Adaptive Operator Selection in EAs with Extreme - Dynamic Multi-Armed Bandits

Álvaro Fialho and Marc Schoenauer
Microsoft Research – INRIA Joint Centre, Orsay, France
FirstName.LastName@inria.fr

August 17, 2009

Abstract

The performance of evolutionary algorithms is highly affected by the selection of the variation operators to solve the problem at hand. This paper presents a brief review of the results that have been recently obtained using the "Extreme - Dynamic Multi-Armed Bandit" (Ex-DMAB), a technique used to automatically select the operator to be applied between the available ones, while searching for the solution. Experiments on three well-known unimodal artificial problems of the EC community, namely the OneMax, the Long k-Path and the Royal Road, and on a set of a SAT instances, are briefly presented, demonstrating some improvements over both any choice of a single-operator alone, and the naive uniform choice of one operator at each application.

1 Adaptive Operator Selection

Adaptive methods use information from the history of evolution to modify parameters while solving the problem. This paper focuses on the *Adaptive Operator Selection* (AOS), i.e., the definition of an on-line strategy able to autonomously select between different variation operators each time one needs to be applied. Fig. 1 illustrates the general scheme for achieving this goal, from which we can derive the need of defining two main components: the *Credit Assignment* - how to assess the performance of each operator based on the impact of its application on the progress of the search; and the *Operator Selection* rule - how to select between the different operators based on the rewards that they have received so far.

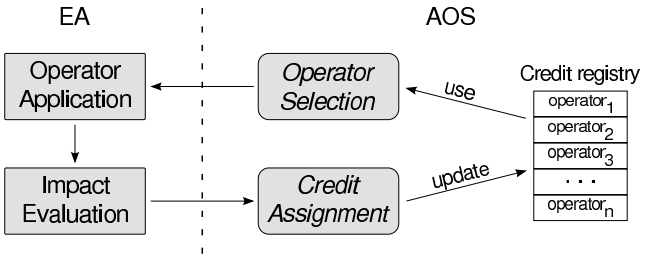

Figure 1: General *Adaptive Operator Selection* scheme.

2 Extreme - Dynamic Multi-Armed Bandit

The two ingredients of the *Adaptive Operator Selection* method proposed by us are: an *Operator Selection* rule based on the Multi-Armed Bandit paradigm, and a *Credit Assignment* mechanism based on extreme values.

2.1 Operator Selection: Dynamic Multi-Armed Bandits

The explored idea, firstly proposed in [3], is that the selection of an operator can be seen as yet another Exploration vs. Exploitation dilemma, but this time at operator-selection level: there is the need of applying as much as possible the operator known to have brought the best results so far, while nevertheless exploring the other possibilities, in case one of the other operators becomes the best option at some point. Such dilemma has been intensively studied in the context of *Game Theory*, in the so-called Multi-Armed Bandit (MAB) framework. Among the existent MAB variants, the *Upper Confidence Bound* (UCB) [1] was chosen to be used, for being proved optimal w.r.t. maximization of the cumulative reward.

More formally, the UCB algorithm works as follows. Each variation operator is viewed as an *arm* of a MAB problem. Let $n_{i,t}$ denote the number of times the i^{th} arm has been played up to time t, and let $\hat{p}_{i,t}$ denote the average empirical reward received until time t by arm i. At each time step t, the algorithm selects the arm maximizing the following quantity:

$$\hat{p}_{j,t} + C\sqrt{\frac{2\log\sum_k n_{k,t}}{n_{j,t}}} \tag{1}$$

The first term of this equation favors the best empirical arm (exploitation) while the second term ensures that each arm is selected infinitely often (exploration); this algorithm has also been described briefly as "be opti-

mistic when facing the unknown", as the second term of Equation 1 can also be viewed as some kind of variance, and the user should choose the arm that might lead to the highest value.

In the original setting [1], all rewards, and hence also their empirical means $\hat{p}_{j,t}$ are in $[0, 1]$. However, since this is not the case in the AOS context, a *Scaling factor* C is needed, in order to properly balance the trade-off between both terms.

Another important issue is that the original MAB setting is static, while the AOS scenario is dynamic, i.e., the quality of the operators is likely to change along the different stages of the search. Even though the exploration term in the UCB algorithm ensures that all operators will be tried infinitely many times, after a change in their ordering, it might take a long time before the new best operator catches up. To cope with dynamic environments, it was proposed [7] to use a statistical test that efficiently detects changes in time series, the *Page-Hinkley* (PH) test [10], coupled with the UCB algorithm. Basically, as soon as this test detects a change in the rewards distribution, the MAB algorithm is restarted from scratch, allowing it to quickly re-discover the new best operator.

2.2 Credit Assignment: Extreme Value Based

The idea of using extreme values was proposed as the *Credit Assignment* mechanism, based on the assumption that attention should be paid to extreme, rather than average events, in agreement with [11]. The credit assigned to the operator is the maximum of the impacts caused by the operator application over a sliding window of the last $\mathcal{W}$ applications.

