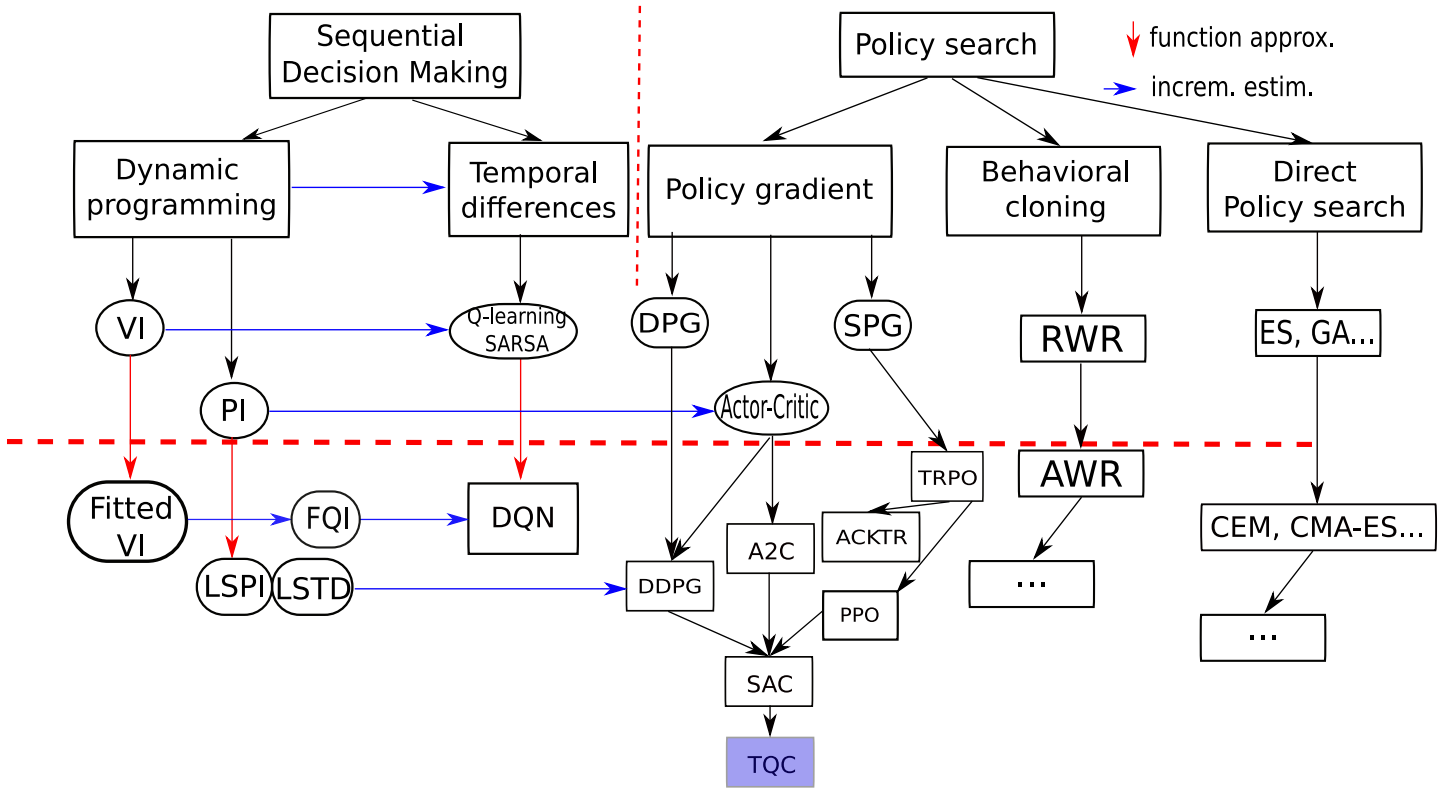
Deep Reinforcement Learning

Wrap-up, Take Home Messages

