An Adaptive Pursuit Strategy for Allocating Operator Probabilities

Dirk Thierens

Department of Computer Science
Universiteit Utrecht, The Netherlands

Outline

Adaptive Operator Allocation: What ?

- **Given:**
  1. Set of K operators $\mathcal{A} = \{a_1, \ldots, a_K\}$
  2. Probability vector $\mathcal{P}(t) = \{\mathcal{P}_1(t), \ldots, \mathcal{P}_K(t)\}$:
     operator a_i applied at time t in proportion to probability $\mathcal{P}_i(t)$
  3. Environment returns rewards $\mathcal{R}_i(t) \geq 0$
- **Goal:** Adapt $\mathcal{P}(t)$ such that the expected value of the cumulative reward $\mathcal{E}[\mathcal{R}] = \sum_{t=1}^{T} \mathcal{R}_i(t)$ is maximized

Adaptive Operator Allocation: Why ?

Probability of applying an operator

1. is difficult to determine a priori
2. depends on current state of the search process

$\rightarrow$ Adaptive allocation rule specifies how probabilities are adapted according to the performance of the operators

Adaptive Operator Allocation: Requirements

1. Non-stationary environment $\Rightarrow$ operator probabilities need to be adapted continuously
2. Stationary environment $\Rightarrow$ operator probabilities should converge to best performing operator

$\rightarrow$ conflicting goals !

Probability Matching: Main Idea

- Adaptive allocation rule often applied in GA literature: probability matching strategy
- Main idea: update $\mathcal{P}(t)$ such that the probability of applying operator a_i matches the proportion of the estimated reward $Q_i(t)$ to the sum of all reward estimates $\sum_{a=1}^{K} \mathcal{Q}_a(t)$

Probability Matching: Reward Estimate

- The adaptive allocation rule computes an estimate of the rewards received when applying an operator
- In non-stationary environments older rewards should get less influence
- Exponential, recency-weighted average ($0 < \alpha < 1$):

$$\mathcal{Q}_a(t+1) = \mathcal{Q}_a(t) + \alpha[\mathcal{R}_a(t) - \mathcal{Q}_a(t)]$$

Probability Matching: Probability Adaptation

- In non-stationary environments the probability of applying any operator should never be less than some minimal threshold $P_{min} > 0$
- For K operators maximal probability $P_{max} = 1 - (K - 1)P_{min}$
- Updating rule for $\mathcal{P}(t)$:

$$\mathcal{P}_a(t) = P_{min} + (1 - K \cdot P_{min}) \frac{\mathcal{Q}_a(t)}{\sum_{i=1}^{K} \mathcal{Q}_i(t)}$$

Probability Matching: Algorithm

ProbabilityMatching$(\mathcal{P}, \mathcal{Q}, K, P_{min}, \alpha)$

for $i \leftarrow 1$ **to** K
do $\mathcal{P}_i(0) \leftarrow \frac{1}{K}$; $\mathcal{Q}_i(0) \leftarrow 1.0$
while NotTerminated?()
do $a^s \leftarrow$ ProportionalSelectOperator$(\mathcal{P})$
    $R_{a^s}(t) \leftarrow$ GetReward(a^s)
    $\mathcal{Q}_{a^s}(t+1) = \mathcal{Q}_{a^s}(t) + \alpha[R_{a^s}(t) - \mathcal{Q}_{a^s}(t)]$
    for $a \leftarrow 1$ **to** K
    do $\mathcal{P}_a(t+1) = P_{min} + (1 - K \cdot P_{min}) \frac{\mathcal{Q}_a(t+1)}{\sum_{i=1}^{K} \mathcal{Q}_i(t+1)}$

Probability Matching: Problem

- Assume one operator is consistently better
- For instance, 2 operators a_1 and a_2 with constant rewards $\mathcal{R}_1 = 10$ and $\mathcal{R}_2 = 9$
- If $P_{min} = 0.1$ we would like to apply operator a_1 with probability $\mathcal{P}_1 = 0.9$ and operator a_2 with $\mathcal{P}_2 = 0.1$.
- Yet, the probability matching allocation rule will converge to $\mathcal{P}_1 = 0.52$ and $\mathcal{P}_2 = 0.48$!

