Dynamic Multi-Armed Bandits and Extreme Value Rewards for Adaptive Operator Selection in Evolutionary Algorithms

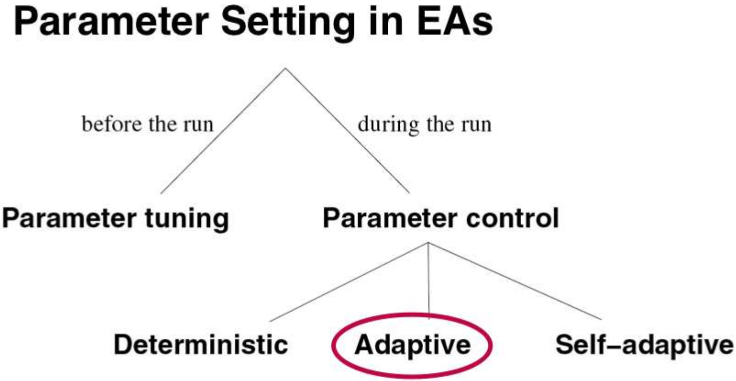
Álvaro Fialho[1], Luís Da Costa[2], Marc Schoenauer[1,2], Michèle Sebag[1,2]

[1]Microsoft Research – INRIA Joint Centre
Orsay, France

[2]Project-Team TAO, LRI / INRIA Saclay - Île-de-France
Orsay, France

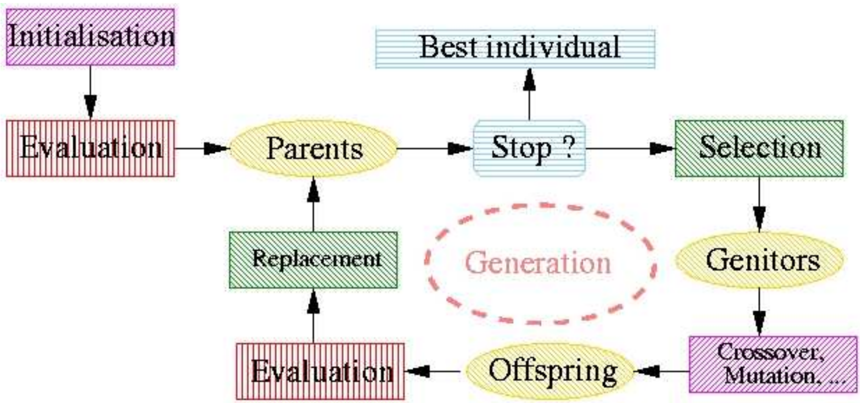
Learning and Intelligent OptimizatioN Conference - LION 3
January 15th, 2009

Parameter Setting in EAs

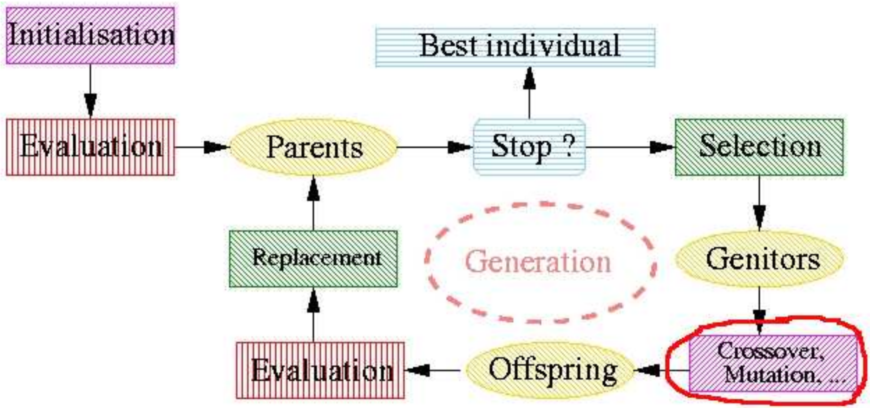


(from [Eiben et al., 2007])