Olivier Sigaud

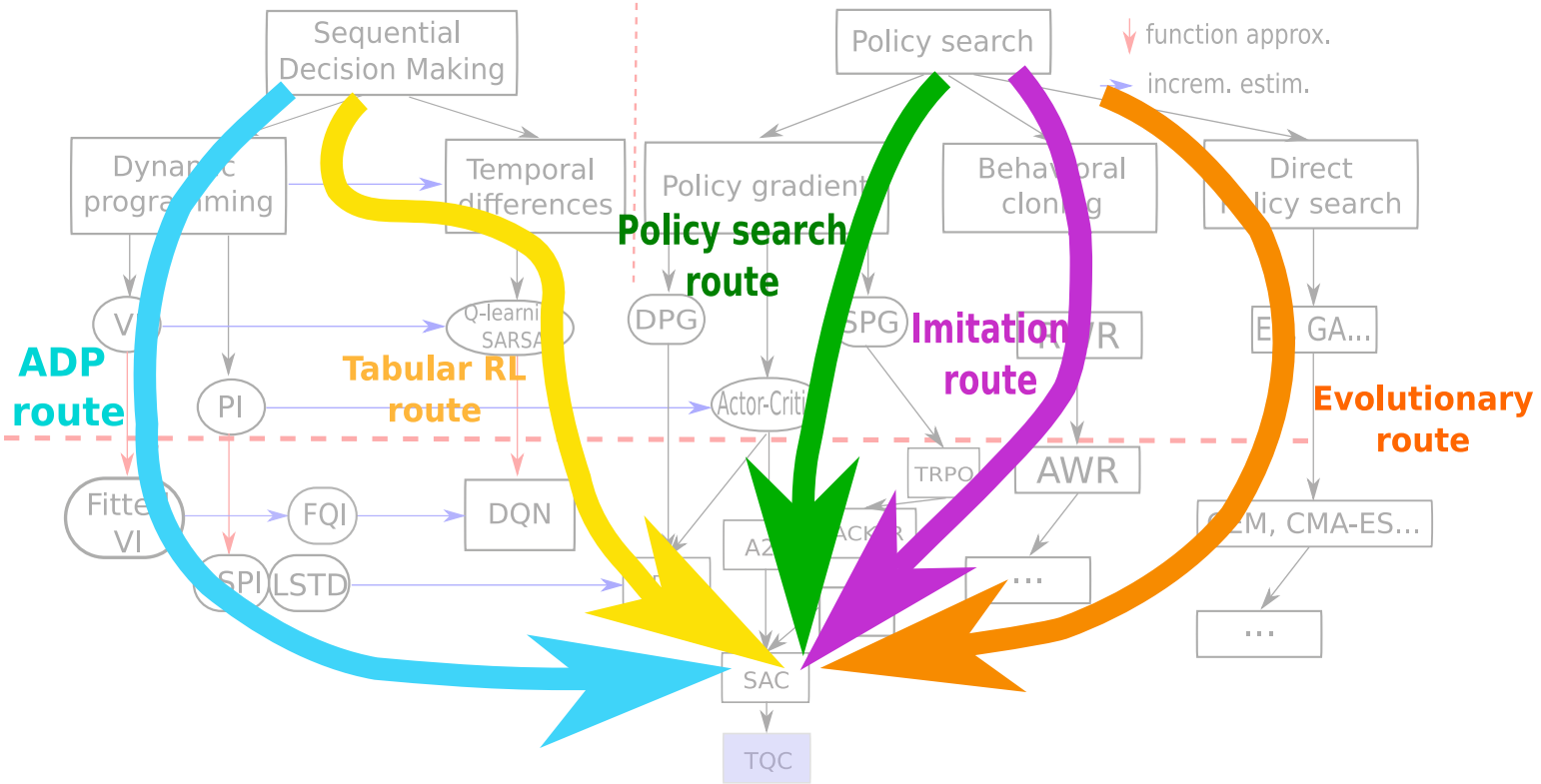
Sorbonne Université
http://people.isir.upmc.fr/sigaud

