The Hierarchical Discrete Learning Automaton Suitable for Environments with Many Actions and High Accuracy Requirements

Rebekka Olsson Omslandseter[1], Lei Jiao[1], Xuan Zhang[2], Anis Yazidi[3], and B. John Oommen[4]

[1]Dept. of Information and Communication Technology, University of Agder (UiA), Grimstad, Norway
[2]Norwegian Research Centre (NORCE), 4879, Grimstad, Norway
[3]Oslo Metropolitan University, 0167, Oslo, Norway
[4]Carleton University, Ottawa, Canada

Abstract

Since its early beginning, the paradigm of Learning Automata (LA), has attracted much interest. The concept of incorporating *structure* into the ordering of the LA's actions is one of the latest advancements to the field, leading to the ε −optimal Hierarchical Continuous Pursuit LA (HCPA) that has superior performance to other LA variants when the number of actions is *large*. Although the previously proposed HCPA is powerful, it has a slow convergence when the required action probability of an action is approaching unity. Therefore, we propose the novel Hierarchical Discrete Learning Automata (HDPA) in this paper, which does not possess the same impediment as the HCPA. The proposed machine infuse the principle of discretization into the action probability vector's updating functionality, where this type of updating is invoked recursively at every depth within a hierarchical tree structure, and we pursue the best estimated action in all iterations through utilization of the Estimator phenomenon. The HDPA is ε-optimal, and our experimental results demonstrate that the number of iterations required before convergence is significantly reduced for the HDPA, when compared with the HCPA.

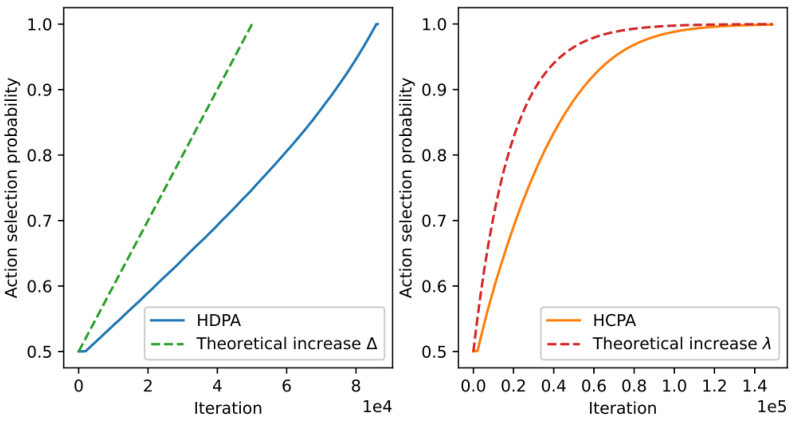

Fig. 2. The action probability of the optimal action per iteration compared to the theoretical increase in the action probability vector for the different schemes. The proposed HDPA on the left side, and the existing HCPA (state-of-the-art) on the right side.

Contributions

- We propose the novel HDPA that converges faster than the state-of-the-art HCPA algorithm, when the accuracy requirement is high. The advantages become more pronounced when a large number of actions exist, and the higher the accuracy requirement is.
- Via extensive simulations, we demonstrate how much faster the HDPA converged than the HCPA for Environments with *many* actions and high accuracy requirements.

The Proposed Concept

The concept of the HDPA is to utilize VSSA, discretizing the probability space, structuring the LA instances in a hierarchical tree structure and incorporating the Estimator concept. In more detail, we organize a set of Discrete Pursuit Automata (DPA) instances in a tree structure, where each instance has a set of actions corresponding to the possible paths down the tree structure from that automaton. The probabilities of these actions are maintained through vectors that are updated in a discretized manner. At the bottom level of the tree, we have the actions that directly interact with the Environment. The HDPA maintains reward estimates of all the actions throughout the tree structure, and we pursue the action with the currently best reward estimate in all iterations according to the pursuit paradigm. The reader should note that, in reality, the reward estimates are only necessary for the actions at the leaf level. We utilize the reward estimates in this way, because the proof of the algorithm's convergence needs the reward estimates along the path. In Fig. 1, we have an example of the HDPA for a four actions Environment. Each LA, denoted by $\mathcal{A}$ with the subscript representing its level in the hierarchy and the LA number at that level. At the bottom level we have the four actions interacting directly with the Environment, denoted by α, with the same subscripting as for the LAs.

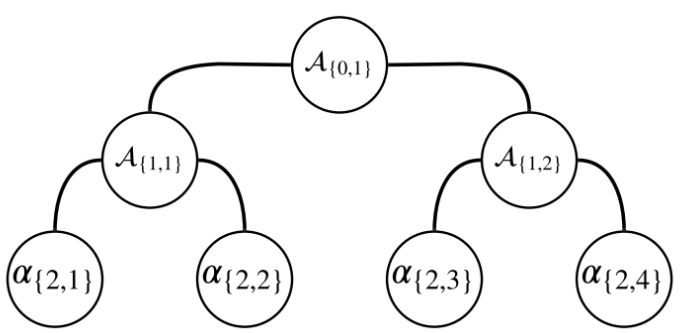

Fig. 1: Example of the proposed HDPA hierarchical tree structure of DPA instances for a four actions Environment.

Experimental Results

In the field of LA, a learning scheme's performance is often measured through the number of iterations required before the algorithm has converged. Due to the stochastic nature of the Environments that LA operates in, we normally measure the average number of iterations, i.e., we conduct many experiments and report the average number of iterations required before convergence over a number of experiments. In VSSA, the LA has achieved convergence once the action probability of any one of the actions has reached a certain threshold. The convergence criterion threshold is often configured close to unity. In these simulations, we configured this threshold to 0.992, and considered the average of the schemes' performance for 600 experiments. As presented in Tables 1 and 2, the proposed HDPA required considerably less number of iterations on average, when compared with the state-of-the-art HCPA.

Tab. 1: Experimental results for the state-of-the-art HCPA.

Number of actions	Mean	Standard deviation (Std.)
16	1,366.61	121.14
32	10,281.84	681.82
64	169,839.67	13,687.48
128	155,088.62	10,613.21

Tab. 2: Experimental results for the proposed HDPA.

Number of actions	Mean	Standard deviation (Std.)
16	868.25	135.50
32	6,172.38	744.84
64	100,638.41	17,653.41
128	97,795.59	13,266.12

Conclusion

In this paper, we proposed the HDPA scheme. The HDPA incorporates all the major phenomena within LA that have improved these algorithms over the last six decades. By implementing VSSA probability updating functionality and discretizing the probability space, utilizing the Estimator phenomenon, and structuring the LA in a hierarchical tree structure akin to the concept of binary search, the HDPA outperforms the state-of-the-art HCPA when the convergence criterion is close to unity, i.e., when the accuracy requirement is high. Our simulations demonstrated the clear advantage of the HDPA over the HCPA for different Environments with many actions.