

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

Intro to Reinforcement Learning

Data Science Utrecht

Rob Romijnders

Eindhoven, University of Technology

RomijndersRob@gmail.com

March 23, 2017

Overview

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient
Gradient
Variance
reduction

Case: Alpha
Go

Questions

1 Why

- Examples
- Why not

2 State and policy

3 Policy gradient

- Gradient
- Variance reduction

4 Case: Alpha Go

5 Questions

Popular examples

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and

policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions



(a) Atari



(b) Alpha Go

Figure: Examples of RL

Robots

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

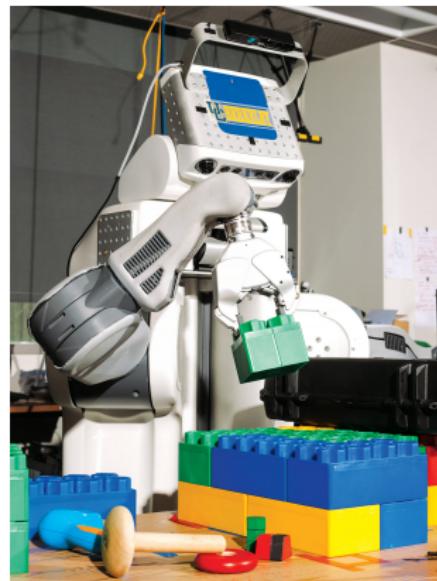


Figure: credits: bloomberg.com/features/2015-preschool-for-robots/

Poker

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions



Figure: credits: cardschat.com Jan, 2017

Where are we?

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

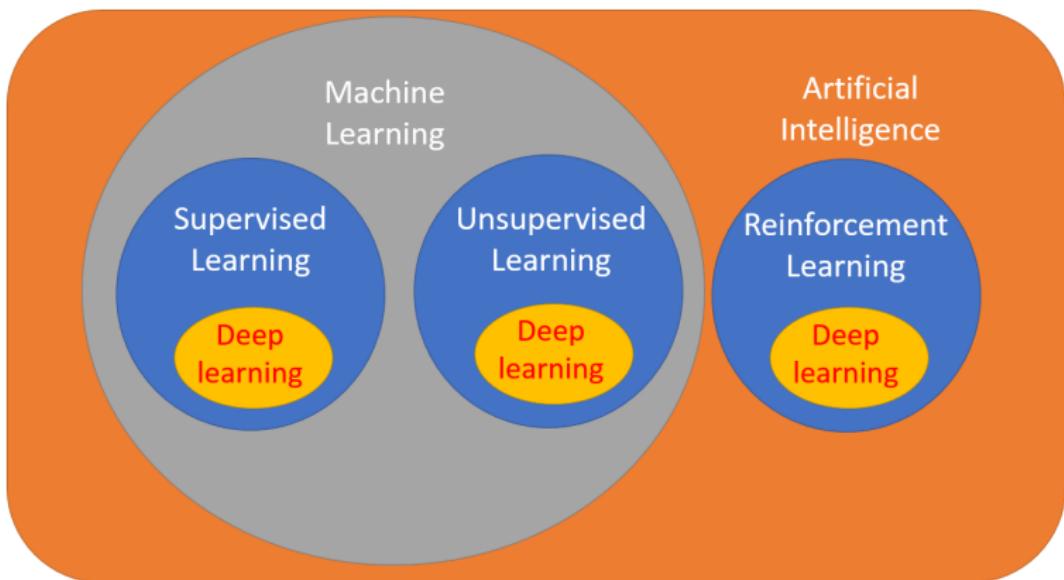
State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions



Why not RL?