The Big Picture



- A very partial view of the whole RL literature

The five routes to deep RL

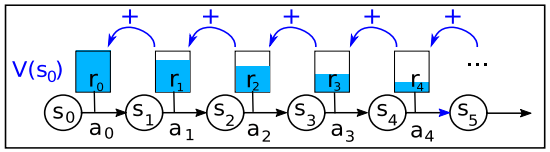


- Five different ways to come to Deep RL

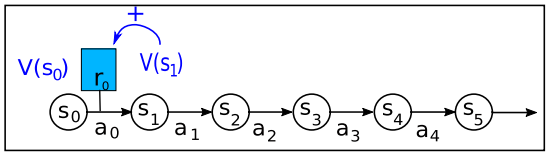
Key Policy Gradient Steps

- 1. Splitting the trajectory into steps: Markov Hypothesis required
- Key difference to Direct Policy Search methods
- Makes it possible to optimize trajectories using a gradient over policy params
- 2. Introducing the an estimator of the V/Q function
- Makes it possible to perform policy updates from a single step
- Opens the way to the replay buffer, critic networks, partly off-policy methods
- 3. Using baselines
- Makes it possible to reduce variance
- When learning critics from bootstrap, becomes actor-critic

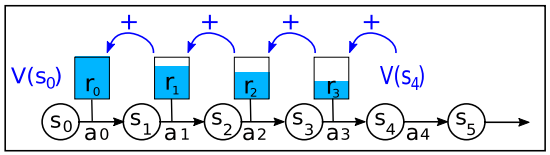
Bias-variance, Being Off-policy



Monte Carlo

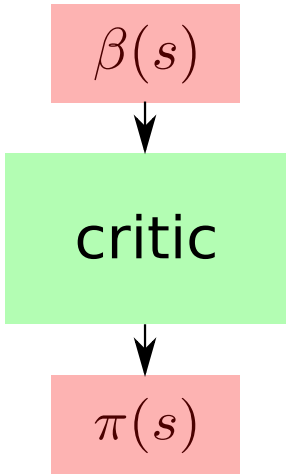


One-step TD



N-step TD

λ return



- Continuum between Monte Carlo methods and bootstrap methods
- Playing on the continuum helps finding the right bias-variance trade-off
- Using a replay buffer requires bootstrap
- Using a replay buffer requires being partly off-policy
- No deep RL algorithm is truly off-policy, it's a matter of degree

Marcin Andrychowicz, Anton Raichuk, Piotr Stańczyk, Manu Orsini, Sertan Girgin, Raphael Marinier, Léonard Hussenot, Matthieu Geist, Olivier Pietquin, Marcin Michalski, et al. (2020) What matters in on-policy reinforcement learning? a large-scale empirical study. *arXiv preprint arXiv:2006.05990*