The measurement of such impact depends on the nature of the problem at hand. In unimodal problems, we have been using the fitness improvement; while in the multimodal SAT problems, an engineered aggregation of fitness improvement and diversity, called Compass [9], was applied.

3 Summary of Results

Experiments with *Ex-DMAB* in unimodal benchmark problems have been presented in [4, 5, 6], in which the fitness improvement was used to measure the impact of the operator application. The *Ex-DMAB* has been used to adapt a (1+λ)-EA, by efficiently choosing on-line between 4 mutation operators for solving the OneMax problem [3]; and has been tried on yet another unimodal benchmark problem, the Long k-Path [5], this time efficiently selecting between 5 mutation operators. In both cases, the optimal operator

selection strategy was extracted by means of Monte-Carlo simulations, and the Ex-DMAB showed to perform statistically equivalent to it; while significantly improving over the naive (uniform selection) approach. It has also been used to adapt a (100,100)-EA with 4 crossover and 1 mutation operators on the Royal Road problem [6], also performing significantly better than the naive approach. For the three problems, we have also used other AOS combinations as baseline for comparison, namely Adaptive Pursuit, Probability Matching and the static MAB (without restarts), coupled with Extreme or Average rewards. Ex-DMAB was shown to be the best option.

A different analysis was also done in the light of SAT problems, in [8]. Since these problems are mostly multimodal, the reward used was the Compass [9], which aggregates both the fitness improvement of the offspring, and the diversity that this offspring brought by being inserted in the population. Significantly better results were achieved w.r.t. the naive approach, and also to the original Compass and Ex-DMAB combinations. The application of such approach in multimodal functions should be more deeply investigated in the future.

4 Discussion and Perspectives

A current drawback concerns the tuning of the *Ex-DMAB* hyper-parameters, the window size $\mathcal{W}$, the scaling factor C and the change detection test threshold γ – actually, they are being off-line tuned by means of F-Race [2]. Although its good performances rely on such expensive procedure, *Ex-DMAB* was found to outperform the main options opened to the naive EA user, namely (i) using a fixed or deterministic strategy (including the naive, uniform selection, strategy; or (ii) using a different AOS strategy. Furthermore, *Ex-DMAB* involves a fixed and limited number of parameters, whereas the number of operator rates increases with the number of operators.

Further research will aim at addressing the above weaknesses. Firstly, we shall investigate better how the threshold γ and the scaling factor C relate, as both cooperate to control the exploration vs exploitation trade-off. Another possible direction is the analysis of a rank-based reward, instead of the absolute value that is currently being used; in this way, it will not be problem-dependent anymore, and a robust setting for these parameters might be found. A different direction that might also be analyzed is the use of this selection technique in another context, selecting between different algorithms instead of operators.

References

[1] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2/3):235–256, 2002.

[2] M. Birattari, T. Stützle, L. Paquete, and K. Varrentrapp. A racing algorithm for configuring metaheuristics. In *Proc. GECCO'02*, pages 11–18. Morgan Kaufmann, 2002.

[3] L. Da Costa, A. Fialho, M. Schoenauer, and M. Sebag. Adaptive operator selection with dynamic multi-armed bandits. In *Proc. GECCO'08*, pages 913–920. ACM, 2008.

[4] A. Fialho, L. Da Costa, M. Schoenauer, and M. Sebag. Extreme value based adaptive operator selection. In *Proc. PPSN'08*, pages 175–184. Springer, 2008.

[5] A. Fialho, L. Da Costa, M. Schoenauer, and M. Sebag. Dynamic multi-armed bandits and extreme value-based rewards for adaptive operator selection in evolutionary algorithms. In *Proc. LION'09*. Springer, 2009 (to appear).

[6] A. Fialho, M. Schoenauer, and M. Sebag. Analysis of adaptive operator selection techniques on the royal road and long k-path problems. In *Proc. GECCO'09*, pages 779–786. ACM, 2009.

[7] C. Hartland, N. Baskiotis, S. Gelly, O. Teytaud, and M. Sebag. Change point detection and meta-bandits for online learning in dynamic environments. In *Proc. CAp'07*, pages 237–250. Cepadues, 2007.

[8] J. Maturana, A. Fialho, F. Saubion, M. Schoenauer, and M. Sebag. Extreme compass and dynamic multi-armed bandits for adaptive operator selection. In *Proc. CEC'09*, pages 365–372. IEEE, 2009.

[9] J. Maturana and F. Saubion. A compass to guide genetic algorithms. In *Proc. PPSN'08*, pages 256–265. Springer, 2008.

[10] E. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.

[11] J. M. Whitacre, T. Q. Pham, and R. A. Sarker. Use of statistical outlier detection method in adaptive evolutionary algorithms. In *Proc. GECCO'06*, pages 1345–1352. ACM, 2006.