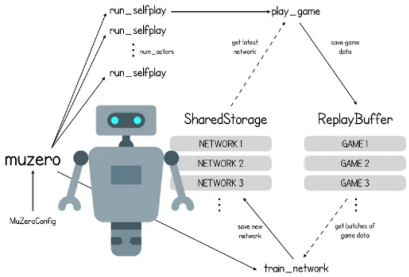
Published in Applied Data Science

You have **1** free member-only story left this month. Upgrade for unlimited access.

David Foster

Dec 2, 2019 · 7 min read · Member-only · Listen



MuZero: The Walkthrough (Part 1/3)

Teaching a machine to play games using self-play and deep learning... without telling it the rules 😱

> *If you want to learn how one of the most sophisticated AI systems ever built works, you've come to the right place.*

In this three part series, we'll explore the inner workings of the DeepMind MuZero model — the younger (and even more impressive) brother of AlphaZero.

👉 Part 2

👉 Part 3

Also check out my latest post, about how to train reinforcement learning agents for multi-player board games, using self-play!

👉 Self-Play in Multiplayer Environments

We'll be walking through the pseudocode that accompanies the MuZero paper — so grab yourself a cup of tea and a comfy chair and let's begin.

α

The story so far...

On 19th November 2019 DeepMind released their latest model-based reinforcement learning algorithm to the world — MuZero.

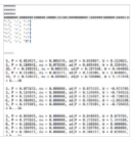
This is the fourth in a line of DeepMind reinforcement learning papers that have continually smashed through the barriers of possibility, starting with AlphaGo in 2016.

To read about the full history from AlphaGo through to AlphaZero — check out my previous blog 👇

How to build your own AlphaZero AI using Python and Keras

Teach a machine to learn Connect4 strategy through self-play and deep learning.

medium.com



AlphaZero was hailed as the general algorithm for getting good at something, quickly, without any prior knowledge of human expert strategy.

So...what now?

μ

MuZero

Get unlimited access

David Foster
5.1K Followers

Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

Help Status Writers Blog Careers Privacy Terms About
David Foster
5.1K Followers

Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

Help Status Writers Blog Careers Privacy Terms About
Text to speech
David Foster
5.1K Followers

Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

Help Status Writers Blog Careers Privacy Terms About
Text to speech
David Foster
5.1K Followers

Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

[Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model]

MuZero takes the ultimate next step. Not only does MuZero deny itself human strategy to learn from. It isn't even shown the rules of the game.

In other words, for chess, AlphaZero is set the following challenge:

> *Learn how to play this game on your own — here's the rulebook that explains how each piece moves and which moves are legal. Also it tells you how to tell if a position is checkmate (or a draw).*

MuZero on the other hand, is set this challenge:

> *Learn how to play this game on your own — I'll tell you what moves are legal in the current position and when one side has won (or it's a draw), but I won't tell you the overall rules of the game.*

Alongside developing winning strategies, MuZero must therefore also develop its own dynamic model of the environment so that it can understand the implications of its choices and plan ahead.

Imagine trying to become better than the world champion at a game where you are never told the rules. MuZero achieves precisely this.

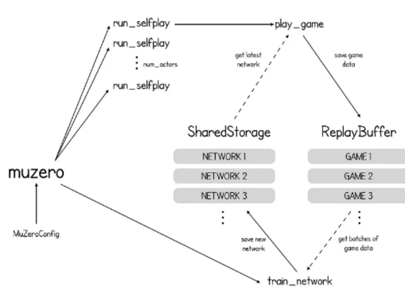
In the next section we will explore how MuZero achieves this amazing feat, by walking through the codebase in detail.

The MuZero pseudocode

Alongside the MuZero preprint paper, DeepMind have released Python pseudocode detailing the interactions between each part of the algorithm.

In this section, we'll pick apart each function and class in a logical order, and I'll explain what each part is doing and why. We'll assume MuZero is learning to play chess, but the process is the same for any game, just with different parameters. All code is from the open-sourced DeepMind pseudocode.

Let's start with an overview of the entire process, starting with the entrypoint function, `muzero`.



