平成 16 年 1 月 9 日

知識情報工学専攻	学籍番号	981714
申請者氏名	水野　純也	

指導教官氏名	村越　一支

論　文　要　旨（修士）

論文題目	予期せぬ環境変化にも素早く追従する強化学習のパラメータ制御法

試行錯誤を通じて制御規則を自動的に獲得する学習制御の枠組として強化学習がある．これは行動した結果の報酬という情報を手がかりにして学習する教師なし学習である．ニューラルネットワークのような教師あり学習では，お手本となる教師信号を設計者が与える必要があるが，強化学習では環境とのフィードバックループを使って自ら学習していくことができるため，自律的かつ柔軟な学習が可能である．しかし，ある程度学習が収束した状態で，突然今までとは異なった環境が現れたら，再学習に時間がかかることが予想される．従来，このような問題に対して強化学習内部パラメータのいずれかを変化させる制御法がいくつか提案されている．その中のひとつとして Schweighofer *et al.*(2003)は強化学習内にある三つのパラメータα，β，γを少ない計算コストで柔軟に変化させる方法を提案している．しかし，この方法は確率的に各パラメータを変化させるために，素早い行動改善は期待できない．

そこで本研究では，強化学習内の三つのパラメータα，β，γを各々どの方向に変化させれば良いかをあらかじめ決定づけ，報酬の減り具合を用いてそれらのパラメータを動かすことで，予期せぬ環境変化に対しても素早く行動を改善させる方法を提案する．パラメータの変化させる方向は，強化学習パラメータと対応づけられる神経修飾物質系の，緊急時におけるふるまいに基づき決定した．

この提案手法を用いて，実際に学習問題に適用して計算機実験を行った．問題設定，学習手法に依存しないアルゴリズムであることが望ましいことから，学習手法としては，強化学習の代表的手法であるＱ学習，Actor-Critic を取り扱い，問題設定としては，状態行動空間が連続的なアームロボット前進問題，離散的な迷路問題を取り扱った．計算機実験の結果，制御を行わない従来の強化学習，確率的に三つのパラメータを変化させる Schweighofer *et al.*の制御法に比べ,提案手法が素早い行動改善に有効であった.さらに，提案手法内で三つのパラメータ制御のそれぞれが報酬獲得に貢献していることも示すことができた.

これらの結果から，提案手法を組み込んだ強化学習では，予期せぬ環境変化に対しても素早く行動を改善させられることが示された.