Adaptive Pursuit Strategy: Pursuit Method

- The pursuit algorithm is a rapidly converging algorithm applied in learning automata
- Main idea: update $\mathcal{P}(t)$ such that the operator a^* that currently has the maximal estimated reward $\mathcal{Q}_{a^*}(t)$ is pursued
- To achieve this, the pursuit method increases the selection probability $\mathcal{P}_{a^*}(t)$ and decreases all other probabilities $\mathcal{P}_a(t), \forall a \neq a^*$
- Adaptive pursuit algorithm is extension of the pursuit algorithm to make it applicable in non-stationary environments

Adaptive Pursuit Strategy: Adaptive Pursuit Method

- Similar to probability matching:
  1. The adaptive pursuit algorithm proportionally selects an operator to execute according to the probability vector $\mathcal{P}(t)$
  2. The estimated reward of the selected operator is updated with:

$$\mathcal{Q}_a(t+1) = \mathcal{Q}_a(t) + \alpha[\mathcal{R}_a(t) - \mathcal{Q}_a(t)]$$

- Different from probability matching:
  1. Selection probability vector $\mathcal{P}(t)$ is adapted in a greedy way

Adaptive Pursuit Strategy: Probability Adaptation

- The selection probability of the current best operator $a^* = \text{argmax}_a[\mathcal{Q}_a(t+1)]$ is increased ($0 < \beta < 1$):

$$\mathcal{P}_{a^*}(t+1) = \mathcal{P}_{a^*}(t) + \beta[P_{max} - \mathcal{P}_{a^*}(t)]$$

- The selection probability of the other operators is decreased:

$$\forall a \neq a^* : \mathcal{P}_a(t+1) = \mathcal{P}_a(t) + \beta[P_{min} - \mathcal{P}_a(t)]$$

Note

$$\begin{aligned}
&\sum_{a=1}^{K} \mathcal{P}_a(t+1) \\
&= \mathcal{P}_{a^*}(t) + \beta[P_{max} - \mathcal{P}_{a^*}(t)] + \sum_{a=1, a \neq a^*}^{K} (\mathcal{P}_a(t) + \beta[P_{min} - \mathcal{P}_a(t)]) \\
&= (1-\beta) \sum_{a=1}^{K} \mathcal{P}_a(t) + \beta[P_{max} + (K-1)P_{min}] \\
&= (1-\beta) \sum_{a=1}^{K} \mathcal{P}_a(t) + \beta \\
&= 1
\end{aligned}$$

Adaptive Pursuit Strategy: Algorithm

$\textsc{AdaptivePursuit}(\mathcal{P}, \mathcal{Q}, K, P_{min}, \alpha, \beta)$

$P_{max} \leftarrow 1 - (K-1)P_{min}$
for $i \leftarrow 1$ **to** K
do $\mathcal{P}_i(0) \leftarrow \frac{1}{K}$; $\mathcal{Q}_i(0) \leftarrow 1.0$
while $\textsc{NotTerminated?}()$
do $a^s \leftarrow \textsc{ProportionalSelectOperator}(\mathcal{P})$
    $R_{a^s}(t) \leftarrow \textsc{GetReward}(a^s)$
    $\mathcal{Q}_{a^s}(t+1) = \mathcal{Q}_{a^s}(t) + \alpha[R_{a^s}(t) - \mathcal{Q}_{a^s}(t)]$
    $a^* \leftarrow \textsc{Argmax}_a(\mathcal{Q}_a(t+1))$
    $\mathcal{P}_{a^*}(t+1) = \mathcal{P}_{a^*}(t) + \beta[P_{max} - \mathcal{P}_{a^*}(t)]$
    for $a \leftarrow 1$ **to** K
    do if $a \neq a^*$
        then $\mathcal{P}_a(t+1) = \mathcal{P}_a(t) + \beta[P_{min} - \mathcal{P}_a(t)]$

Adaptive Pursuit Strategy: Example

- Consider again the 2-operator stationary environment with $\mathcal{R}_1 = 10$, and $\mathcal{R}_2 = 9$ ($P_{min} = 0.1$)
- As opposed to the probability matching rule, the adaptive pursuit method will play the better operator a_1 with maximum probability $P_{max} = 0.9$
- It also keeps playing the poorer operator a_2 with minimal probability $P_{min} = 0.1$ in order to maintain its ability to adapt to any change in the reward distribution

