



主結果

提案手法(TR-MAX)のサンプル量

R-MAXのサンプル量

[Brafman+, JMLR2002]

[Strehl+, JMLR2009]

$$O\left(\frac{|S||A|}{\epsilon^3(1-\gamma)^3} \left(|S| + \ln \frac{|S||A|}{\delta}\right) V_{\max}^3 \ln \frac{1}{\delta} \ln \frac{1}{\epsilon(1-\gamma)}\right)$$

$$O\left(\frac{|S||A| - |P_\epsilon|}{\epsilon^3(1-\gamma)^3} \left(|S| - \min_{(s,a) \in S \times A} |Z_{s,a}| + \ln \frac{|S||A| - |P_\epsilon|}{\delta}\right) V_{\max}^3 \ln \frac{1}{\delta} \ln \frac{1}{\epsilon(1-\gamma)}\right)$$

$Z_{s,a}$: 状態 s で行動 a を取ったとき遷移することがない状態の集合

P_ϵ : Source TaskとTarget Taskで状態遷移確率と報酬確率が似ている状態と行動の組の集合