CS4400
DEEP REINFORCEMENT LEARNING

Lecture 13: Course Summary

Wendelin Böhmer
<j.w.bohmer@tudelft.nl>

TU Delft

23rd of January 2024

Content of this lecture

13.1

Course Summary

Course summary

13.1 Course summary

- Modern neural network architectures
  - modules represent structural constraints (equivariances)
  - implicit equivariances mostly responsible for generalization
- Modern deep reinforcement learning methods
  - on-policy, off-policy, offline, multi-agent, exploration
  - how to implement the core parts
- Underlying deep reinforcement learning concepts
  - continuous actions, partial observations
  - value propagation, robustness, deep exploration
  - how to formalize, derive and prove these things
- When reinforcement learning fails
  - instability, distribution shift, value bias, catastrophic forgetting
  - random expl., relative overgen., communication, zero-shot coord.
  - sim2real transfer, out-of-distribution generalization

13.2

Course Summary

Beyond this course

13.2 Is reinforcement learning realistic?

- Standard setup: train and test in stationary environment
  - online learning makes overfitting impossible
  - learns after many many environmental interactions
  - but does it *generalize* or *memorize*?

Is reinforcement learning realistic?

- Standard setup: train and test in stationary environment
  - online learning makes overfitting impossible
  - learns after many many environmental interactions
  - but does it *generalize* or *memorize*?
- Real world applications differ significantly
  - finite training samples
  - catastrophic actions
  - observation noise
  - varying background activity
  - one agent multiple tasks

13.2 How would "realistic RL" look like?

- Conventional machines
  - robustness/multi-functionality/generalization
  - guarantees/constraints/ethics
  - reliability/explainability/responsibility

  → application breakthroughs within 10 years?

13.2 How would "realistic RL" look like?

- Conventional machines
  - robustness/multi-functionality/generalization
  - guarantees/constraints/ethics
  - reliability/explainability/responsibility

  → application breakthroughs within 10 years?

- Biology-like machines
  - life-long learning: keeps learning in the wild
  - learning without tasks and without terminal states
  - develops abstractions and switches between them

  → basic terminology develops right now

13.2 How would “realistic RL” look like?

- Conventional machines
  - robustness/multi-functionality/generalization
  - guarantees/constraints/ethics
  - reliability/explainability/responsibility

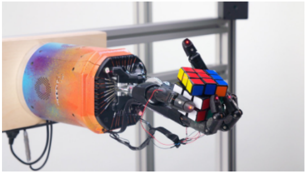
  → application breakthroughs within 10 years?
- Biology-like machines
  - life-long learning: keeps learning in the wild
  - learning without tasks and without terminal states
  - develops abstractions and switches between them

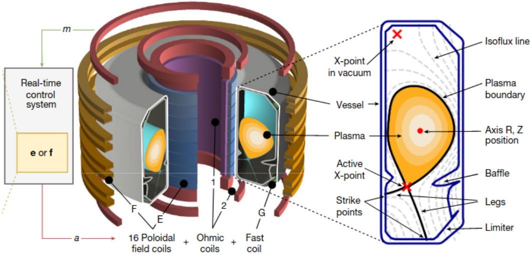
  → basic terminology develops right now
- Sentience (artificial general intelligence)
  - merges sequential (symbolic) and reflexive (pattern recognition) AI

  → we don’t even know what questions to ask

The coolest deep RL applications

- Atari (Mnih et al., 2013, 2015, and many, many follow-ups)
- Go, Shogi, Chess (Silver et al., 2016, 2017, 2018)
- StarCraft II (Vinyals et al., 2019)
- MOBAs (Berner et al., 2019; Ye et al., 2020)
- Robotic Rubik's cube (OpenAI et al., 2019)
- Traffic signals (Cabrejas-Egea et al., 2021)
- Fusion reactors (Degrave et al., 2022)
- Parkour robot (Cheng et al., 2023)





13.2 What do we need to work on?

- Data efficiency
  - breakthroughs in offline RL will allow RL in robotics soon
  - model-based RL, network architecture, Bayesian optimization
- Safety
  - uncertainty, Bayesian RL, safe exploration
  - constrained RL, interpretable RL, formal verification
- Generalization
  - network architectures, inference from multiple abstractions
- Lifelong learning
  - learning w/o task boundaries, self-motivated learning
- Social agents
  - zero-shot coordination, interaction with humans, communication

13.2 What to do next?

- RL research @ TU-Delft

SDM Wendelin Böhmer: anything deep RL
SDM Matthijs Spaan: RL and uncertainty
SDM Frans Oliehoek: Bayesian RL, model-based RL, multi-agent RL
ALG Anna Lukina: verifiable ML & RL
INSY Luciano Siebert: inverse RL for responsible and ethical AI
INSY Pradeep Murukannaiah: interactive AI
INSY Catharine Oertel: dialogue AI
3ME Jens Kober: RL from demonstrations for robots
3ME Javier Alonso-Mora: RL for robot motion planning
3ME Laura Ferranti: RL for reliable robot control
AS/EWI BIOlab: AI and RL for neuroscience & biomedical applications