Overview of the MuZero self-play and training process

```python
def muzero(config: MuZeroConfig):
  storage = SharedStorage()
  replay_buffer = ReplayBuffer(config)

  for _ in range(config.num_actors):
    launch_job(run_selfplay, config, storage, replay_buffer)

  train_network(config, storage, replay_buffer)

  return storage.latest_network()
```

pseudocode.py hosted with ❤ by GitHub   view raw

The entrypoint function `muzero` is passed a `MuZeroConfig` object, which stores important information about the parameterisation of the run, such as the `action_space_size` (number of possible actions) and `num_actors` (the number of parallel game simulations to spin up). We'll go through these parameters in more detail as we encounter them in other functions.

At a high level, there are two independent parts to the MuZero algorithm — *self-play* (creating game data) and *training* (producing improved versions of the neural network). The `SharedStorage` and `ReplayBuffer` objects can be accessed by both halves of the algorithm and store neural network versions and game data respectively.

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

Help Status Writers Blog Careers Privacy Terms About Text to speech

David Foster
5.1K Followers
Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

Help Status Writers Blog Careers Privacy Terms About Text to speech

David Foster
5.1K Followers
Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

Help Status Writers Blog Careers Privacy Terms About Text to speech

David Foster
5.1K Followers
Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Shared Storage and the Replay Buffer

The `SharedStorage` object contains methods for saving a version of the neural network and retrieving the latest neural network from the store.

```
class SharedStorage(object):

  def __init__(self):
    self._networks = {}

  def latest_network(self) -> Network:
    if self._networks:
      return self._networks[max(self._networks.keys())]
    else:
      # policy -> uniform, value -> 0, reward -> 0
      return make_uniform_network()

  def save_network(self, step: int, network: Network):
    self._networks[step] = network
```

pseudocode.py hosted with ❤ by GitHub view raw

We also need a `ReplayBuffer` to store data from previous games. This takes the following form:

```
class ReplayBuffer(object):

  def __init__(self, config: MuZeroConfig):
    self.window_size = config.window_size
    self.batch_size = config.batch_size
    self.buffer = []

  def save_game(self, game):
    if len(self.buffer) > self.window_size:
      self.buffer.pop(0)
    self.buffer.append(game)

  ...
```

pseudocode.py hosted with ❤ by GitHub view raw

Notice how the `window_size` parameter limits the maximum number of games stored in the buffer. In MuZero, this is set to the latest 1,000,000 games.

Self-play (run_selfplay)

After creating the shared storage and replay buffer, MuZero launches `num_actors` parallel game environments, that run independently. For chess, `num_actors` is set to 3000. Each is running a function `run_selfplay` that grabs the latest version of the network from the store, plays a game with it (`play_game`) and saves the game data to the shared buffer.

```
# Each self-play job is independent of all others; it takes the latest network
# snapshot, produces a game and makes it available to the training job by
# writing it to a shared replay buffer.
def run_selfplay(config: MuZeroConfig, storage: SharedStorage,
                 replay_buffer: ReplayBuffer):
  while True:
    network = storage.latest_network()
    game = play_game(config, network)
    replay_buffer.save_game(game)
```

pseudocode.py hosted with ❤ by GitHub view raw

So in summary, MuZero is playing thousands of games against itself, saving these to a buffer and then training itself on data from those games. So far, this is no different to AlphaZero.

To end Part 1, we will cover one of the key differences between AlphaZero and MuZero — why does MuZero have three neural networks, whereas AlphaZero only has one?

The 3 Neural Networks of MuZero

Both AlphaZero and MuZero utilise a technique known as **Monte Carlo Tree Search (MCTS)** to select the next best move.