Experiments: Environment

- We consider an environment with 5 operators $a_i : i = 1 \ldots 5$
- Each operator a_i receives a uniformly distributed reward $\mathcal{R}_i$ between the boundaries $\mathcal{R}_i = \mathcal{U}[i-1 \ldots i+1]$:

Operator reward	[0..1]	[1..2]	[2..3]	[3..4]	[4..5]	[5..6]
$\mathcal{R}_1$						
$\mathcal{R}_2$						
$\mathcal{R}_3$						
$\mathcal{R}_4$						
$\mathcal{R}_5$						

- After a fixed time interval ΔT the reward distributions are randomly reassigned to the operators

Upper bounds to performance

- If we had full knowledge of the reward distributions and their switching pattern we could always pick the optimal operator a^* and achieve an expected reward $\mathcal{E}[\mathcal{R}^{Opt}] = 5$.
- The performance in the stationary (non-switching) environment of a correctly converged operator allocation scheme represents an upper bound to the optimal performance in the switching environment.
- 3 allocation strategies:
  1. Non-adaptive, equal-probability allocation rule
  2. Probability matching allocation rule ($P_{min} = 0.1$)
  3. Adaptive pursuit allocation rule ($P_{min} = 0.1$)

Non-adaptive, equal-probability allocation rule

The probability of choosing the optimal operator a^*_{Fixed}:

$$\text{Prob}[a^s = a^*_{Fixed}] = \frac{1}{K} = 0.2$$

The expected reward:

$$\begin{aligned}
\mathcal{E}[\mathcal{R}^{Fixed}] &= \sum_{a=1}^{K} \mathcal{E}[\mathcal{R}_a]\text{Prob}[a^s = a] \\
&= \frac{\sum_{a=1}^{K} \mathcal{E}[\mathcal{R}_a]}{K} \\
&= 3
\end{aligned}$$

Probability matching allocation rule

The probability of choosing the optimal operator $a^*_{ProbMatch}$:

$$\begin{aligned}
&\mathrm{Prob}[a^s = a^*_{ProbMatch}] \\
&= \quad P_{min} + (1 - K.P_{min})\frac{\mathcal{E}[\mathcal{R}_{a^*}]}{\sum_{a=1}^{K}\mathcal{E}[\mathcal{R}_a]} = 0.2666\ldots
\end{aligned}$$

The expected reward:

$$\begin{aligned}
&\mathcal{E}[\mathcal{R}^{ProbMatch}] \\
&= \quad \sum_{a=1}^{K}\mathcal{E}[\mathcal{R}_a]\mathrm{Prob}[a^s = a] \\
&= \quad \sum_{a=1}^{K} a[P_{min} + (1 - K \cdot P_{min})\frac{\mathcal{E}[\mathcal{R}_a]}{\sum_{a=1}^{K}\mathcal{E}[\mathcal{R}_a]}] \\
&= \quad 3.333\ldots
\end{aligned}$$

Adaptive pursuit allocation rule

The probability of choosing the optimal operator $a^*_{AdaPursuit}$:

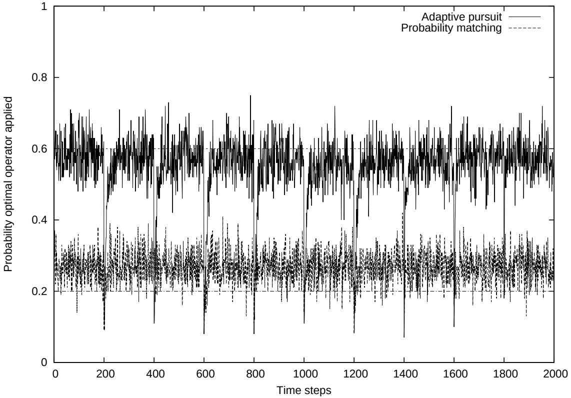
$$\begin{aligned}\text{Prob}[a^s = a^*_{AdaPursuit}] &= 1 - (K-1) \cdot P_{min} \\ &= 0.6\end{aligned}$$

