Learning Step-Size Adaptation in CMA-ES

Gresa Shala, André Biedenkapp, Noor Awad, Steven Adriaensen, Marius Lindauer, and Frank Hutter

In a Nutshell

- Step-size in CMA-ES must be adapted dynamically
- Using Guided Policy Search (GPS) learn to control step-size offline in an automated, data-driven way
- Learned policies generalize beyond training setting
  o higher dimensions
  o longer runs
  o other function classes

Related Work

- Algorithm Configuration
  o Static [e.g. Ansótegui et al. 2009, Hutter et al. 2011, López-Ibáñez et al. 2011]
  o Dynamic [e.g. Adriaensen et al. 2016, Biedenkapp et al. 2020]
- Parameter Control Using Reinforcement Learning (RL)
  o Offline (Bartoli and Campigotto 2012, Sharma et al. 2019)
- Learning to Optimize [Li and Malik 2017]

GPS for DAC

Dynamic Algorithm Configuration (DAC)

- Configure per time-step & per-instance
- Learn a configuration policy
- Can be posed as RL problem
- Prior-art: Value-based RL (DQN)
  o Not sample-efficient
  o Focus on categorical parameters
  o Learning from scratch

Guided Policy Search (GPS)

- Sample-efficient RL method from robotics
- Learn arbitrary parameterized policies
- Represent policies as neural networks
- Learn policies offline
- Easily warm-start from demonstration
  o Imitation learning (supervised ML)
  o Learning from a reward signal (RL)

Learning Step-Size Adaptation

- Learn from Cumulative Step-size Adaptation (CSA)
- Vanilla GPS uses example trajectories only once, in the beginning, to warm-start the search
- We repeatedly query the hand-crafted baseline
  o Continuous use of expert knowledge
  o Learn from the teacher in many more situations
  o Sampling rate:
    0.0 → Vanilla GPS
    1.0 → Pure imitation learning
    0.3 → Good trade-off

Learning from a Hand-Crafted Heuristic

- Continuous use of expert knowledge
- Learn from the teacher in many more situations
- Sampling rate:
  0.0 → Vanilla GPS
  1.0 → Pure imitation learning
  0.3 → Good trade-off

Performance & Generalization

The learned policies ...
- are capable of producing well-performing step-sizes
- generalize to
  o longer trajectories
  o higher dimensions
  o other function classes