Evolutionary Algorithms



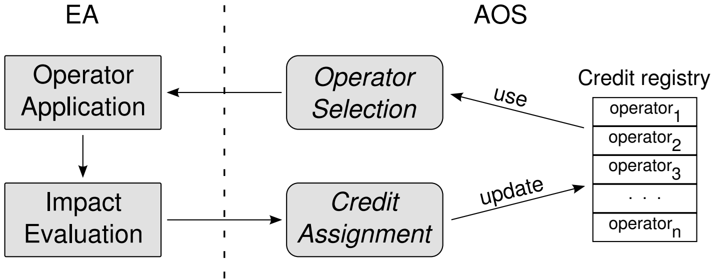
Evolutionary Algorithms



Adaptive Operator Selection

Objective

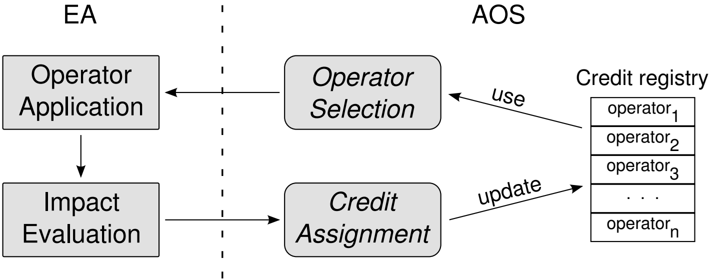
Autonomously select the operator to be applied amongst available ones, based on its impact in the past.



Adaptive Operator Selection

Objective

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



This work:

- Operator selection: Dynamic Multi-Armed Bandits
- Credit Assignment: Extreme Value Based

AOS: A (kind of) Multi-Armed Bandit problem

Multi-Armed Bandits

At time t, *gambler* plays arm j

$$\textit{reward at } t \; : r_t = \begin{cases} 1 & \text{with prob } = p_j \\ 0 & \text{with prob } = 1 - p_j \end{cases}$$

Goal: maximize cumulated reward

AOS: A (kind of) Multi-Armed Bandit problem

Multi-Armed Bandits

At time t, *gambler* plays arm j

$$\textit{reward at } t \ : r_t = \begin{cases} 1 & \text{with prob } = p_j \\ 0 & \text{with prob } = 1 - p_j \end{cases}$$

Goal: maximize cumulated reward

State-of-the-art: UCB1 [Auer et al., 2002]

- At time t, choose arm j maximizing:

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

AOS with Multi-Armed Bandits: the true story I

Scaling

MAB framework:

- $\hat{r}_{j,t} \in [0, 1]$;

AOS framework:

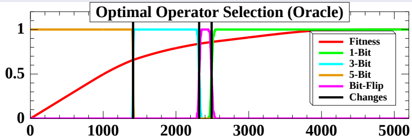
- $\hat{r}_{j,t} \in [a, b]$ (e.g. fitness improvement)

UCB1's EvE balance is broken, **Scaling** is needed:

$$\hat{q}_{i,t} = \mathcal{C} * \hat{r}_{j,t} + \sqrt{\frac{2 \log \sum_k n_{k,t}}{n_{j,t}}}$$

AOS with Multi-Armed Bandits: the true story II

AOS is a dynamic context

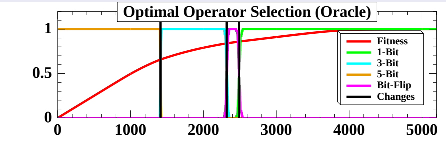


UCB1: too long to recover.

- Detect change;
- Restart the MAB.

AOS with Multi-Armed Bandits: the true story II

AOS is a dynamic context



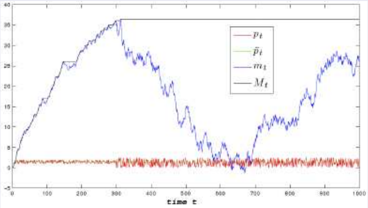
UCB1: too long to recover.