The idea is that in order to select the next best move, it makes sense to 'play

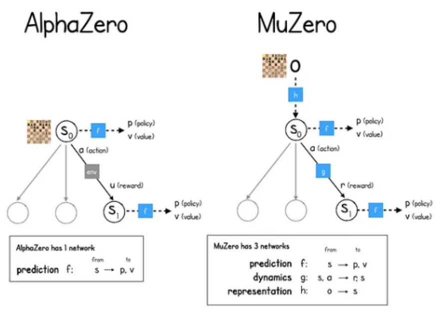
Terence Shin
All Machine Learning Algorithms You Should Know for 2023
Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work
Help Status Writers Blog Careers Privacy Terms About Text to speech
David Foster
5.1K Followers
Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners
Follow
More from Medium
Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot
Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why
Terence Shin
All Machine Learning Algorithms You Should Know for 2023
Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work
Help Status Writers Blog Careers Privacy Terms About Text to speech
David Foster
5.1K Followers
Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners
Follow
More from Medium
Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot
Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why
Terence Shin
All Machine Learning Algorithms You Should Know for 2023
Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work
Help Status Writers Blog Careers Privacy Terms About Text to speech
David Foster
5.1K Followers
Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners
Follow
More from Medium
Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot
Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why
Terence Shin
All Machine Learning Algorithms You Should Know for 2023
Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work
Help Status Writers Blog Careers Privacy Terms About Text to speech
David Foster

out' likely future scenarios from the current position, evaluate their value using a neural network and choose the action that maximises the future expected value. This seems to be what we humans are doing in our head when playing chess, and the AI is also designed to make use of this technique.

However, MuZero has a problem. As it doesn't know the rules of the game, it has no idea how a given action will affect the game state, so it cannot imagine future scenarios in the MCTS. It doesn't even know how to work out what moves are legal from a given position, or whether one side has won.

> *The stunning development in the MuZero paper is to show that this doesn't matter. MuZero learns how to play the game by creating a dynamic model of the environment within its own imagination and optimising within this model.*

The diagram below shows a comparison between the MCTS processes in AlphaZero and MuZero:



Whereas AlphaZero only has only one neural network (**prediction**), MuZero needs three (**prediction**, **dynamics**, **representation**)

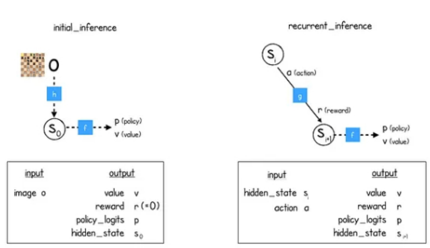
The job of the AlphaZero **prediction** neural network `f` is to predict the policy `p` and value `v` of a given game state. The policy is a probability distribution over all moves and the value is just a single number that estimates the future rewards. This prediction is made every time the MCTS hits an unexplored leaf node, so that it can immediately assign an estimated value to the new position and also assign a probability to each subsequent action. The values are backfilled up the tree, back to the root node, so that after many simulations, the root node has a good idea of the future value of the current state, having explored lots of different possible futures.

MuZero also has a **prediction** neural network `f`, but now the 'game state' that it operates on is a hidden representation that MuZero learns how to evolve through a **dynamics** neural network `g`. The dynamics network takes the current hidden state `s` and chosen action `a` and outputs a reward `r` and new state. Notice how in AlphaZero, moving between states in the MCTS tree is simply a case of asking the environment. MuZero doesn't have this luxury, so needs to build its own dynamic model!

Lastly, in order to map from the current observed game state to the initial representation, MuZero uses a third **representation** neural network, `h`.

There are therefore two inference functions MuZero needs, in order to move through the MCTS tree making predictions:

- `initial_inference` for the current state. `h` followed by `f` (representation followed by prediction) .
- `recurrent_inference` for moving between states inside the MCTS tree. `g` followed by `f` (representation followed by dynamics).



The two types of inference in MuZero

The exact models aren't provided in the pseudocode, but detailed descriptions are given in the accompanying paper.

5.1K Followers
Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners
Follow
More from Medium
Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot
Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why
Terence Shin
All Machine Learning Algorithms You Should Know for 2023
Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work
Help Status Writers Blog Careers Privacy Terms About Text to speech
David Foster
5.1K Followers
Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners
Follow
More from Medium
Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot
Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why
Terence Shin
All Machine Learning Algorithms You Should Know for 2023
Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work
Help Status Writers Blog Careers Privacy Terms About Text to speech
David Foster
5.1K Followers
Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners
Follow
More from Medium
Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot
Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why
Terence Shin
All Machine Learning Algorithms You Should Know for 2023
Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work
Help Status Writers Blog Careers Privacy Terms About Text to speech
David Foster
5.1K Followers
Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners
Follow
More from Medium
Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

```python
class NetworkOutput(typing.NamedTuple):
  value: float
  reward: float
  policy_logits: Dict[Action, float]
  hidden_state: List[float]


class Network(object):

  def initial_inference(self, image) -> NetworkOutput:
    # representation + prediction function
    return NetworkOutput(0, 0, {}, [])

  def recurrent_inference(self, hidden_state, action) -> NetworkOutput:
    # dynamics + prediction function
    return NetworkOutput(0, 0, {}, [])

  def get_weights(self):
    # Returns the weights of this network.
    return []

  def training_steps(self) -> int:
    # How many steps / batches the network has been trained for.
    return 0
```

