

[RL22] Q&A Session on Model-based RL #1

30.06.2022

Christopher Mutschler

Let's play Kahoot! again...

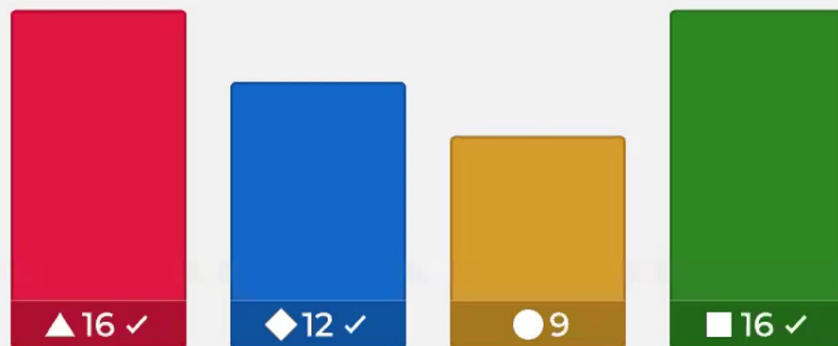
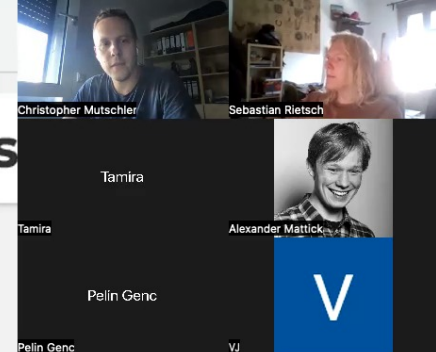
Kahoot!

Let's play Kahoot!

The image shows a screenshot of the Kahoot! website homepage. The browser address bar shows 'kahoot.com'. The navigation menu includes 'Kahoot!', 'News', 'School', 'Work', 'Home', 'Study', 'Academy', 'AccessPass', 'Contact sales', 'Explore content', 'Play', 'Sign up', 'Log in', and 'EN'. The 'Play' button is circled in red. Below the navigation are four promotional cards:

- Make learning awesome!**
Kahoot! delivers engaging learning to billions.
[Sign up for free!](#)
- Make your team superstar presenters**
Set your whole team up to deliver awesome presentations with Kahoot! 360 Spirit, our best plan from only \$16 per month.
[Learn more >](#)
[Buy now](#)
- NEW! Create a branded experience with Kahoot! themes**
Boost audience engagement by customizing your kahoots for your work setting.
[Choose Kahoot! 360 Pro Max](#)
- Meet Kahoot! Kids!**
Spark your child's curiosity for learning with our new playful app experience.
[Get started today](#)

Models: which kind of knowledge could we model and make use



Medien anzeigen

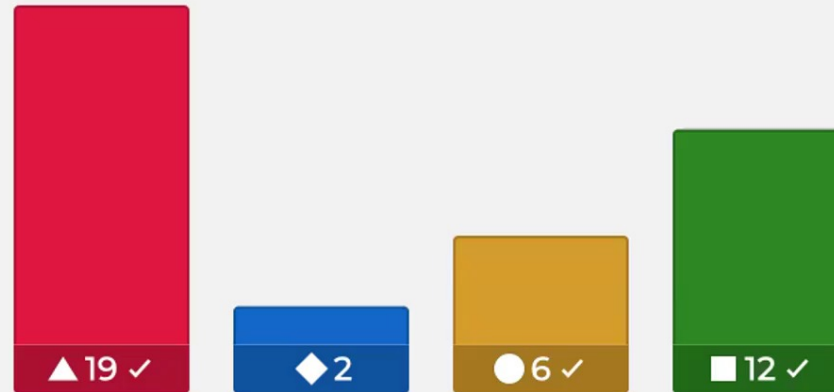
▲ Rewards ✓

◆ Inverse Dynamics ✓

● Policy ✗

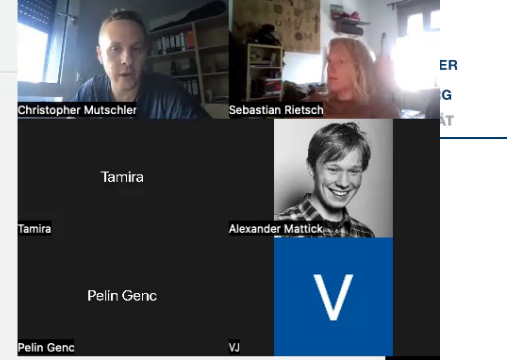
■ Transition Probabilities ✓

Policy Backpropagation...