- Detect change;
- Restart the MAB.

How to detect a change in a distribution?

[Page-Hinkley test, 1954]

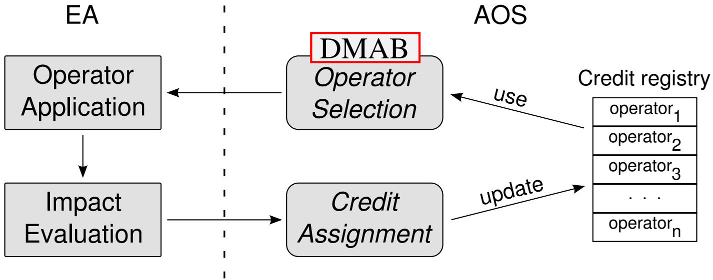
1. $\bar{r}_t = \frac{1}{t}\sum_{i=1}^{t} r_i$
2. $m_t = \sum_{i=1}^{t}(r_i - \bar{r}_i + \delta)$,
3. $M_t = \max\{|m_i|,\ i = 1 \ldots t\}$
4. Return $(M_t - |m_t| > \lambda)$



Operator Selection: Dynamic Multi-Armed Bandits

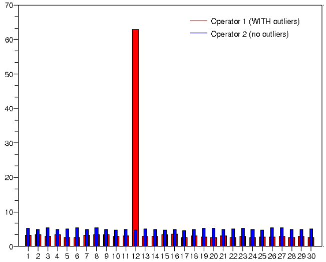
DMAB: UCB1 + Scaling + Page-Hinkley

- Proposed by other members of our group [Hartland, 2007]
- Won the Pascal Network challenge on "On-line Trading of Exploration and Exploitation" (OTEE)



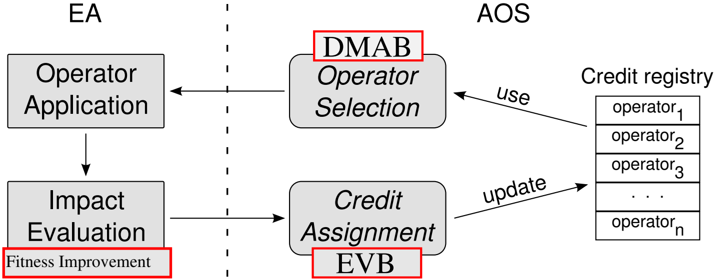
Credit Assignment: Extreme Value-Based (EVB)

- Fitness improvement: $(\mathcal{F}(o(x)) - \mathcal{F}(x))_+$

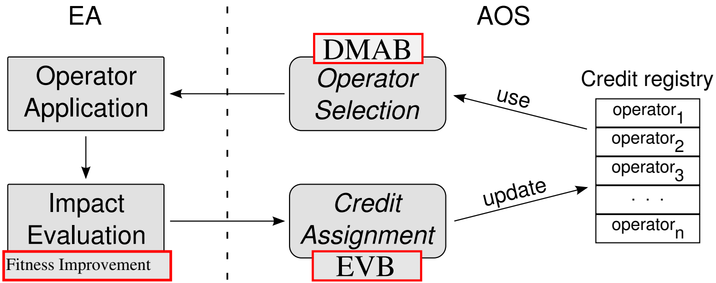


- Outlier operators are rarely considered - smaller expectation.
- EC: Focus on extreme, rather than average events
  - Complex systems, e.g. epidemic propagation, financial markets
- Extreme Value-Based (EVB) Credit Assignment.
  - $\mathcal{R}$ = Extreme value over a Window

Ex-DMAB Adaptive Operator Selection



Ex-DMAB Adaptive Operator Selection



Meta-Parameters

- Operator selection (DMAB):
  - $\mathcal{C}$, for scaling ;
  - λ and δ, for the PH test ($\delta \equiv 0.15$)
- Credit assignment (EVB):
  - $\mathcal{W}$, the size of the sliding window