The expected reward:

$$\begin{aligned}&\mathcal{E}[\mathcal{R}^{AdaPursuit}] \\ &= \sum_{a=1}^{K} \mathcal{E}[\mathcal{R}_a]\text{Prob}[a^s = a] \\ &= P_{max}\, \mathcal{E}[\mathcal{R}_{a^*}] + P_{min} \sum_{a=1, a \neq a^*}^{K} \mathcal{E}[\mathcal{R}_a] \\ &= 4\end{aligned}$$

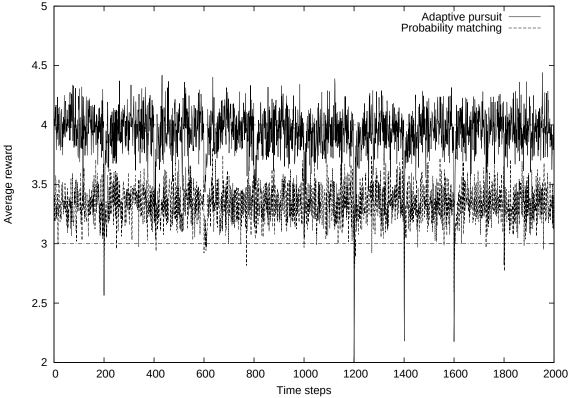
Probability of Selecting the Optimal Operator

($\Delta T = 200$; $\alpha = 0.8$; $\beta = 0.8$; $P_{min} = 0.1$; $K = 5$)



Reward Received

($\Delta T = 200;\ \alpha = 0.8;\ \beta = 0.8;\ P_{min} = 0.1;\ K = 5$)



Probability of Selecting the Optimal Operator

$(\Delta T = 200;\ P_{min} = 0.1;\ K = 5)$

α	*Probab. Match.*	*Adaptive Pursuit:* (β) 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.10	0.247	0.399	0.414	0.416	0.422	0.423	0.427	0.422	0.423	0.429
0.20	0.257	0.491	0.498	0.508	0.508	0.509	0.515	0.514	0.511	0.516
0.30	0.260	0.520	0.530	0.537	0.537	0.538	0.542	0.540	0.543	0.547
0.40	0.264	0.534	0.546	0.550	0.551	0.554	0.556	0.555	0.559	0.558
0.50	0.265	0.539	0.553	0.557	0.557	0.559	0.559	0.561	0.561	0.562
0.60	0.264	0.537	0.552	0.556	0.558	0.561	0.562	0.565	0.564	0.563
0.70	0.264	0.538	0.552	0.555	0.556	0.560	0.560	0.561	0.560	0.561
0.80	0.267	0.528	0.541	0.549	0.550	0.552	0.557	0.554	0.556	0.560
0.90	0.266	0.521	0.537	0.538	0.546	0.547	0.547	0.549	0.550	0.553

Reward received

$(\Delta T = 200;\ P_{min} = 0.1;\ K = 5)$

α	Probab. Match.	Adaptive Pursuit: (β)								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.10	3.233	3.719	3.757	3.767	3.768	3.775	3.778	3.780	3.776	3.789
0.20	3.287	3.834	3.853	3.877	3.879	3.879	3.893	3.891	3.887	3.892
0.30	3.302	3.873	3.896	3.916	3.912	3.914	3.922	3.921	3.923	3.934
0.40	3.315	3.886	3.915	3.926	3.932	3.933	3.939	3.942	3.948	3.938
0.50	3.320	3.891	3.925	3.940	3.939	3.945	3.940	3.946	3.946	3.950
0.60	3.323	3.890	3.926	3.936	3.941	3.949	3.947	3.956	3.955	3.951
0.70	3.322	3.894	3.928	3.936	3.943	3.948	3.948	3.947	3.947	3.951
0.80	3.333	3.878	3.912	3.934	3.937	3.934	3.946	3.940	3.945	3.951
0.90	3.329	3.881	3.916	3.913	3.933	3.933	3.933	3.938	3.936	3.944

Conclusion

- Probability matching
  $\Rightarrow$ low probability of applying best operator and low expected reward
- Adaptive pursuit
  $\Rightarrow$ high probability of applying best operator and high expected reward
  $\Rightarrow$ able to react swiftly at changes of the reward distribution