pseudocode.py hosted with ❤ by GitHub view raw

In summary, in the absence of the actual rules of chess, MuZero creates a new game inside its mind that it can control and uses this to plan into the future. The three networks (**prediction**, **dynamics** and **representation**) are optimised together so that strategies that perform well inside the imagined environment, also perform well in the real environment.

Amazing stuff.

. . .

This is the end of Part 1 — in Part 2, we'll start by walking through the `play_game` function and see how MuZero makes a decision about the next best move at each turn.

Please clap if you've enjoyed this post and I'll see you in Part 2!

. . .

APPLIED DATA SCIENCE

This is the blog of Applied Data Science Partners, a consultancy that develops innovative data science solutions for businesses. To learn more, feel free to get in touch through our website.

Machine Learning | Deep Learning | Data Science | Artificial Intelligence | AI

1.5K 7

Get an email whenever David Foster publishes.

Emails will be sent to dvmixalkin@gmail.com. Not you?

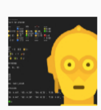
Subscribe

More from Applied Data Science Follow

Cutting edge data science, machine learning and AI projects

David Foster · Jan 25, 2021

How To Build Your Own AI To Play Any Board Game



NEW reinforcement learning Python package SIMPLE — Self-play In MultiPlayer Environments — ✏ The Plan In November, I set out to write a Python package that can train AI agents to play any board...

Deep Learning 10 min read

Share your ideas with millions of readers. Write on Medium

Daniel Sharp · Jan 17, 2022

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

Help Status Writers Blog Careers Privacy Terms About Text to speech

David Foster
5.1K Followers

Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

Help Status Writers Blog Careers Privacy Terms About Text to speech

David Foster
5.1K Followers

Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

Help Status Writers Blog Careers Privacy Terms About Text to speech

David Foster
5.1K Followers

Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

How To Build You Own Slack Bot

Full Stack Data Scientist: Part 7— How to set up a Slack Bot for instant, automatic notifications — Dashboards are the most popular way of presenting data in businesses — but they're not the only way!...

Data Science 5 min read

David Foster · Jan 26, 2018

How to build your own AlphaZero AI using Python and Keras

Teach a machine to learn Connect4 strategy through self-play and deep learning — Update! (2nd December 2019) I've just released a series on...

Artificial Intelligence 11 min read

David Foster · Sep 28, 2017

NEW R package that makes XGBoost interpretable

xgboostExplainer makes your XGBoost model as transparent and 'white-box' as a single decision tree — In this post I'm going to do three things: Show you how a single decision tree is not great at predicting,...

Machine Learning 9 min read

David Foster · Oct 29, 2017

AlphaGo Zero Explained In One Diagram

Download the AlphaGo Zero cheat sheet — Get the full cheat sheet here Update! (2nd December 2019) I've just released a series on MuZero — AlphaZero's younger and cooler brother. Check it out 👇 How to Build...

Machine Learning 2 min read

Read more from Applied Data Science

David Foster
5.1K Followers

Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

David Foster
5.1K Followers

Author of the Generative Deep Learning book :: Founding Partner of Applied Data Science Partners

Follow

More from Medium

Molly Ru... in Towards Data Scie...
How ChatGPT Works: The Models Behind The Bot

Somnath ... in JavaScript in Plai...
Coding Won't Exist In 5 Years. This Is Why

Terence Shin
All Machine Learning Algorithms You Should Know for 2023

Josep Ferrer in Geek Culture
5 ChatGPT features to boost your daily work

Help Status Writers Blog Careers Privacy Terms About
Text to speech