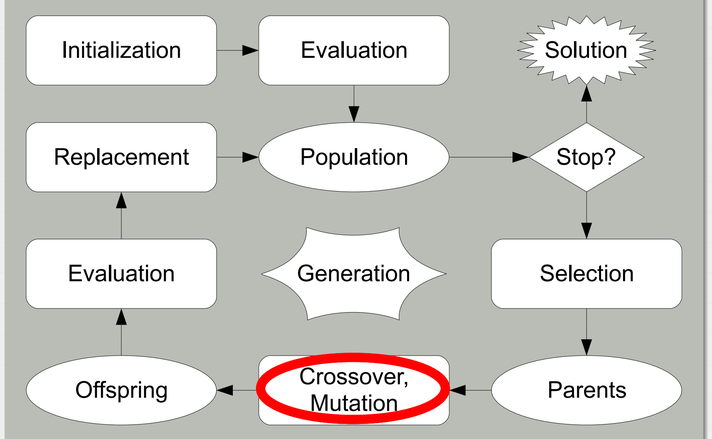
Adaptive Operator Selection in Evolutionary Algorithms with Extreme - Dynamic Multi-Armed Bandits

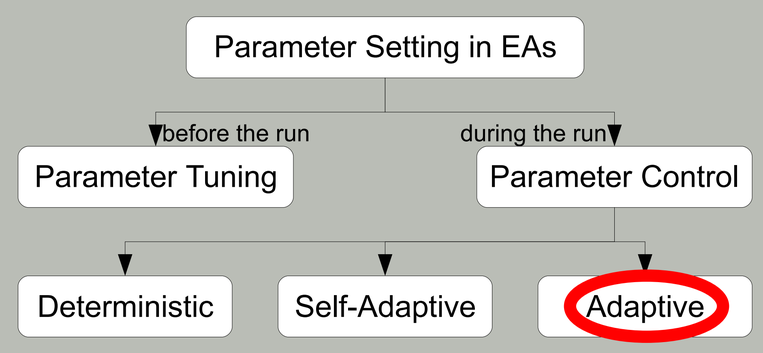
Álvaro Fialho and Marc Schoenauer

CENTRE DE RECHERCHE COMMUN
INRIA MICROSOFT RESEARCH

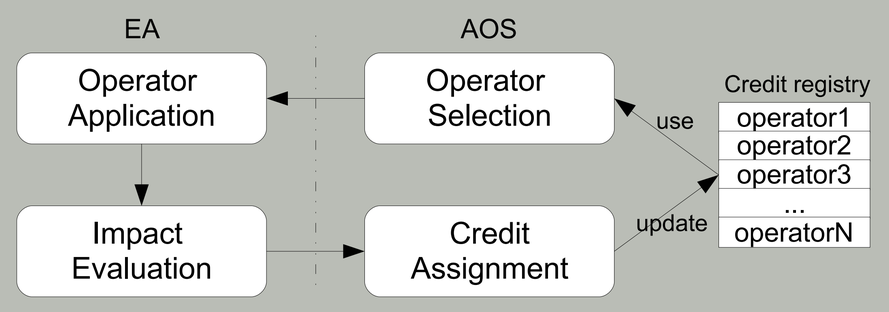
Evolutionary Algorithms



Parameter Setting in EAs



Adaptive Operator Selection



Autonomously select the operator to be applied amongst available ones, based on the past.

Operator Selection: Dynamic Multi-Armed Bandit *[GECCO'08]*

Yet another *Exploration vs. Exploitation* dilemma.

Multi-Armed Bandits

At time t, *gambler* plays arm j

$$r_{j,t} = \begin{cases} 1 \text{ with prob } = p_j \\ 0 \text{ with prob } = 1 - p_j \end{cases}$$

MAB UCB1 algorithm

- ***Be optimistic in face of the unknown.***
- At time t, choose arm j maximizing:

$$\hat{r}_{j,t} + \sqrt{\frac{2 \log \sum_k n_{k,t}}{n_{j,t}}}, \text{ where } \begin{cases} \hat{r}_{j,t} & \text{,estimated reward arm } j \\ n_{j,t} & \text{,chosen times for arm } j \end{cases}$$

EvE balance

- In MAB, $\hat{r}_{j,t} \in [0,1]$, not true in AOS, Scaling S needed.

Dynamic Context

- UCB1: too long to recover.
- Page-Hinkley statistical test
  - Detect changes in the reward distribution (threshold γ),
  - Restart the MAB from scratch.