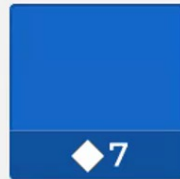


Medien anzeigen

- ▲ uses Backpropagation through time (BPTT) to calculate the policy gradient ✓
- ◆ Explicitly addresses compounding errors in large prediction horizons ✗
- can be engineered to take long term credit assignment into account ✓
- is prone to local minima ✓



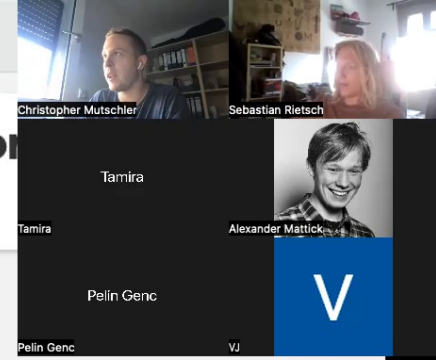
"Background Planning" finds / optimizes for the best sequence of actions for current situation



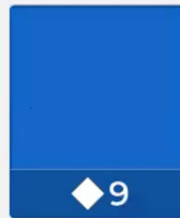
Medien anzeigen

◆ Wahr ✕

▲ Falsch ✓



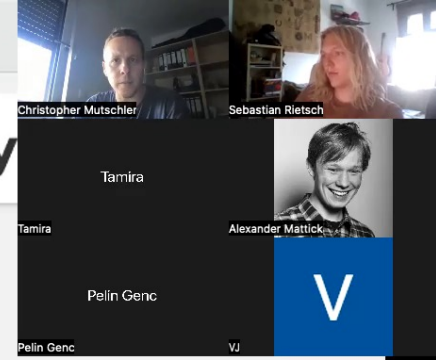
Decision-time planning is computationally lightweight at deploy



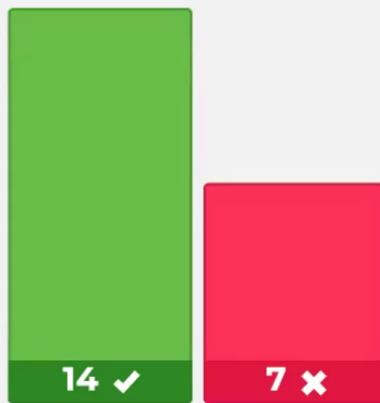
Medien anzeigen

◆ Wahr ✕

▲ Falsch ✓



Monte Carlo Tree Search: sort the steps within an iteration

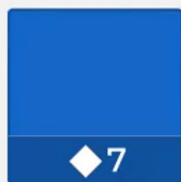
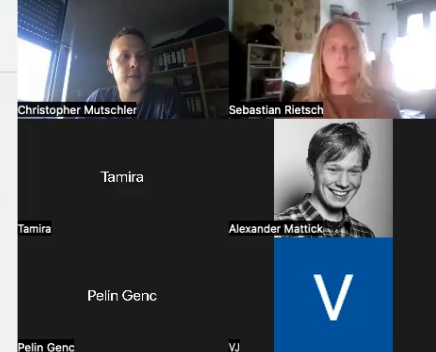


Medien anzeigen

- Select ✓
- Expand ✓
- Simulate ✓
- Backup ✓

Sebastian Rietsch
Christopher Mutschler
Tamira
Pelin Genc
Alexander Mattick
VJ

MCTS/Selection: Exploration is done via ϵ -greedy

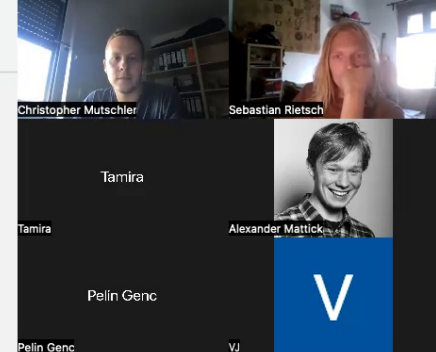


Medien anzeigen

Wahr

Falsch

MCTS/Simulation: what is correct?



