Automatic Serial Femtosecond Crystallography Online Analysis with Reinforcement Learning

D. Ferreira de Lima   A. Davtyan   L. Gelisio

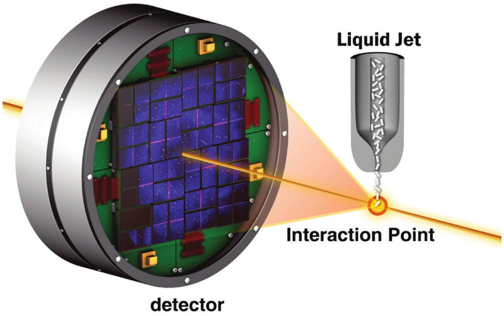
European XFEL

15 October 2021

Serial Femtosecond Crystallography analysis

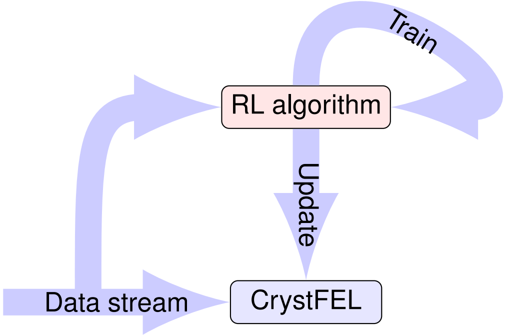
- Serial Femtosecond Crystallograph's pipeline:
  - Pre-select relevant frames.
  - Identify Bragg peaks.
  - Identify Miller indices corresponding to peaks and crystal orientation.
  - Integrate intensity of the Bragg peaks.
- Requires tuning many parameters depending on the sample and experimental conditions.
  - Can we automatize the parameter tuning?

CrystFEL



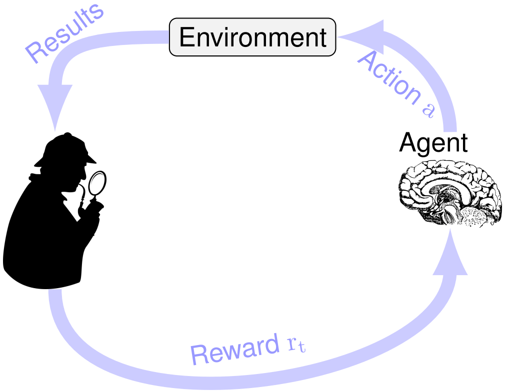
(Image from the SPB TDR, TR-2013-004)

Automatizing SFX



- Collect a sample of the incoming data to search optimal parameters.
- While the parameter search runs, the standard pipeline runs in parallel.
  - The parameter search explores randomly other possible parameter values.
- When improved parameters are found, update the parameters in the pipeline to improve the results as data is taken.

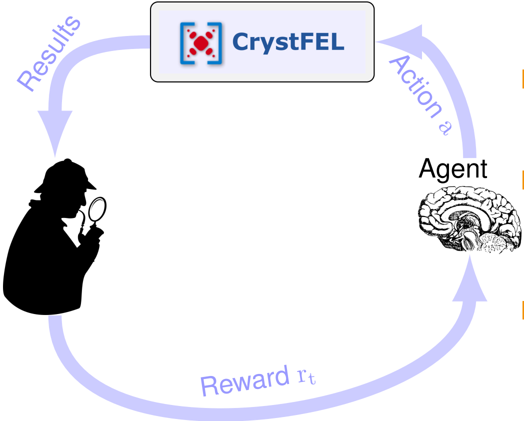
Reinforcement Learning



- Environment $\rightarrow$ analysis pipeline.
- Model-free RL $\rightarrow$ environment-independent.
  - Input = last + noise (*perform action*).
  - Was there an improvement (*collect reward*)?
  - Update agent.
- Objective: maximize total returns G after T attempts.
  - At $t = T$, reset state to the best.

$$G = \sum_{t=1}^{T} \gamma^t r_t$$

Online SFX tuning



- What is the environment?
  - The environment is the algorithm to be optimized: CrystFEL.
- What are the actions?
  - Actions are changes in the parameter values set by CrystFEL.
  - A current state (parameter set) is kept.
- What are the rewards?
  - Change in fraction of indexed frames.
  - Positive reward means the number of indexed frames increased.