Experimental Conditions

- (1+50)-EA applied to: One-Max; Long k-Path
- Mutation Operators:
  - 1-bit, 3-bit, 5-bit
  - $1/n$ bit-flip (and also k/n bit-flip for the 2nd problem)
- Initial individual is set to $(0, \ldots, 0)$
- Meta-parameters tuned off-line
  - One-Max: complete DOE campaign
  - Long k-Path: F-RACE [Birattari et al., 2002], a Racing technique using the Friedman's 2-way ANOVA by ranks

Extreme-DMAB *versus*

- Probability Matching and Adaptive Pursuit [Thierens, 2007]
- Average-DMAB *
- Naive (uniform) and Optimal operator selection *

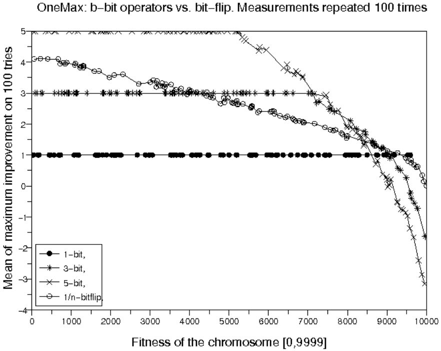
The One-Max Problem

- 10^4 bits
- Fitness is the number of “1”s in the bitstring
- Very simple unimodal problem, the “drosophila of EC”

The One-Max Problem

- 10^4 bits
- Fitness is the number of "1"s in the bitstring
- Very simple unimodal problem, the "drosophila of EC"

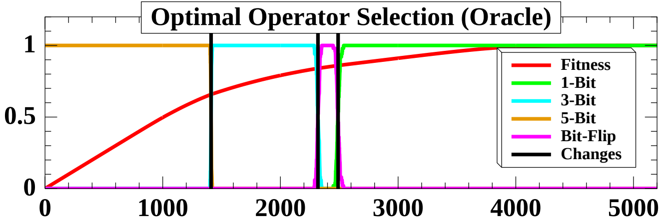
The One-Max "Oracle"



Best operator:

1. 5-Bit for
   $\mathcal{F} \in [0 : 6579]$
2. 3-Bit for
   $\mathcal{F} \in [6580 : 8400]$
3. $\frac{1}{n}$Bit-Flip for
   $\mathcal{F} \in [8401 : 8600]$
4. 1-Bit for
   $\mathcal{F} \in [8601 : 10000]$

Results Extreme - DMAB on the One-Max



Results Extreme - DMAB on the One-Max

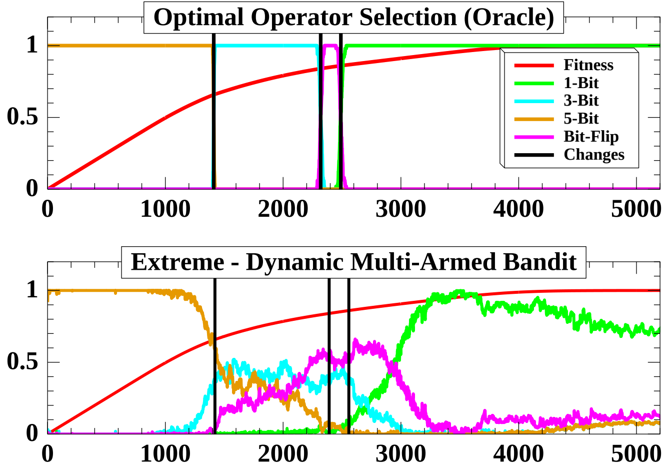


Figure: Extreme - DMAB behavior averaged over 50 runs.