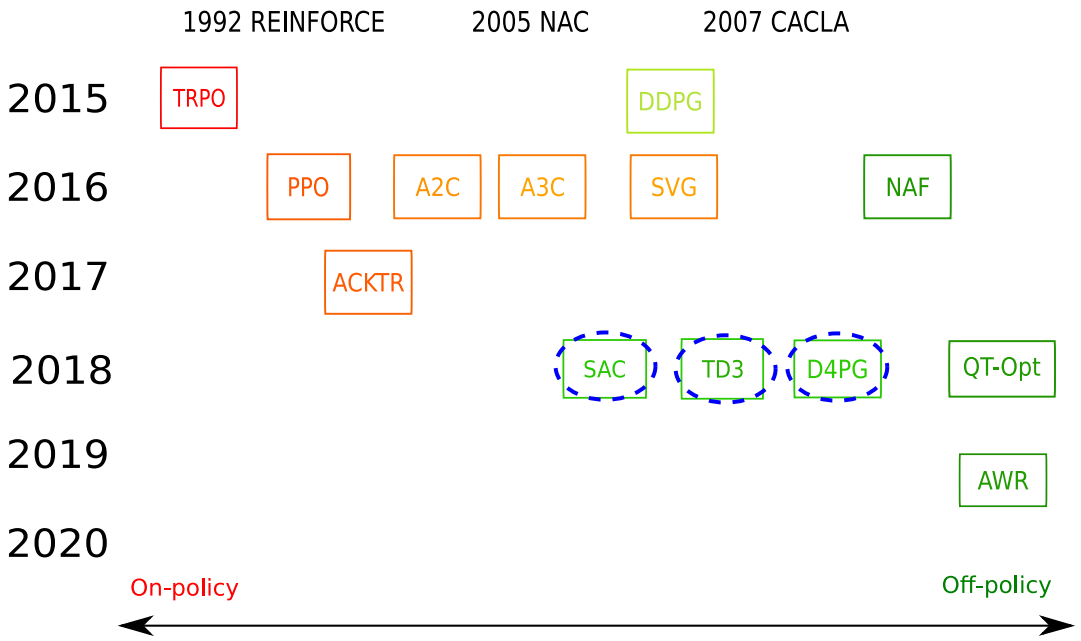
The deadly triad

- RL algorithms combining function approximators, off-policy learning and bootstrap can run instable
- Studied in DQN (does not occur much), but also true for continuous actions
- Three options: no function approximators (tabular RL), being on-policy (TRPO, PPO, ACKTR, A3C), no bootstrap (AWR)
- Being on-policy, TRPO, PPO are not sample efficient
- Without bootstrap, AWR offers an original solution

Richard Sutton. Introduction to reinforcement learning with function approximation. In *Tutorial at NIPS*, 2015

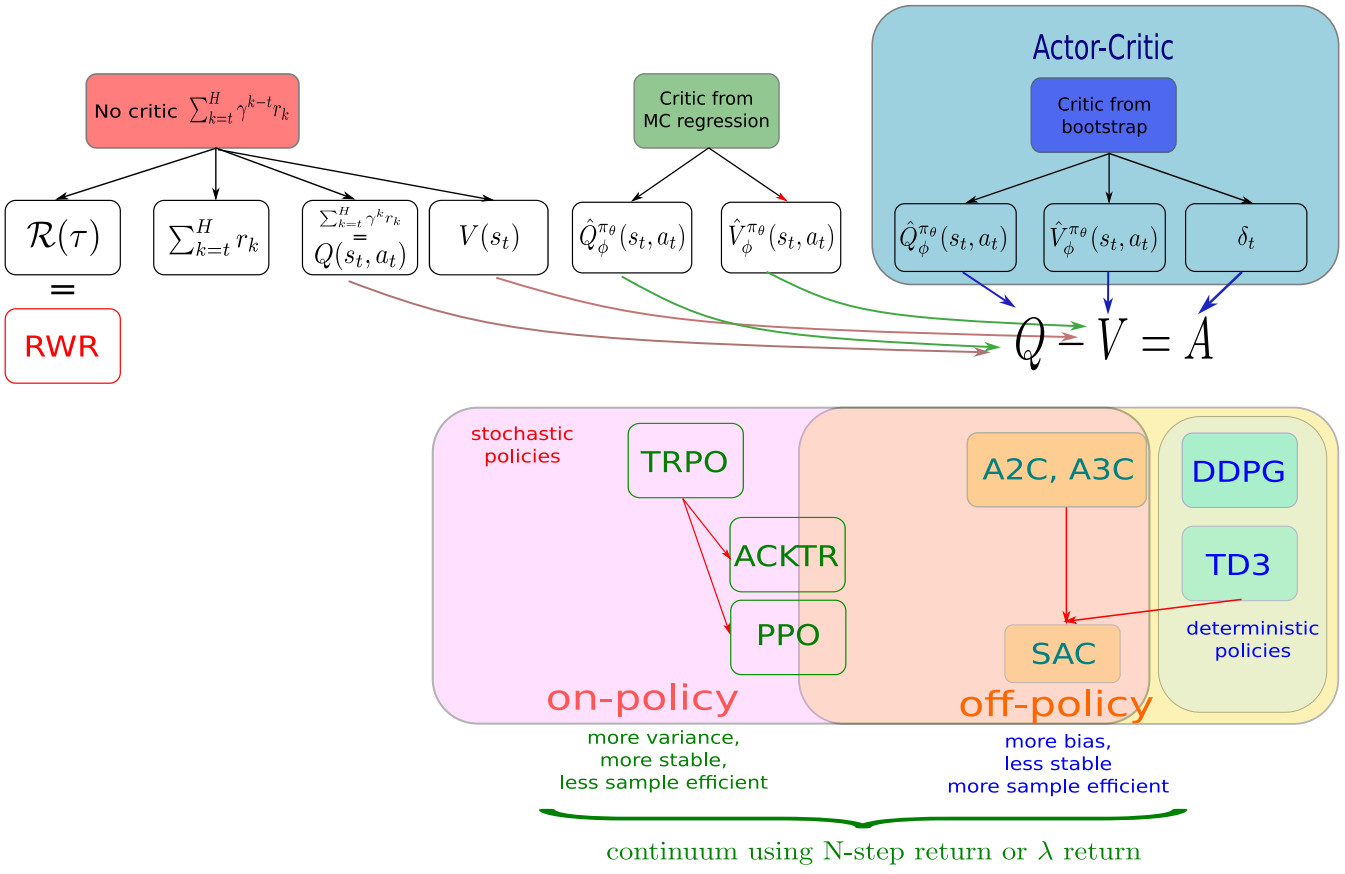
Hado van Hasselt, Yotam Doron, Florian Strub, Matteo Hessel, Nicolas Sonnerat, and Joseph Modayil. Deep Reinforcement Learning and the Deadly Triad. *arXiv preprint arXiv:1812.02648*, 2018