What is the agent?

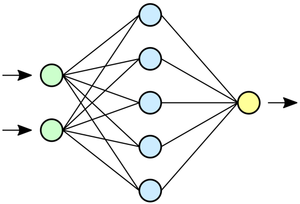
- The agent is just a parametrized function.
  - It takes the current parameter set and outputs the change in parameter values needed to optimize the return.
  - We fit the parameters of the function.
- Neural networks are universal function approximators.

Theorem 1. *Let σ be any continuous discriminatory function. Then finite sums of the form*

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j) \qquad (2)$$

are dense in $C(I_n)$. In other words, given any $f \in C(I_n)$ and $\varepsilon > 0$, there is a sum, $G(x)$, of the above form, for which

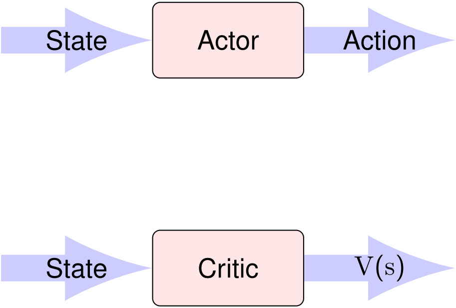
$$|G(x) - f(x)| < \varepsilon \qquad \text{for all} \quad x \in I_n.$$



G. Cybenko, Math. Control Signals Systems (1989) **2**, 303 – 314

Actor-critic models

- One possible RL setup: actor-critic model.
- After randomly exploring actions around the current optimal:
  - Critic $\mathrm{C(s)}$ fits the obtained returns G, given the state s.
    - Advantage function definition: $\mathrm{A(s) = C(s) - G}$
    - Critic minimizes $\mathcal{L}_\mathrm{C} = [\mathrm{A(s)}]^2$
  - Actor optimizes returns, based on the critic's predictions.



Advantage Actor Critic (A3C/A2C) and ACKTR

- How to optimize the actor?
  - Maximize returns $\rightarrow$ move parameters in the direction of $\nabla_\theta \mathbb{E}_{\pi(\tau|\theta)}\left[\mathrm{G}(\tau)\right]$.
- Williams (1992) showed that:
  - $\mathbb{E}_{\pi(\tau|\theta)}\left[\nabla_\theta \log \pi(\tau|\theta)\mathrm{G}(\tau)\right]$ is an unbiased estimator for $\nabla_\theta \mathbb{E}_{\pi(\tau|\theta)}\left[\mathrm{G}(\tau)\right]$.
  - Probability of taking a set of actions $\tau = \{(\mathrm{s}_1, \mathrm{a}_1), \ldots, (\mathrm{s}_n, \mathrm{a}_n)\}$ is $\pi(\tau)$.
  - Parameters of the neural network θ.
- In A3C: substitute G with A to reduce variance.
  - ACKTR: Use second-order optimization algorithms to improve convergence speed.

Machine Learning, 8, 229-256 (1992)
© 1992 Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.

Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning

RONALD J. WILLIAMS rjw@corwin.ccs.northeastern.edu
College of Computer Science, 161 CN, Northeastern University, 360 Huntington Ave., Boston, MA 02115

Asynchronous Methods for Deep Reinforcement Learning

Volodymyr Mnih[1] VMNIH@GOOGLE.COM
Adrià Puigdomènech Badia[1] ADRIAP@GOOGLE.COM
Mehdi Mirza[1,2] MIRZAMOM@IRO.UMONTREAL.CA
Alex Graves[1] GRAVESA@GOOGLE.COM
Tim Harley[1] THARLEY@GOOGLE.COM
Timothy P. Lillicrap[1] COUNTZERO@GOOGLE.COM
David Silver[1] DAVIDSILVER@GOOGLE.COM
Koray Kavukcuoglu[1] KORAYK@GOOGLE.COM
[1] Google DeepMind
[2] Montreal Institute for Learning Algorithms (MILA), University of Montreal

Scalable trust-region method for deep reinforcement learning using Kronecker-factored approximation