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient
Gradient
Variance
reduction

Case: Alpha
Go

Questions

- 1 Know when you are doing supervised learning. And don't force it into RL
- 2 Consider imitation learning

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

State Policy

March 23,
2017

Rob
Romijnders

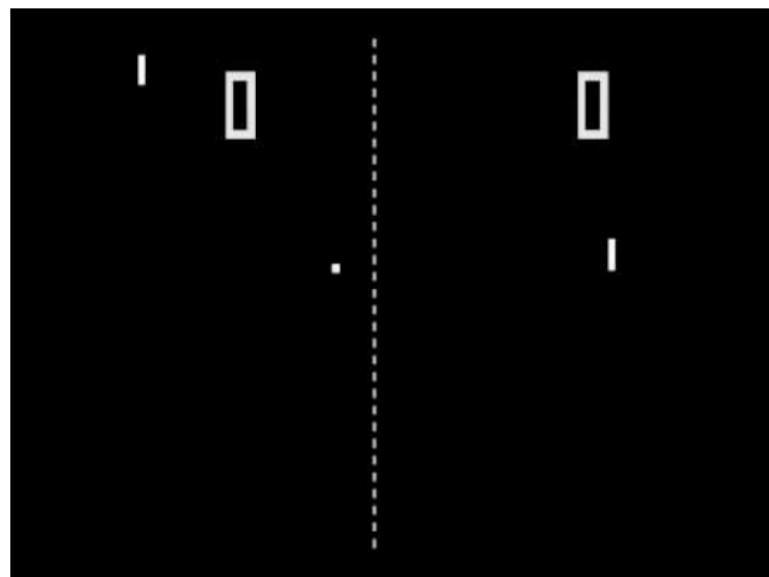
Why
Examples
Why not

State and
policy

Policy
gradient
Gradient
Variance
reduction

Case: Alpha
Go

Questions



March 23,
2017

State and policy

Questions



Figure: Playing FPS Games with Deep Reinforcement Learning (Lample, Chaplot. ArXiv 2016)