Ranking per year



- It all started in 2015, 2018 was a major year
- Since 2019, the focus is more on other questions: multitask RL, intrinsic motivations, exploration, meta-RL...

Final view



- Many algorithms have not been covered: CACLA, ACER, soft-DQN, C51, D4PG, ...

Status

- Large computational ressources are necessary
- Good engineers help a lot
- Grad student descent
- Big actors are ruling the game: Deepmind, OpenAI, Berkeley, Microsoft, FAIR...
- Focus more on performance than on understanding
- Deep RL that matters: instabilities, hard to compare, sensitivity to hyper-parameters
- Empirical comparisons based mostly on openAI gym, mujoco, Deepmind Control Suite
- The reproducibility crisis, and challenge
- Lack of controlled experiments
- Still fast performance progress, but progress is now more in exploration, multitask learning, curriculum learning, offline learning, using transformers, etc.


Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, Doina Precup, and David Meger (2017) Deep reinforcement learning that matters. *arXiv preprint arXiv:1709.06560*


Artemij Amiranashvili, Alexey Dosovitskiy, Vladlen Koltun, and Thomas Brox (2018) TD or not TD: Analyzing the role of temporal differencing in deep reinforcement learning. In *International Conference on Learning Representations (ICLR)*.

Any question?

Send mail to: Olivier.Sigaud@upmc.fr

Amiranashvili, A., Dosovitskiy, A., Koltun, V., and Brox, T. (2018).

TD or not TD: analyzing the role of temporal differencing in deep reinforcement learning.
In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

Andrychowicz, M., Raichuk, A., Stańczyk, P., Orsini, M., Girgin, S., Marinier, R., Hussenot, L., Geist, M., Pietquin, O., Michalski, M., et al. (2020).

What matters in on-policy reinforcement learning? a large-scale empirical study.
arXiv preprint arXiv:2006.05990.

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., and Meger, D. (2018).

Deep reinforcement learning that matters.
In McIlraith, S. A. and Weinberger, K. Q., editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 3207–3214. AAAI Press.

Sutton, R. (2015).

Introduction to reinforcement learning with function approximation.
In *Tutorial at NIPS*.

van Hasselt, H., Doron, Y., Strub, F., Hessel, M., Sonnerat, N., and Modayil, J. (2018).

Deep Reinforcement Learning and the Deadly Triad.
arXiv preprint arXiv:1812.02648.