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

Anderson Rocha Tavares¹, Sivasubramanian Anbalagan²,
Leandro Soriano Marcolino², Luiz Chaimowicz¹
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 $\tau$ 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 $\tau$ changes with problem size ($|A|$), number of algorithms ($|X|$), $u$ and $\mu$, 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: $\tau$ increases with statistical significance under all parameters considered.

Additionally, we note that $\tau$ tends to converge as algorithm set size ($|X|$) grows, instead of dropping after $|X| > |A|$; 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 $\tau$ seems to be slowly dropping (when considering the reward or cumulative reward) for the uniform model, as $|X|$ gets much greater than $|A|$. This is expected, since it gets harder for the agent to find the best algorithm.

![Graph](image_url)  
Figure 1: Example of reward and cumulative reward curves, from the synthetic experiments.
Figure 2: $\tau$ as number of actions, algorithms, $u$ and $\mu$ grows, in terms of reward.

Figure 3: $\tau$ as number of actions, algorithms, $u$ and $\mu$ grows, in terms of cumulative reward.