March 23,
2017

Rob
Romijnders

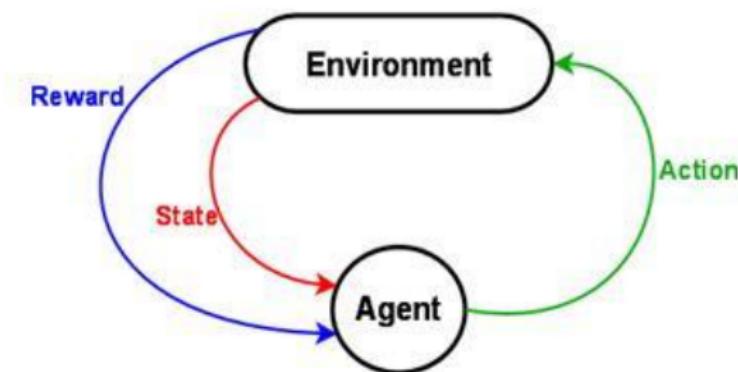
Why
Examples
Why not

State and
policy

Policy
gradient
Gradient
Variance
reduction

Case: Alpha
Go

Questions



March 23,
2017

Rob
Romijnders

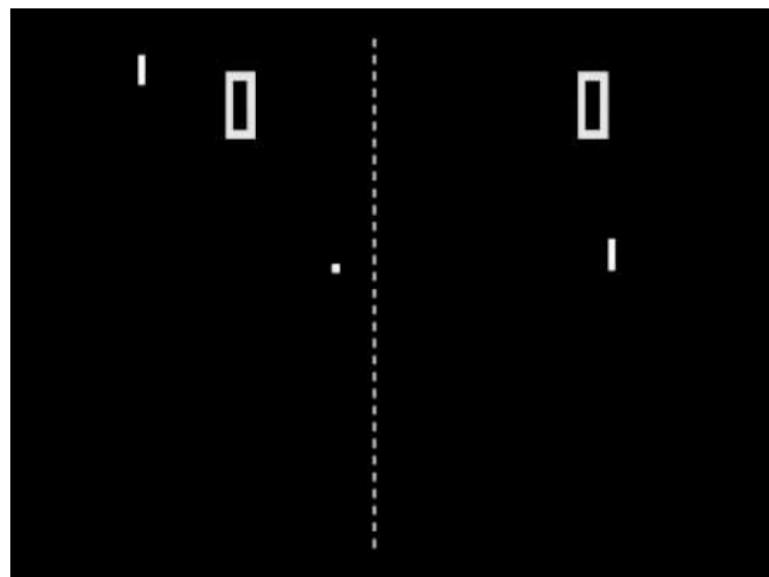
Why
Examples
Why not

State and
policy

Policy
gradient
Gradient
Variance
reduction

Case: Alpha
Go

Questions



March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

state → s

$a \sim \pi_\theta(a|s)$

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

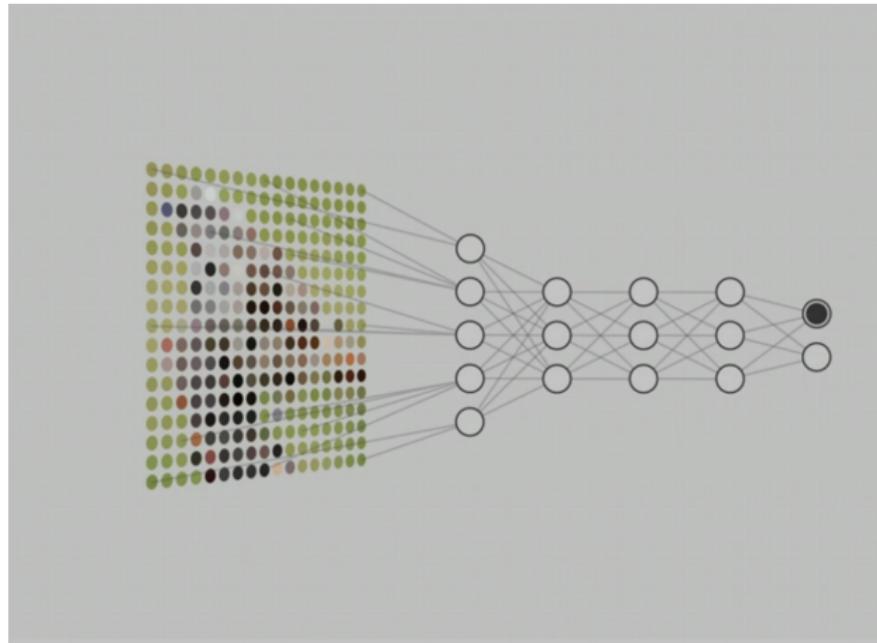


Figure: credits: Blaise Aguera y Arcas

Python code

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient
Gradient
Variance
reduction

Case: Alpha
Go

Questions

```
h = np.tanh(np.dot(W1, s) + b1))  
logpi = np.dot(W2, h) + b2  
p = softmax(logpi)
```

Architecture

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

How to train a policy?

Python code

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy

gradient

Gradient

Variance
reduction

Case: Alpha

Go

Questions

```
h = np.tanh(np.dot(W1, s) + b1))  
logpi = np.dot(W2, h) + b2  
p = softmax(logpi)
```

Gradient Descent

March 23,
2017

Rob
Romijnders

Why
Examples
Why not
State and
policy
Policy
gradient
Gradient
Variance
reduction
Case: Alpha
Go
Questions