▲ 11

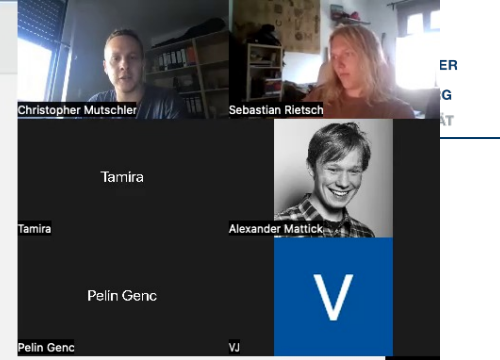
◆ 10 ✓

Medien anzeigen

▲ We follow our tree policy ✕

◆ We follow a random policy ✓

MCTS/Simulation: what is correct?



▲ 1

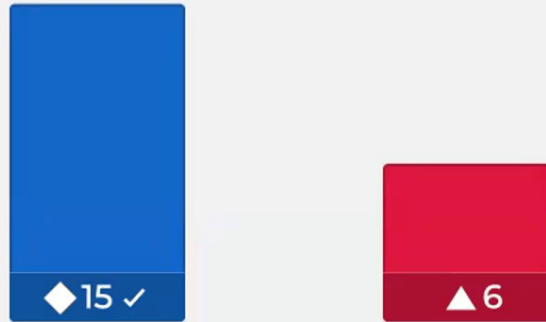
◆ 19 ✓

Medien anzeigen

▲ We choose a child node and run a single trajectory ✕

◆ We choose a child node and run trajectories until a time budget is reached ✓

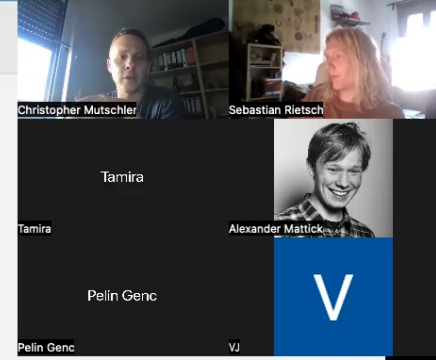
Backup: improves the selection policy



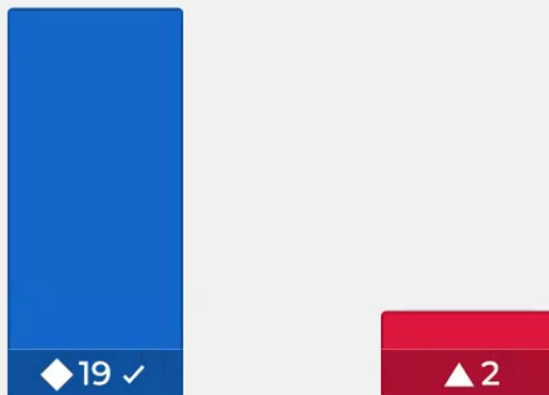
Medien anzeigen

◆ Wahr ✓

▲ Falsch ✗



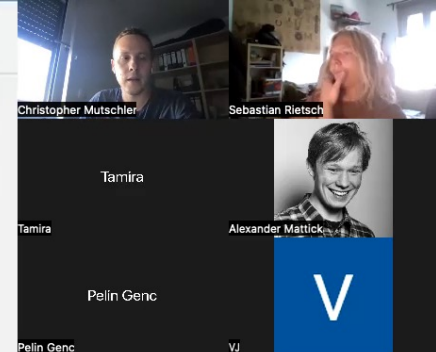
AlphaGo uses a fast rollout policy for simulations



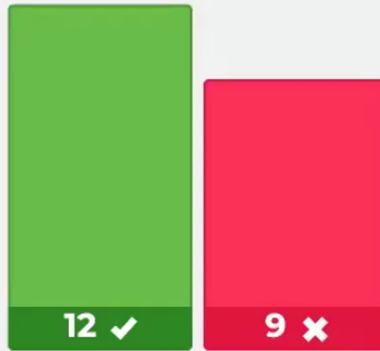
Medien anzeigen

◆ Wahr ✓

▲ Falsch ✗



Sort the Evolution of AlphaGo (top: oldest, bottom: recent)



Medien anzeigen

● AlphaGo ✓

▲ AlphaGo Zero ✓

◆ AlphaZero ✓

■ MuZero ✓

A Zoom meeting interface showing several participant thumbnails. Visible names include Sebastian Rietsch, Christopher Mutschler, Tamira, Pelin Genc, Alexander Mattick, and VJ. A large white 'V' is overlaid on the bottom right thumbnail.