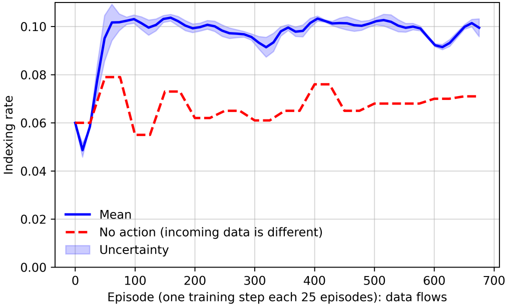
Yuhuai Wu[*] University of Toronto Vector Institute ywu@cs.toronto.edu
Elman Mansimov[*] New York University mansimov@cs.nyu.edu
Shun Liao University of Toronto Vector Institute sliao3@cs.toronto.edu

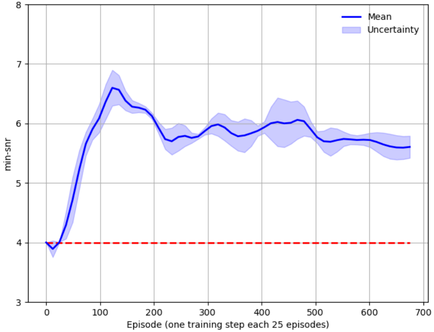
Tests

- Tested the method with a dataset of:
  - hen egg white lysozyme;
  - collected on an AGIPD detector at the SPD experiment.
- Used a cache of 1000 images to explore parameter space in the RL algorithm.
- Optimization on 10 parallel cores.
- RL algorithm takes only $\sim$ 1 second out of 1 minute: most time spent on the analysis pipeline itself.

Results

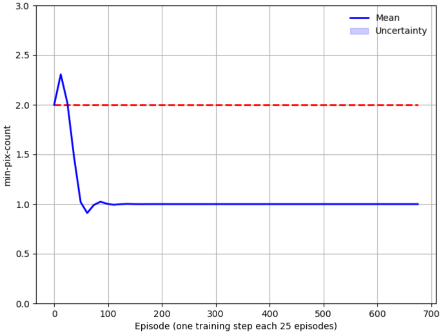
- Indexed fraction increases when compared to maintaining the initial setup.
- The minimum signal-to-noise ratio parameter is tuned away from the initial value.

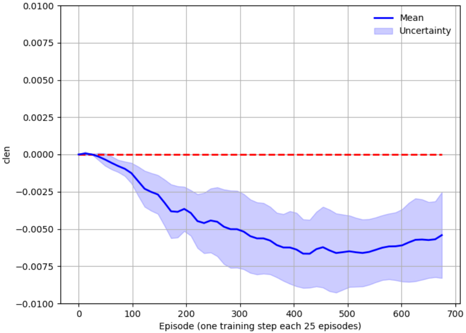




Other tuned variables

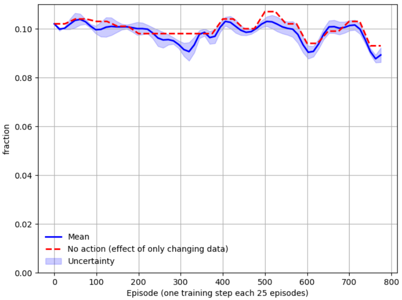
- The detector to sample distance and minimum number of hits can also be tuned.
- Uncertainty band: root-mean-squared-error of 4 random NN initializations.

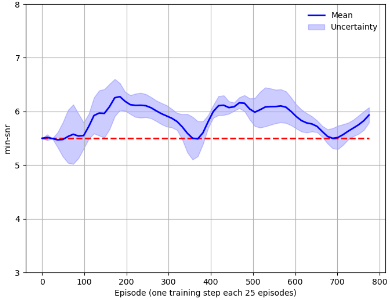




Stability test

- What if we already start close to the ideal conditions?
- Focus on speed $\rightarrow$ using a small number of images to explore parameters.
  - Could increase the number of frames for smaller uncertainty on the rewards.





Summary

- Apply on-policy reinforcement learning algorithm to optimize SFX pipeline online.
  - Clear goal: maximize number of indexed frames.
- Main concept:
  - Set initial parameters and try to change it randomly.
  - Fit the obtained rewards with the critic.
  - Fit actor to optimize rewards.
  - Automatically adapt parameters as the data flows.
- System tested offline with further tests on new datasets coming.