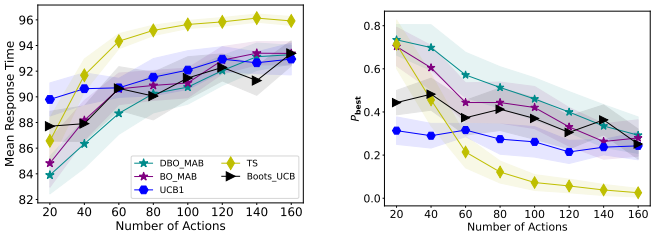
Technical Appendix
An Online Incremental Learning Approach for Configuring Multi-arm Bandits Algorithms

Mohammad Essa Alsomali[a,*], **Roberto Rodrigues-Filho**[b], **Leandro Soriano Marcolino**[a] **and Barry Porter**[a]

[a]School of Computing and Communications, Lancaster University, United Kingdom
[b]Department of Computing, Federal University of Santa Catarina, Brazil

Table 1: Computational Efficiency (s)

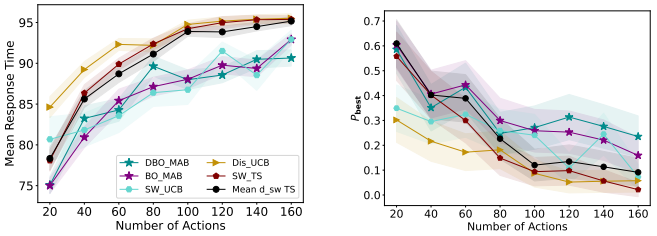
Method	80	100	120	140	160
Update_BO	937	868	854	953	907
Discard_BO	345	371	362	384	392
DBO_MAB	439	470	441	466	490
BO_MAB	400	428	416	489	482



(a) Average performance of the learning agent

(b) Probability of selecting the best action

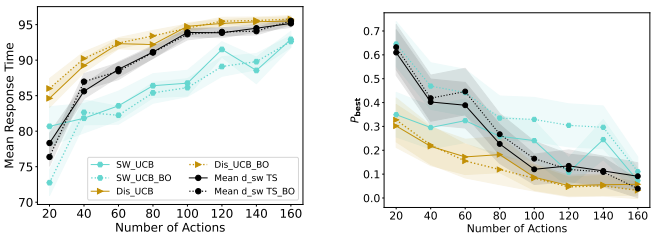
Figure 1: Comparison of the proposed method against parameter-free strategies and UCB1 with c default value, across an average workload range of 75-100.



(a) Average performance (lower is better) of the learning agent

(b) Probability (higher is better) of selecting the best action

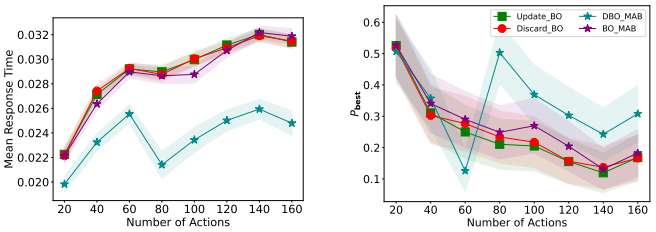
Figure 2: Comparing best-performing agent across multiple actions for dynamic high average workload between 60-100.



(a) Average performance (lower is better) of the learning agent

(b) Probability (higher is better) of selecting the best action

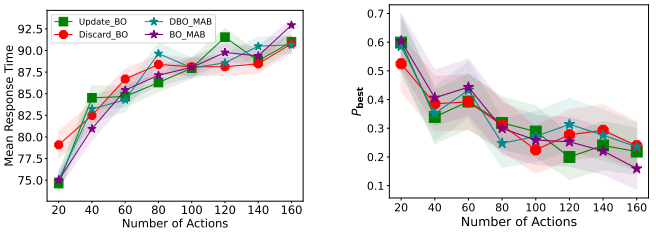
Figure 3: Comparing best-performing agent across multiple actions for dynamic high average workload between 60-100 by applying DBO-MAB method.



(a) Average performance of $i = 160$.

(b) Average performance of the learning agent for dynamic workload.

Figure 4: Performance analysis of DBO-MAB (and its variant) against ablation methods for a dynamic workload $(0 - 0.1)$.

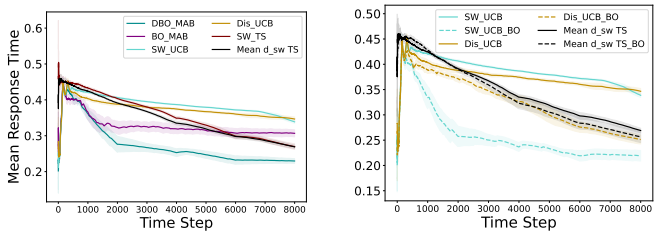


(a) Average performance (lower is better) of the learning agent

(b) Probability (higher is better) of selecting the best action

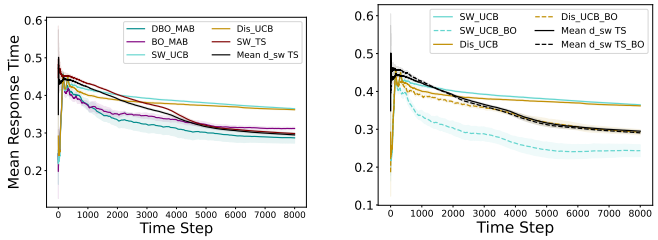
Figure 5: Comparing best-performing agent across multiple actions for dynamic high average workload between 60-100.

* Corresponding Author. Email: {m.alsomali, l.marcolino, b.f.porter}@lancaster.ac.uk, roberto.filho@ufsc.br



(a) Sudden environmental changes (b) Sudden environmental changes

Figure 6: Scalability of MAB algorithms under sudden environmental changes. The graph shows mean response time for 160 actions for optimized and non-optimized versions of MAB.



(a) Incremental environmental changes (b) Incremental environmental changes

Figure 7: Scalability of MAB algorithms under incremental environmental changes. The graph shows mean response time for 160 actions for optimized and non-optimized versions of MAB.

Table 2: SW-UCB-BO

Component	Sampling Range
c	$[0.0, 1.0]$
$windowsize$	$[10.0, 500.0]$

Table 3: Dis-UCB-BO

Component	Sampling Range
c	$[0.0, 1.0]$
γ	$[0.95, 0.99]$

Table 4: Mean d-sw TS-BO

Component	Sampling Range
$windowsize$	$[10.0, 500.0]$
γ	$[0.95, 0.99]$
α	$[0.0, 1.0]$
β	$[0.0, 1.0]$