DMAB: UCB1 + Scaling + Page-Hinkley

Credit Assignment: Extreme Value-Based *[PPSN'08]*

What is the impact measure?

- Unimodal: Fitness improvement
- Multimodal: Fitness and Diversity

How the impact history should be used?

- Usually, average performance is considered.
- Outlier operators are rarely used – smaller expectation.
- EC: focus on extreme, rather than average events.

$\mathcal{R}$ = Extreme value over a Window

On Unimodal Problems *[PPSN'08, LION'09, GECCO'09]*

Impact = ΔFitness

- (1+50)-EA with 4 mutation operators on the OneMax.
- (1+50)-EA with 5 mutation operators on the Long K-Path.
- (100, 100)-EA with 4 Xover and 1 mut. on the Royal Road.
- Equivalent to Optimal strategy (not available for RR).
- Significantly better than the Naive (uniform) strategy.
- Better than other AOS techniques (AP, PM, MAB combined with either Extreme or Average rewards).

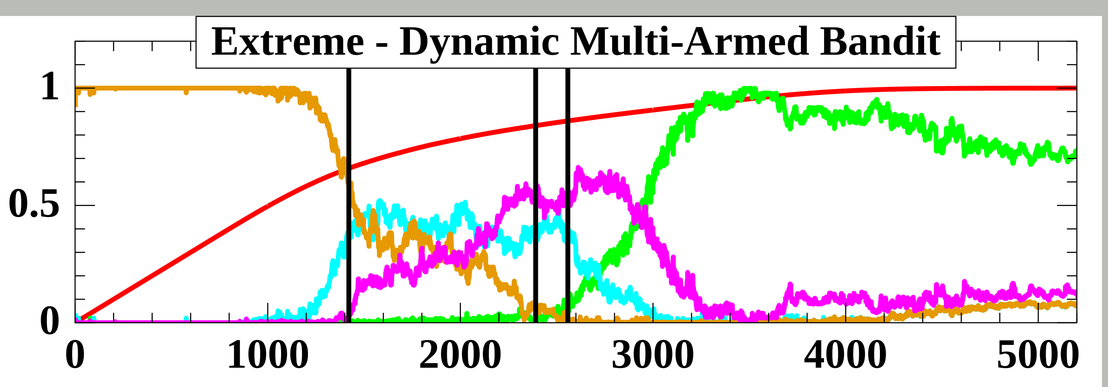


Figure: Performance on the 10k bits OneMax, averaged over 50 runs.

On (multimodal) SAT problems *[CEC'09]*

Impact = F(ΔFitness, ΔDiversity)

- How to aggregate them? Compass!
- SSGA with 6 local search operators.
- 22 instances: SATlib & SAT Race'06.
- Significantly better than Naive and both original DMAB and Compass.

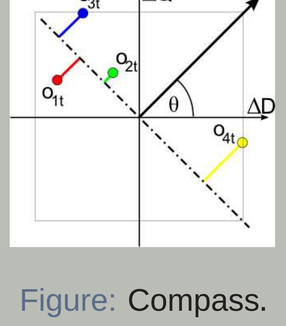


Figure: Compass.

Final Considerations

- Off-line tuning of meta-parameters is still expensive.
- In real problems, optimal behavior is not known.
- Better than fixed, naive and known adaptive approaches.
- Next: Generalization (meta-parameters and rewarding).

References

- GECCO'08: L. Da Costa, A. Fialho, M. Schoenauer, and M. Sebag. *Adaptive Operator Selection with Dynamic Multi-Armed Bandis.*
- PPSN'08: A. Fialho, L. Da Costa, M. Schoenauer, and M. Sebag. *Extreme value based Adaptive Operator Selection.*
- LION'09: A. Fialho, L. Da Costa, M. Schoenauer, and M. Sebag. *DMABs and Extreme value-based rewards for AOS in EAs.*
- CEC'09: J. Maturana, A. Fialho, F. Saubion, M. Schoenauer, and M. Sebag. *Extreme Compass and DMABs for Adaptive Operator Selection.*
- GECCO'09: A. Fialho, M. Schoenauer, and M. Sebag. *Analysis of AOS techniques on the Royal Road and Long k-Path problems.*

{Alvaro.Fialho,Marc.Schoenauer}@inria.fr Microsoft Research - INRIA Joint Centre, Orsay, France