Algorithms or Actions? A Study in Large-Scale Reinforcement Learning Supplemental material

Anderson Rocha Tavares^{*1}, **Sivasubramanian Anbalagan**^{*2},
Leandro Soriano Marcolino^{2}, **Luiz Chaimowicz**^{1}
^{1} Computer Science Department – Universidade Federal de Minas Gerais
^{2} School of Computing and Communications – Lancaster University
anderson@dcc.ufmg.br, siva.anbalagan@gmail.com,
l.marcolino@lancaster.ac.uk, chaimo@dcc.ufmg.br

Abstract

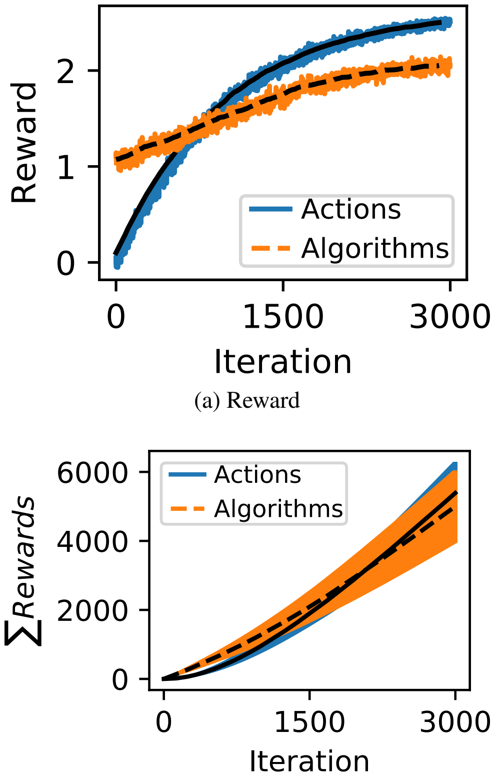
This document presents supplemental material, with further results associated with the original paper.

A Appendix

In this section we evaluate our results in terms of rewards and cumulative rewards. First, in Figure 1 we show examples of reward and cumulative reward graphs for a Gaussian model, similarly as in Figure 2 of the original paper. We can see a similar result as in our theoretical study: learning over algorithms outperforms learning over actions for a finite number of training iterations. Here we will define τ as the iteration where the reward (or cumulative reward) of learning over actions meets the reward (or cumulative reward) of learning over algorithms.

As before, we evaluate how τ changes with problem size ($|\mathbf{A}|$), number of algorithms ($|\mathbf{X}|$), u and μ, but now in terms of rewards and cumulative rewards (Figures 2 and 3, respectively). We can observe similar results as when evaluating the probability of playing the best action: τ increases with statistical significance under all parameters considered.

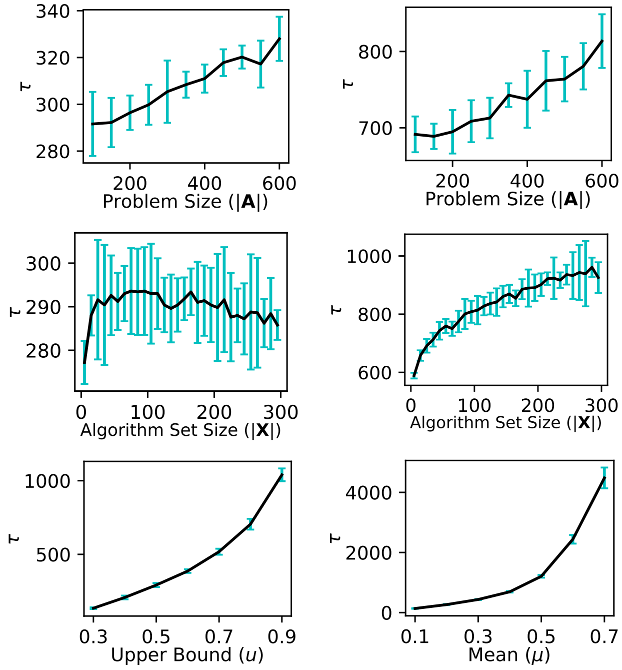
Additionally, we note that τ tends to converge as algorithm set size ($|\mathbf{X}|$) grows, instead of dropping after $|\mathbf{X}| > |\mathbf{A}|$; in a similar fashion as when we evaluated the probability of playing the best action (p_{a^*}) in Section 3 of the original paper. It is interesting to note, however, that τ seems to be slowly dropping (when considering the reward or cumulative reward) for the uniform model, as $|\mathbf{X}|$ gets much greater than $|\mathbf{A}|$. This is expected, since it gets harder for the agent to find the best algorithm.



(a) Reward

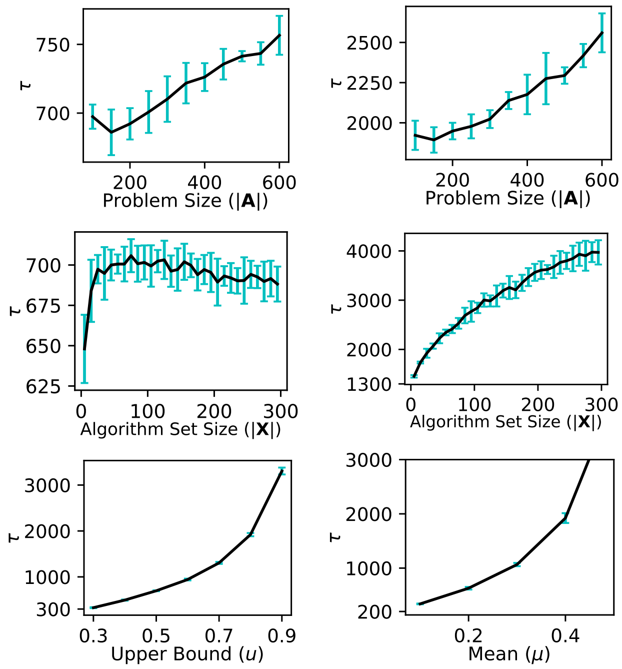
(b) Cumulative Reward

Figure 1: Example of reward and cumulative reward curves, from the synthetic experiments.



(a) Uniform Model (b) Gaussian Model

Figure 2: τ as number of actions, algorithms, u and μ grows, in terms of reward.



(a) Uniform Model (b) Gaussian Model

Figure 3: τ as number of actions, algorithms, u and μ grows, in terms of cumulative reward.