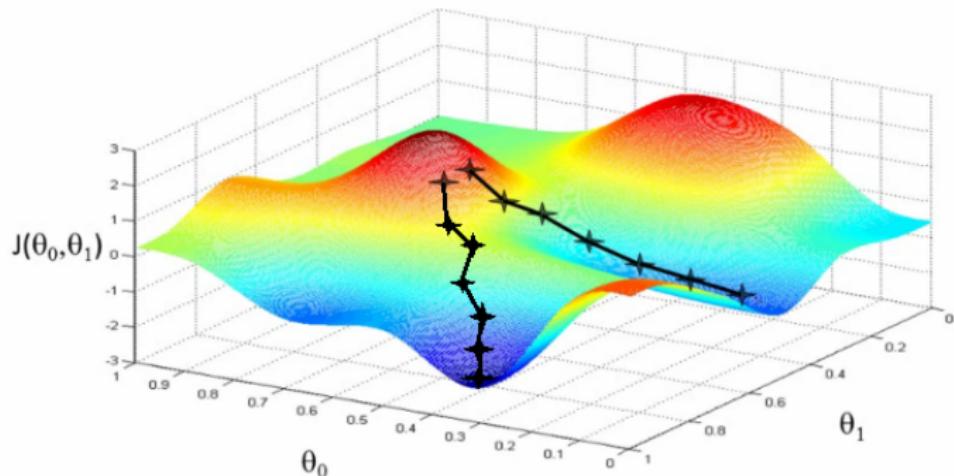


Figure: Credits: ML blog Vasilis Vryniotis

Policy Gradient

March 23,
2017

Rob
Romijnders

Why
Examples
Why not
State and policy
Policy gradient
Gradient Variance reduction
Case: Alpha Go
Questions

1 Reinforcement learning

$$grad = \sum_t \nabla_{\theta} \log \pi(a_t | s_t) R$$

2 Supervised learning

$$grad = \sum_i \nabla_{\theta} \log p(y_i | x_i)$$

Roll outs

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

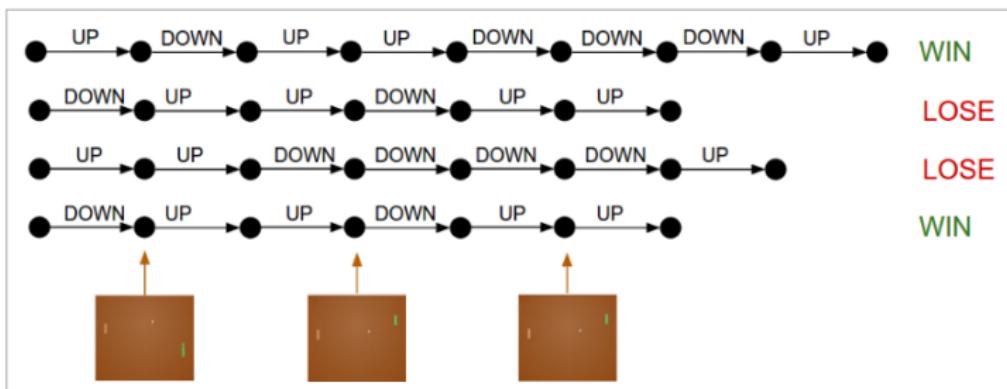


Figure: Credits: karpathy.github.io

Policy Gradient

March 23,
2017

Rob
Romijnders

Why
Examples
Why not
State and policy
Policy gradient
Gradient Variance reduction
Case: Alpha Go
Questions

