

情報・知能工学専攻	学籍番号	093733
氏名	中島 健志	

指導教員氏名	村越 一支
--------	-------

論文要旨 (修士)

論文題目	未知の情報へ行動を誘発する拡散的好奇心を導入した強化学習手法
------	--------------------------------

強化学習における目的は環境からの報酬を最大にすることであるが、その目的には環境同定と報酬獲得の二つの性質がある。環境同定は最終的にできるだけ多くの報酬を得るため時間をかけてでも、最良の解を探索しようとするが報酬獲得は学習の途中でもなるべく報酬を得ることを目的とし、現時点で最も報酬の得られる行動をとろうとする。そのため最終的に得られる報酬と、学習が収束するまでの時間がトレードオフの関係にある。

この関係を解消するために提案された手法として、岡田ら (2001) の二次元評価強化学習がある。二次元評価強化学習では、報酬性評価と嫌悪性評価のそれぞれの予測された報酬との差である、TD(Temporal-Difference) 誤差の絶対値の和を重要度 Interest と呼び、報酬獲得と環境同定との比率を動的に決定することを可能にした。しかし Interest は行動選択時の行動のランダム性に影響するため、問題によって Interest が適切な値にならなかった場合、ランダム性が高くなりすぎて、不必要な探索を行ってしまう場合が考えられる。

学習をより早く収束させ、報酬値と学習時間のバランスがとれたより効率的な学習手法を目指すために、本研究では岡田ら (2001) が環境同定と報酬獲得の比率を決定する Interest と呼ぶ量を、生物における好奇心だと考えることができると述べている点に注目した。好奇心は、心理学の分野で拡散的好奇心と特殊的好奇心に大別され、本研究においては Interest を定義とは完全には一致しないが拡散的好奇心だとみなした。拡散的好奇心は未知の情報を求める好奇心であるのに対して、岡田らの手法における Interest はランダム性を高め、既知の情報、未知の情報に関わらず学習対象を決定してしまう。よって、より拡散的好奇心の定義に合った、未知の情報に対して行動を誘発するための拡散的好奇心の導入を検討する。

拡散的好奇心に関する研究として、章 (2011) は粒子群最適化の初期収束問題の解消に拡散的好奇心を利用する手法の研究を、Shimo ら (2006) は二種類の好奇心を学習に取り入れる研究を行っている。しかし、それぞれの研究は目的に特化した拡散的好奇心の利用であり、汎用的に拡散的好奇心による学習の効率化を提供するものではない。

本研究では、強化学習内部に拡散的好奇心の定義に従い未知の情報へ行動を誘発する行動選択手法を実現することで、問題に依存せず学習をより効率化することを目標とする。具体的には行動選択時に使用する行動優先度を、対応する情報を探索した回数に応じて高め、未知の情報へ行動を誘発する行動選択を実現する。提案した手法は、2 関節アームロボット前進問題において二次元評価強化学習の 4 割程度の時間で収束し、拡散的好奇心を実装することで、学習を高速化することが確認できた。また、迷路問題に実装することで、従来手法と比較して高い確率で真のゴールへの経路を学習し有効性が確認できた。