P4-24　階層型方策勾配法の学習則

五十嵐　治一
芝浦工業大学
arashi50@sic.shibaura-it.ac.jp

◎大規模な強化学習問題に対する階層化モデルの提案

【特徴】

1. 方策勾配法(Williams'92)を使用
   - ➢環境ダイナミクス，方策，報酬に関するマルコフ性は要求しない
   - ➢方策の表現が柔軟・・・ヒューリスティクスの活用
2. 状態・行動の階層化・・・内部報酬の代わりにサブゴールを方策中に設定
3. 粗視化，記号化を統一的に扱う・・・空間分割，タスク分割に対応
4. 階層別の部分学習と全体の同時学習の両方が可能・・・全体の最適化

方策パラメータの学習則

報酬 r など

特徴的適正度

上層：$\frac{\partial}{\partial\theta'}E_{\pi,\pi'}\left[X(\sigma,\sigma')\right]\equiv\sum_{\sigma}\sum_{\sigma'}\frac{\partial}{\partial\theta'}P^{\pi,\pi'}(\sigma,\sigma')\cdot X(\sigma,\sigma')=E_{\pi,\pi'}\left[X(\sigma,\sigma')\sum_{\tau=0}^{L(\sigma')-1}e'_{\theta'}(\tau)\right]$

下層：$\frac{\partial}{\partial\theta}E_{\pi,\pi'}\left[X(\sigma,\sigma')\right]\equiv\sum_{\sigma}\sum_{\sigma'}\frac{\partial}{\partial\theta}P^{\pi,\pi'}(\sigma,\sigma')\cdot X(\sigma,\sigma')$

$=E_{\pi,\pi'}\left[X(\sigma,\sigma')\sum_{\tau=0}^{L(\sigma)-1}\left[\sum_{t=t(\tau)}^{t(\tau)+L(\sigma_\tau)-1}e_\theta(t;g_\tau)\right]\right]$

両層エピソードの出現確率の勾配

確率的勾配法

用いた方策中のパラメータだけを更新！

上層：$\Delta\theta'=+\varepsilon\cdot X(\sigma,\sigma')\sum_{\tau=0}^{L(\sigma')-1}e'_{\theta'}(\tau)$

下層：$\Delta\theta=+\varepsilon\cdot X(\sigma,\sigma')\sum_{\tau=0}^{L(\sigma)-1}\left[\sum_{t=t(\tau)}^{t(\tau)+L(\sigma_\tau)-1}e_\theta(t;g_\tau)\right]$

＊n層モデル，マルチエージェント系へ
そのまま拡張可能