1 Reinforcement learning

$$grad = \sum_t \nabla_{\theta} \log \pi(a_t | s_t) R$$

2 Supervised learning

$$grad = \sum_i \nabla_{\theta} \log p(y_i | x_i)$$

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

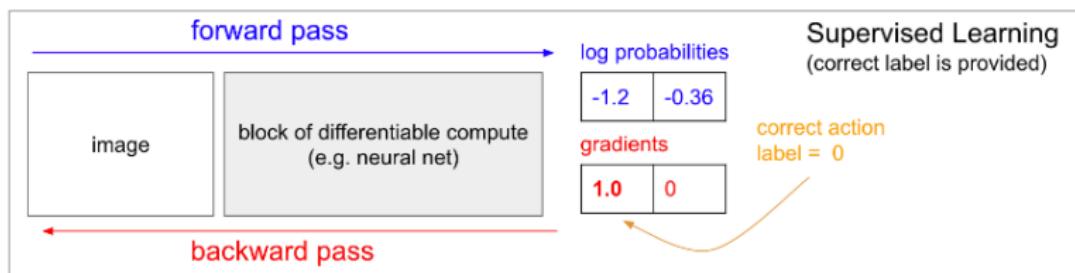


Figure: credits: karpathy.github.io

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

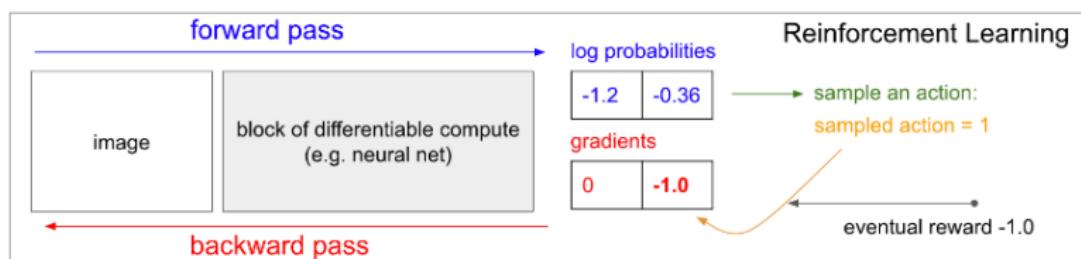


Figure: credits: karpathy.github.io

Policy Gradient

March 23,
2017

Rob
Romijnders

Why
Examples
Why not
State and policy
Policy gradient
Gradient Variance reduction
Case: Alpha Go
Questions

1 Reinforcement learning

$$grad = \sum_t \nabla_{\theta} \log \pi(a_t | s_t) R$$

2 Supervised learning

$$grad = \sum_i \nabla_{\theta} \log p(y_i | x_i)$$

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

Playing Pong with Policy Gradients

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions



	Left	Right
+250	+180	
+50	-20	+200

Variance reduction

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient
Gradient
Variance
reduction

Case: Alpha
Go

Questions

$$\text{grad} = \sum_t \nabla_{\theta} \log \pi(a_t | s_t) R$$

$$\text{grad} = \sum_t \nabla_{\theta} \log \pi(a_t | s_t) \sum_{t'} r_{t'}$$

$$\text{grad} = \sum_t \nabla_{\theta} \log \pi(a_t | s_t) \sum_{t'} (r_{t'} - v(s_t))$$

