

Source Users

$(U_s)$



$u_1$



$u_2$



$u_n$

Randomly Sample  
 $C$

Repeat  $t$  times

Candidate Set

$(C)$



$s_1$



$s_2$



$s_m$

State

$(S_L)$

Next State

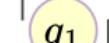
$(S_{L'})$

$\bar{\mu}_t$

DQN ( $Q_L$ )

Q values

$q_1$



$q_2$



$q_m$

$R_t = +1$

Selected Sample ( $a$ )



$s_i$

Cosine Relevance ( $r$ )

$r \geq \tau$

$r < \tau$

$R_t = -1$

$s_i$

$s_t$

Selected Sample Set

$(S^*)$

Update DQN

Target mean

Early Termination

Target Data ( $D_t$ )

$s_i$

$s_t$

$R_t = -1$

$s_i$

$s_t$

Score

$R_t$

$\odot$

Final Reward

$(R)$

Domain Adaptation ( $f$ )

$s_i$

$s_t$

Score

$R_t$

$\odot$

Final Reward

$(R)$

$(S_L, a, S'_L, R)$

Store transitions in Replay Buffer



Replay Buffer