AlphaTensor

Рецензия

Антон Барышников

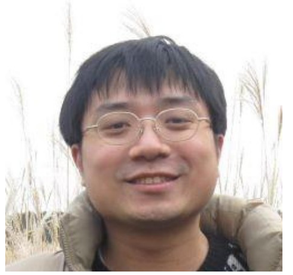
AlphaTensor

Recap

- We want to discover fast matrix multiplication algorithms.
- Let’s represent matrix multiplication as a game.
- And run AlphaZero on it.

Authors

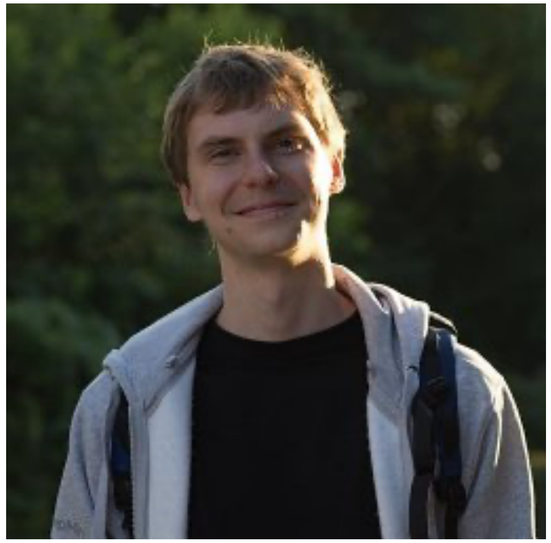





- Alhussein Fawzi (EPFL PhD):
  - ML for mathematics.
  - Adversarial robustness.
- Matej Balog (Cambridge-Tubingen PhD):
  - Program synthesis (DeepCoder).
- Aja Huang (NTNU PhD):
  - Reinforcement learning (AlphaGo, AlphaStar).
- Thomas Hubert (Stanford MS, Financial Mathematics):
  - Reinforcement learning (AlphaGo, AlphaZero).

Authors

- Alexander Novikov (Skolkovo PhD, Bayes Group):
  - Reinforcement Learning.
  - Tensor Decompositions.

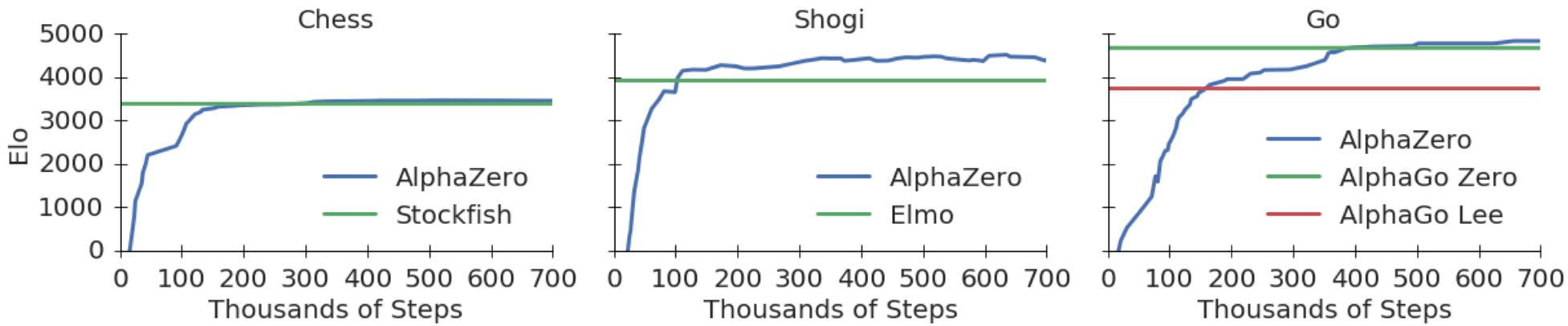


p(B|A)yesgroup.ru

AlphaZero, 2017

Recap

- Key to the success of AlphaTensor.
- Generalist algorithm for turn-based games.
- Deep RL + MCTS.