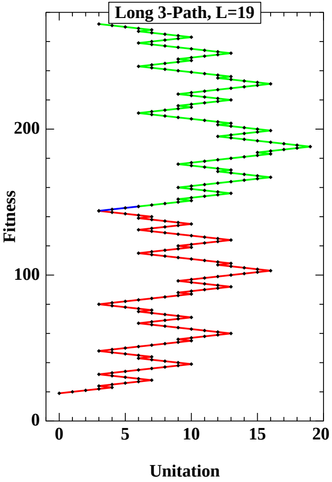
Comparative Results on the One-Max

AOS	Conf.	Gens. to Optimum
Extreme - DMAB	$C = 1, \gamma = 250$	**5467** $\pm$ 513
Average - DMAB	$C = 10, \gamma = 25$	7727 $\pm$ 642
Optimal Strategy	Given by "Oracle"	**5069** $\pm$ 292
Best Naive	$\mathcal{U}$(1-Bit+5-Bit)	6793 $\pm$ 625
Complete Naive	$\mathcal{U}$(4 ops.)	7813 $\pm$ 708

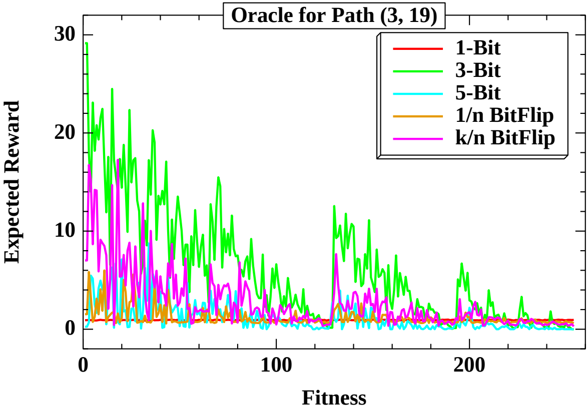
Table: Results on the 10k bits One-Max problem (over 50 runs).

Long k-Path Problems

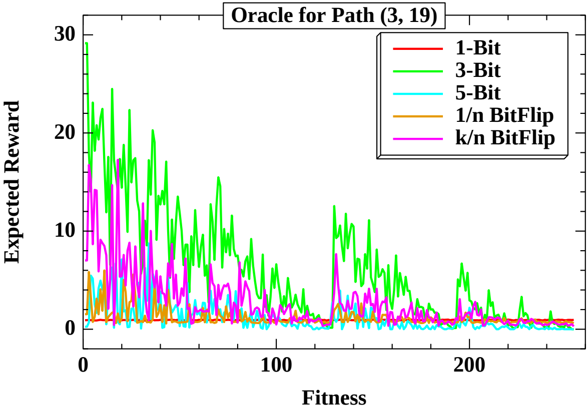
- Unimodal, very long path to the optimum
  - Path length grows exponentially with size of the bitstring (n)
  - Hamming distance between two consecutive points is 1
  - Any other point at distance 1 is off the path
- "Shortcuts" require at least k bit-flips



The Long K-Path Oracle



The Long K-Path Oracle



N	Length
43	$\sim$ 65 K
49	$\sim$ 262 K
55	$\sim$ 1 M
61	$\sim$ 4 M

Comparative Results on the Long K-Path Problem I

N	DMAB - $\mathcal{W}(\mathcal{C},\lambda)$		Optimal	Uniform
	Extreme 500 (100; 100)	Average 50 (50; .5)		
43	2910/1771	$=$ 3039/2234	$>$ 2134/2414	$=$ 3462/2174
49	4407/2698	$<$ 5950/3694	$=$ 3590/3327	$=$ 5201/3461
55	6257/4535	$<$ 8366/5991	$=$ 4858/5669	$<$ 9778/5245
61	14586/9345	$=$ 16222/8608	$>$ 8608/9907	$=$ 12243/1004

Table: Average/Std. Deviation number of generations to optimum out of 50 runs, using the optimal AOS parameters found by F-Race over ALL instances. Comparison validated by applying both unsigned Wilcoxon rank sum and Kolmogorov-Smirnov non-parametric tests.