- Get in touch with RL @ TU-Delft: reinforceAI.net
  - CS4210-B: AIDM Project (Q4)
  - CS4345: formal methods for learned system (Q3)
  - CS4240: deep learning (no RL, Q3)
  - RL reading group Thursdays 15:00 [mattermost]
  - ELLIS unit: get involved with Delft's AI community

Next lecture

- This is the last lecture!
- Ask questions now!
- Shy? Ask questions here: answers.ewi.tudelft.nl

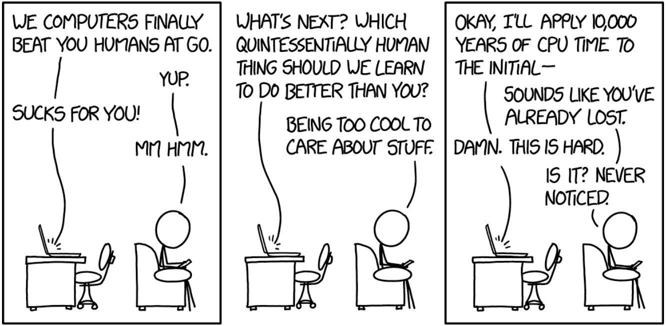


image source: xkcd.com

References I

Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemyslaw Debiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Christopher Hesse, Rafal Józefowicz, Scott Gray, Catherine Olsson, Jakub Pachocki, Michael Petrov, Henrique Pondé de Oliveira Pinto, Jonathan Raiman, Tim Salimans, Jeremy Schlatter, Jonas Schneider, Szymon Sidor, Ilya Sutskever, Jie Tang, Filip Wolski, and Susan Zhang. Dota 2 with large scale deep reinforcement learning. *CoRR*, abs/1912.06680, 2019. URL http://arxiv.org/abs/1912.06680.

Alvaro Cabrejas-Egea, Raymond Zhang, and Neil Walton. Reinforcement learning for traffic signal control: Comparison with commercial systems. *CoRR*, abs/2104.10455, 2021. URL https://arxiv.org/abs/2104.10455.

Xuxin Cheng, Kexin Shi, Ananye Agarwal, and Deepak Pathak. Extreme parkour with legged robots, 2023. URL https://arxiv.org/abs/2309.14341.

Jonas Degrave, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, Timo Ewalds, Roland Hafner, Abbas Abdolmaleki, Diego Casas, Craig Donner, Leslie Fritz, Cristian Galperti, Andrea Huber, James Keeling, Maria Tsimpoukelli, Jackie Kay, Antoine Merle, Jean-Marc Moret, and Martin Riedmiller. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature*, 602:414–419, 02 2022. doi: 10.1038/s41586-021-04301-9.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013. URL http://arxiv.org/abs/1312.5602. NIPS Deep Learning Workshop 2013.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015.

OpenAI, Ilge Akkaya, Marcin Andrychowicz, Maciek Chociej, Mateusz Litwin, Bob McGrew, Arthur Petron, Alex Paino, Matthias Plappert, Glenn Powell, Raphael Ribas, Jonas Schneider, Nikolas Tezak, Jerry Tworek, Peter Welinder, Lilian Weng, Qiming Yuan, Wojciech Zaremba, and Lei Zhang. Solving rubik's cube with a robot hand. *CoRR*, abs/1910.07113, 2019. URL http://arxiv.org/abs/1910.07113.

References II

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, pages 484–489, 2016. doi: 10.1038/nature16961.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of go without human knowledge. *Nature*, 550:354–359, October 2017. URL http://dx.doi.org/10.1038/nature24270.

David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018. URL https://science.sciencemag.org/content/362/6419/1140.

Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H. Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, Laurent Sifre, Trevor Cai, John P. Agapiou, Max Jaderberg, Alexander S. Vezhnevets, Rémi Leblond, Tobias Pohlen, Valentin Dalibard, David Budden, Yury Sulsky, James Molloy, Tom L. Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Yuhuai Wu, Roman Ring, Dani Yogatama, Dario Wünsch, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Koray Kavukcuoglu, Demis Hassabis, Chris Apps, and David Silver. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575:350–354, 2019.

Deheng Ye, Zhao Liu, Mingfei Sun, Bei Shi, Peilin Zhao, Hao Wu, Hongsheng Yu, Shaojie Yang, Xipeng Wu, Qingwei Guo, Qiaobo Chen, Yinyuting Yin, Hao Zhang, Tengfei Shi, Liang Wang, Qiang Fu, Wei Yang, and Lanxiao Huang. Mastering complex control in MOBA games with deep reinforcement learning. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI 2020), The Thirty-Second Innovative Applications of Artificial Intelligence Conference (IAAI 2020), The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI 2020)*, pages 6672–6679, 2020. URL https://arxiv.org/abs/1912.09729.