Strassen, 1969

Before AlphaTensor

- Novel algorithm for matrix multiplication.
- 2x2 matrices require 7 operations instead of 8.
- Long time (**50 years**) best for 4x4 over a finite field.
- AT **outperforms** Strassen's algorithm on 4x4.

Size (n, m, p)	Best method known	Best rank known	AlphaTensor rank Modular	AlphaTensor rank Standard
$(2, 2, 2)$	(Strassen, 1969)[2]	7	7	7
$(3, 3, 3)$	(Laderman, 1976)[15]	23	23	23
$(4, 4, 4)$	(Strassen, 1969)[2] $(2, 2, 2) \otimes (2, 2, 2)$	49	47	49
$(5, 5, 5)$	$(3, 5, 5) + (2, 5, 5)$	98	96	98

Other Human Search Approaches

Before AlphaTensor

- Laderman (1976) — best for 3x3.
- Hopcroft and Kerr (1971) — best for a set of small matrices.
- Hand-crafted algorithms.
- Matched by AT.

Size (n, m, p)	Best method known	Best rank known	AlphaTensor rank Modular	AlphaTensor rank Standard
$(2, 2, 2)$	(Strassen, 1969)[2]	7	7	7
$(3, 3, 3)$	(Laderman, 1976)[15]	23	23	23
$(4, 4, 4)$	(Strassen, 1969)[2] $(2, 2, 2) \otimes (2, 2, 2)$	49	47	49
$(5, 5, 5)$	$(3, 5, 5) + (2, 5, 5)$	98	96	98

Other Methods

Before AlphaTensor

- Relatively new: 2013 and 2021.
- Continuous optimisation (Smirnov, Sedoglavic):
  - Rewrite the problem and solve continuously.
  - **Some results beaten by AT.**
- Combinatorial search (Heule, Kauers, Seidl):
  - Convert problem to SAT.
  - Do combinatorial optimisation.

$$\underset{P_i, Q_i, S_i}{\operatorname{Arg\ min}} \left\| \sum_{i=0}^{q} P_i \otimes Q_i \otimes S_i - \sum_{i=0}^{m} \sum_{j=0}^{n} \sum_{k=0}^{p} E_i^j \otimes E_j^k \otimes E_k^i \right\| + \lambda \left(\sum_{i=0}^{q} \left\| P_i - \widetilde{P_i} \right\| + \left\| Q_i - \widetilde{Q_i} \right\| + \left\| S_i - \widetilde{S_i} \right\| \right).$$

(2, 3, 3)	(Hopcroft and Kerr, 1971)[16]	15	15	15
(2, 3, 4)	(Hopcroft and Kerr, 1971)[16]	20	20	20
(2, 3, 5)	(Hopcroft and Kerr, 1971)[16]	25	25	25
(2, 4, 4)	(Hopcroft and Kerr, 1971)[16]	26	26	26
(2, 4, 5)	(Hopcroft and Kerr, 1971)[16]	33	33	33
(2, 5, 5)	(Hopcroft and Kerr, 1971)[16]	40	40	40
(3, 3, 4)	(Smirnov, 2013)[18]	29	29	29
(3, 3, 5)	(Smirnov, 2013)[18]	36	36	36
(3, 4, 4)	(Smirnov, 2013)[18]	38	38	38
(3, 4, 5)	(Smirnov, 2013)[18]	48	47	47
(3, 5, 5)	(Sedoglavic and Smirnov, 2021)[19]	58	58	58

The FBHHRBNRSSSHK-Algorithm for Multiplication is still not the end of the story.

- Published after AlphaTensor (Kauers and Moosbauer, 2022).
- **Outperforms** AT on 5x5 matrices (from 96 to 95).
- Matches AT on 4x4 over finite field.
- Method unknown.

RL is dead long live RL

Other areas

- Video compression:
  - MuZero (2019) used in YouTube compression.
- Job Scheduling:
  - Zhang, Diettenh (1995).
  - Tassel, Gebser, Schekotihin (2021).
- Compiler optimisations:
  - Finding new hash functions (anecdote, not published).