- How do you feel about superintelligence AI? Are you in the Elon Musk or Mark Zuckerberg camp?
Images source:
https://jedionston.wordpress.com/2015/02/14/go-wei-chi-vs-tafl-hnaftafl/