先端科学技術研究科　修士論文要旨

所属研究室 (主指導教員)	ロボットラーニング (松原　崇充　(教授))		
学籍番号	2111149	提出日	令和 5年 1月 19日
学生氏名	高橋　慶一郎		
論文題目	推論問題としての強化学習における非線形なTD誤差がもたらす挙動の解析		

要旨

近年のロボットは複雑な状況の中で最適に動くことが求められている. そのための有力な制御ルールの学習手法として強化学習(RL)が挙げられる.

RLは確率的推論問題として扱いにくいものとされてきたが, 近年ではRLをKullback-Leiblerダイバージェンス最適化問題として扱う新しい手法が存在し有効性が確認されている. しかし, 従来提案されてきた手法では, 報酬に関する未知の上限(あるいは下限)に依存する要素を近似により消去することで価値関数と方策の学習モデルの勾配を導出しているものと解釈できる. この近似によりTemporal Difference(TD)誤差に含まれる非線形項は消去され, 線形項のみが残される. しかし, この非線形項は従来とは異なる学習特性を持っている可能性がある. 想定できる特性として, 衝突や過大なエネルギー消費等のリスクに鋭敏に反応し回避しようとする性質や, 学習の進行に応じてモデルの更新が機敏なものから緩徐なものに変化する性質が考えられる. また, この非線形項には従来の式には存在しないハイパーパラメータが存在する. 非線形項を消去せずそのハイパーパラメータを学習の調整に用いることができれば, 学習率のように勾配全体のスケールを調整するだけではなく学習の細かな挙動調整が可能である. 具体的には, 扱う問題に応じてリスクのある状態に対する鋭敏さや良い状態への貪欲さを調整でき, これが性能向上に寄与する可能性がある.

本研究では, 上下限値が未知である報酬を正の報酬と負の報酬に区別する枠組みを導入することで, 報酬設計に関する一般性を保持したまま近似を回避し非線形項を残す新たなRLを提案する. また, この非線形項を用いたRLと線形項のみを用いた従来のRLを比較し, 非線形項が学習にもたらす挙動を解析する. 本研究ではシミュレーション上の振子の振り上げタスク, ロコモーションタスク, 作業用ロボットのタスクにおいて, この提案手法の学習特性を確認した. 実験の結果, 悪い状態に鋭敏に反応する特性や学習の勢いを緩める特性を確認でき, ハイパーパラメータでこれらを調整できることも確認できた. 特に, リスクのある状態に鋭敏な性質があるため, 従来のRLよりもリスクを抑えてタスクを達成できた.