

Take action $a^{(t)}$, which is sampled according to policy $\pi\left(a_i^{(t)}\mid s^{(t)}\right)$, to adjust temporal boundaries