

 : discrete action

 : continuous parameter action space

 $c_1 \dots c_w$: split each parameter space into W sub-regions

select w.r.t UCB

state

S

select discrete action

1

$Q_d(s, 1)$

2

$Q_d(s, 2)$

...

K

$Q_d(s, K)$

select discrete region

$Q_r(s, K, 1)$

$Q_r(s, K, 2)$

$Q_r(s, K, W)$

$c_1 \dots c_w$

$c_1 \dots c_w$

$c_1 c_2 \dots c_w$

μ_{K1}

μ_{K2}

...

μ_{KW}

sample continuous parameter

+ Gaussian noise

x_{K2}

forward phase

$n(s) += 1$

2

...

K

$Q_d(s, K) = \max_w Q_r(s, K, w)$

$n(s, K) += 1$

$c_1 \dots c_w$

$c_1 c_2 \dots c_w$

$Q_r(s, K, 2) = \max(Q_{K2})$

$n(s, K, 2) += 1$

μ_{K1}

μ_{K2}

...

μ_{KW}

$Q_{K2} \cdot \text{append}(Q(s, K, x_{K2}))$

estimated Q-value $Q(s, K, x_{K2})$

backward phase