Comparative Results on the Long K-Path Problem II

N	DMAB - $\mathcal{W}(\mathcal{C}, \lambda)$		Optimal	Uniform
	Extreme	Average		
43	2457/1945 500(50; 50)	$=$ 2815/1908 50(.5; 100)	$=$ 2134/2414	$<$ 3462/2174
49	3670/2485 500(100; 500)	$<$ 5759/3696 50(.1; 1000)	$=$ 3590/3327	$<$ 5201/3461
55	6257/4535 500(100; 100)	$<$ 8303/5945 50(50; .1)	$=$ 4858/5669	$<$ 9778/5245
61	12380/9733 500(50; 25)	$=$ 14521/8945 50(.5; 50)	$=$ 8608/9907	$=$ 12243/10041

Table: Average/Std. Deviation number of generations to optimum out of 50 runs, using the optimal AOS parameters found by F-Race over EACH instance. Comparison validated by applying both unsigned Wilcoxon rank sum and Kolmogorov-Smirnov non-parametric tests.

Conclusions and Perspectives

Efficiency

- Demonstrated on 2 benchmark problems.
- Still expensive. Real problems – no optimal behavior.
- Better than fixed, naive and known adaptive approaches.

Conclusions and Perspectives

Efficiency

- Demonstrated on 2 benchmark problems.
- Still expensive. Real problems – no optimal behavior.
- Better than fixed, naive and known adaptive approaches.

Deepen our understanding on meta-parameters

- High sensitivity.
  - DMAB's $\mathcal{C}$ and λ affect the EvE balance.
- Self-Adaptation? Off-line? Learning?

Conclusions and Perspectives

Efficiency

- Demonstrated on 2 benchmark problems.
- Still expensive. Real problems – no optimal behavior.
- Better than fixed, naive and known adaptive approaches.

Deepen our understanding on meta-parameters

- High sensitivity.
  - DMAB's $\mathcal{C}$ and λ affect the EvE balance.
- Self-Adaptation? Off-line? Learning?

Generalization

- Rank-based rewarding (aware of fitness variance)
- Consider different measures (e.g. diversity)
  - Current work on SAT problems (Université d'Angers)

References I

- Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2/3):235–256.
- Birattari, M., Stutzle, T., Paquete, L., Varrentrapp, K. (2002). A racing algorithm for configuring metaheuristics. In *Proc. GECCO'02*.
- Da Costa, L., Fialho, A., Schoenauer, M., and Sebag, M. (2008). Adaptive operator selection with dynamic multi-armed bandits. In *Proc. GECCO'08*.
- Eiben, A. E., Michalewicz, Z., Schoenauer, M., and Smith, J. E. (2007). Parameter Control in Evolutionary Algorithms. In *Parameter Setting in Evolutionary Algorithms*, chapter 2, pages 19–46. Springer Verlag.
- Fialho, A., Da Costa, L., Schoenauer, M., and Sebag, M. (2008). Extreme value based adaptive operator selection. In *Proc. PPSN'08*.
- Hartland, C., Baskiotis, N., Gelly, S., Teytaud, O., and Sebag, M. (2007). Change point detection and meta-bandits for online learning in dynamic environments. In *Proc. CAp'07*.

References II

- Thierens, D. (2007). Adaptive Strategies for Operator Allocation. In *Parameter Setting in Evolutionary Algorithms*, pages 77–90. Springer Verlag.
- Whitacre, J. M., Pham, T. Q., and Sarker, R. A. (2006). Credit assignment in adaptive evolutionary algorithms. In *Proc. GECCO'06*.
- Page, E. (1954). Continuous inspection schemes. *Biometrika*, 41:100–115

Dynamic Multi-Armed Bandits and Extreme Value Rewards for Adaptive Operator Selection in Evolutionary Algorithms

Álvaro Fialho[1], Luís Da Costa[2], Marc Schoenauer[1,2], Michèle Sebag[1,2]

[1]Microsoft Research – INRIA Joint Centre
Orsay, France

[2]Project-Team TAO, LRI / INRIA Saclay - Île-de-France
Orsay, France

Learning and Intelligent OptimizatioN Conference - LION 3
January 15th, 2009