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient

Variance
reduction

Case: Alpha
Go

Questions



March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

Actor Critic at work

Go

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

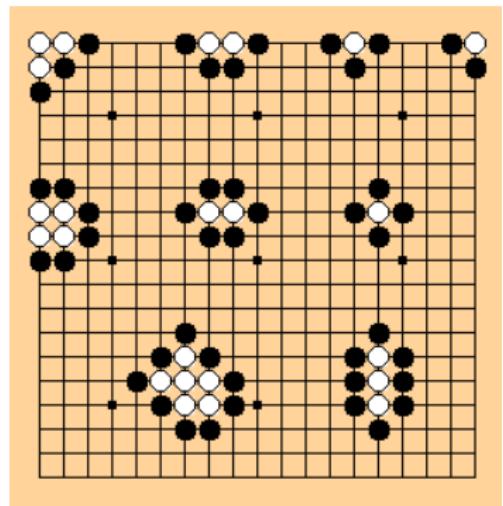
Gradient
Variance
reduction

Case: Alpha
Go

Questions



(a) Go



(b) Capture territory

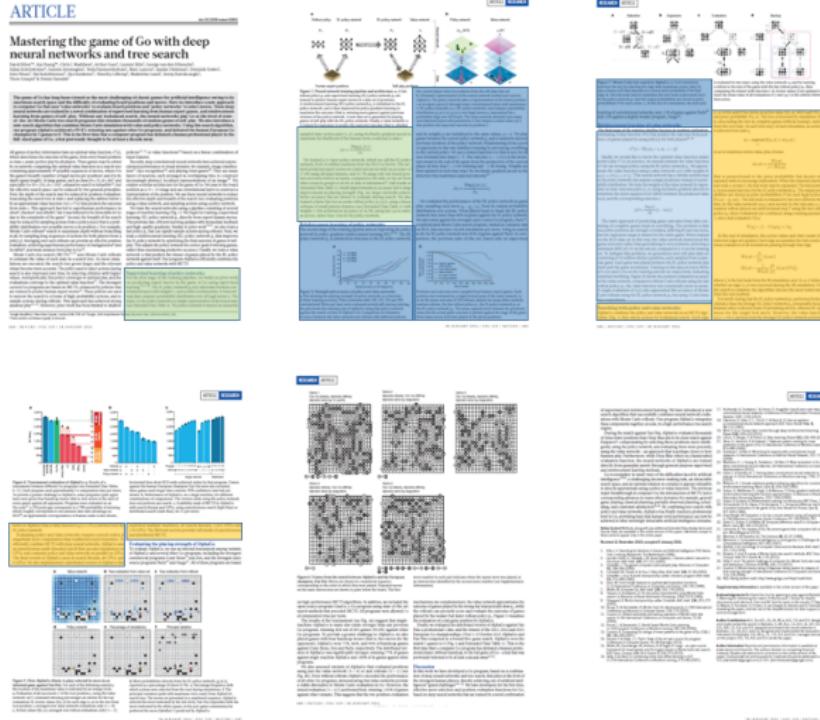
Figure: Go

Alpha Go paper in Nature

March 23,
2017

Case: Alpha Go

Questions



MCTS

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

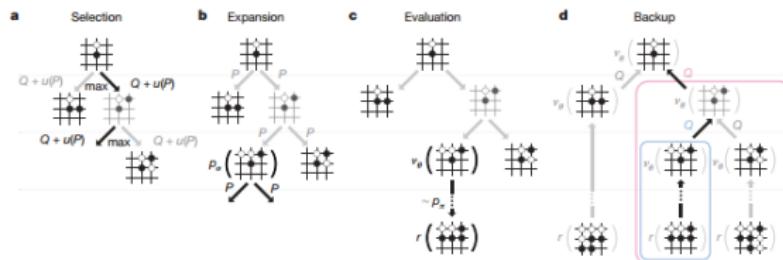
State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions



March 23,
2017

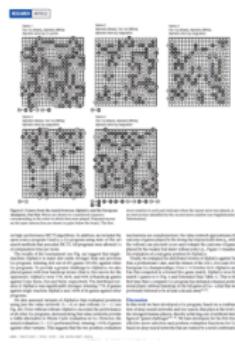
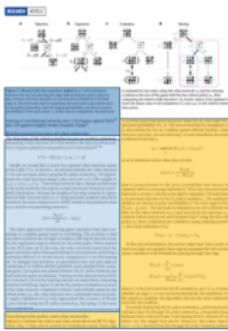
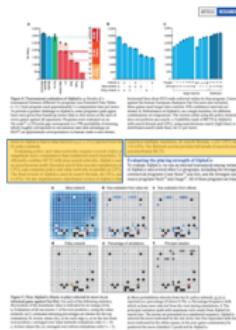
Case: Alpha Go

ARTICLE

Mastering the game of Go with deep neural networks and tree search

The game of Go has long been viewed as the most challenging of chess games for a computer to master. In 2016, Google DeepMind's AlphaGo program became the first computer to beat a professional Go player, and since then it has become a hobby for many people to play against the program. This research project aims to create a Go-playing computer program that can learn from its mistakes and improve its performance over time. The program will use a combination of reinforcement learning, neural network-based game models, and rule-based reasoning to make decisions. The goal is to create a Go-playing computer program that can compete at a professional level.

“*What’s the point of all this?*” I asked. “*What does it mean?*” I asked again. “*What does it all mean?*” I asked again. “*What does it all mean?*” I asked again.



March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

Conclusion

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

Questions

References on next slide

Further reading

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

1 Reinforcement learning

- 1 David Silver's course at UCL
- 2 Book: Reinforcement Learning by Sutton and Barto

2 Deep Reinforcement learning

- 1 Deep Reinforcement Learning: Pong from Pixels, Andrej Karpathy
- 2 Deep Reinforcement Learning, CS 294 at Berkeley

