Bayesian Optimization With Asynchronous Expert Feedback

Agustinus Kristiadi

VECTOR INSTITUTE

Motivation

- Bayesian optimization $\max_{x \in \mathcal{X}} f(x)$
- Human experts might have good "prior" over the problem:
  - In material discovery, chemists have "chemical intuition" on whether a molecule x is preferred.
  - Even though x achieves high $f(x)$, it might not be preferable, e.g. because it's hard synthesize or problematic in terms of regulation or just doesn't feel right
- But this "expert intuition" is very vague. How to incorporate it into BO?
  - How do we even define the prior distribution?
  - How do we even write down the constraint functions?
  - How do we even create a supervised learning dataset?

Desiderata

- Expert feedback asynchronously collected throughout the BO loop
  - Asynchronous $\implies$ non-blocking to the BO campaign; user-friendly for experts
- Modularly built on top of standard BO
  - The expert-feedback component should not interfere the BO loop
  - No feedback given $\implies$ revert back to the standard BO
- The feedback is used to steer the BO's exploration-exploitation
  - With a tuneable discount factor
  - In one limit, we have the standard BO. In other limit, we have preferential BO (Gonzalez et al., ICML 2017)
- Anything else?

Current Idea

Algorithm BO Thread

Input: Initial dataset $\mathcal{D}_1 = \{(x_i, f(x_i))\}_i$, surrogate g, feedback-aware acqf. α

for $t = 1, \ldots, T$ **do**

  Infer $p(g_t \mid \mathcal{D}_t)$

  $x_{\text{next}} = \arg\max_{x \in \mathcal{X}} \alpha(p(g_t(x) \mid \mathcal{D}_t), p(r \mid \mathcal{D}_t^{\text{feedback}}))$

  Compute $f(x_{\text{next}})$

  $\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{(x_{\text{next}}, f(x_{\text{next}}))\}$

end for

return $\arg\max_{x, f(x) \in \mathcal{D}_{T+1}} f(x)$

Algorithm Expert Feedback Thread

Input: Active-learning style acqf. α_{feedback}, Bayesian Bradley-Terry model $r(x, f(x))$

for $t = 1, \ldots, T$ **do**

  Get $\mathcal{D}_t$ from the BO thread

  $\mathcal{D}_t^{\text{feedback}} = \{\}$

  for $k = 1, \ldots, K$ **do**

    Pick $(x_{k1}, f(x_{k1}))$ vs. $(x_{k2}, f(x_{k2}))$ via α_{feedback}

    $\mathcal{D}_t^{\text{feedback}} = \mathcal{D}_t^{\text{feedback}} \cup$ the above

  end for

  Present $\mathcal{D}_t^{\text{feedback}}$ on a web interface

  For some k, the expert pick $\ell_k \in \{1, 2\}$

  Add ℓ_k to $\mathcal{D}_t^{\text{feedback}}$

  Infer $p(r \mid \mathcal{D}_t^{\text{feedback}})$

end for

Current Idea

Open questions:

- What is α_{feedback}?
- What is $\alpha(p(g_t(x) \mid \mathcal{D}_t), p(r \mid \mathcal{D}_t^{\text{feedback}}))$?
- How to simulate expert feedback for benchmarking?
- What are good chemistry datasets to use?
- Know somebody who can help with the web interface?

Previous Work

BO:

- Huang et al., MFPL Workshop @ ICML 23
  - Train a reward model through expert preferences. Then use it as a feature extractor in the BO.
- Tiihonen et al., AI4Mat Workshop @ NeurIPS 23
  - Ask human feedback whether a novel x is good or bad (binary classification).

RL:

- Balsells et al., CoRL 23
  - Occassionally ask non-expert humans which trajectories are closer to the goal. This is asynchronous to the human-informed policy learning.
- Torne et al., 23
  - Similar as above.

Timeline

- Submit an extended abstract to AABI 24 (March 29)
- Submit the full paper at NeurIPS (late May)
- Biweekly meeting on Friday 10 AM EST?