3 Play yourself

- 1 Pol. Grad. in 130 lines using only NumPy
- 2 Open AI gym

Venn diagram

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

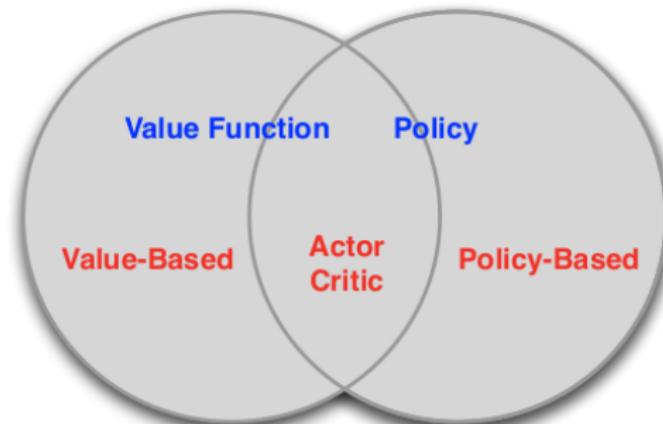


Figure: Credits: David Silver's slides at UCL

Quiz

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient
Gradient
Variance
reduction

Case: Alpha
Go

Questions

- 1 Cluster customers based on their features
- 2 Play chess with an algorithm
- 3 Detect the outlier in a stream of reports
- 4 Predict tomorrow's weather
- 5 Indicate faces in an image
- 6 Have the robot find the exit of a maze

State

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions



Figure: State of a chess board. Credits: chess.com

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

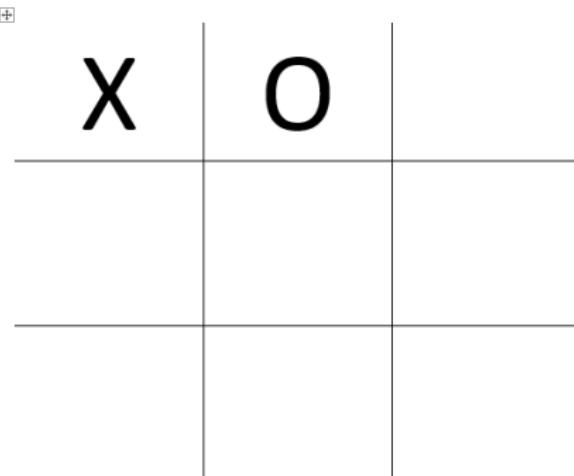
State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions



March 23,
2017

Rob
Romijnders

Why
Examples
Why not
State and
policy
Policy
gradient
Gradient
Variance
reduction
Case: Alpha
Go
Questions

0:X	1:O	2:_
3:_	4:_	5:_
6:_	7:_	8:_

0	X
1	O
2	-
3	-
4	-
5	-
6	-
7	-
8	-

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

$$s = \begin{pmatrix} 1 \\ 0 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}$$

March 23,
2017

Rob
Romijnders

Why
Examples
Why not

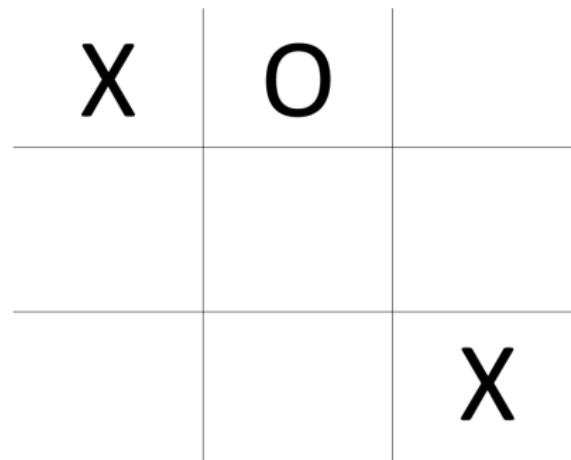
State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions



March 23,
2017

Rob
Romijnders

Why
Examples
Why not
State and
policy
Policy
gradient
Gradient
Variance
reduction
Case: Alpha
Go
Questions

0:X	1:0	2:_
3:_	4:_	5:_
6:_	7:_	8:X

0	X
1	O
2	-
3	-
4	-
5	-
6	-
7	-
8	X

March 23,
2017

Rob
Romijnders

Why

Examples

Why not

State and
policy

Policy
gradient

Gradient
Variance
reduction

Case: Alpha
Go

Questions

$$s = \begin{pmatrix} 1 \\ 0 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$$