論 文 の 内 容 の 要 旨

論文題目　　強化学習への事前知識の組み込みと
システム制御への応用

氏　　名　　青柳　祐基

強化学習とは，事前知識を用いずに，状態，行動および報酬の履歴のみから最適Bellman方程式をon-lineで解くアルゴリズムの総称である．強化学習は高い自律性を持ち，事前知識を必要としないが故の汎用性を長所とする反面，探索の初期に不安定な振る舞いをする，学習に時間を要するといった欠点により，制御問題への応用が進んでいるとは言い難い状況にある．以上を踏まえ，本論文の目的は，事前知識と強化学習の性能の関係について検討した上で，実際に事前知識の組み込んだ強化学習の手法を提案することである．さらに，耐故障制御を想定して宇宙機および航空機の姿勢制御問題に対して提案手法を適用し，その性能について評価する．

まず，探索における No Free Lunch 定理の一つの拡張として，強化学習における No Free Lunch 定理を提案し，強化学習の性能を向上するには何らかの事前知識を組み込むことにより，問題領域を限定する他ないと結論づけた．これは第2の目的である強化学習への事前知識の組み込みを支持する結果である．また，強化学習への事前知識の組み込み手法として，第1に「環境に関する知識の組み込み」として，対象の近似モデルが既知かつ報酬が二次形式で表されるという前提，すなわち環境についての不完全な事前知識が利用可能な場合に，Riccati方程式の解を利用した状態行動価値関数の初期化による学習効率の改善手法を提案している．第2に「ドメイン知識の利用」として，制御問題の多くにみられる，状態遷移が決定論的であるという前提そのものを一種のドメイン知識であると見做し，model-based 強化学習の代表的な枠組みであるDynaと，深層強化学習アルゴリズムであるDQNおよびDDPGを組み合わせた手法を提案した．

さらに，数値計算例として，「環境に関する事前知識の組み込みに」ついては Q-learningによる二次遅れ系の制御，ドメイン知識の利用については，深層強化学習アルゴリズムであるDQNおよびDDPGによるdouble integratorの制御ならびに，より応用的

な例として,DDPGによる宇宙機および航空機の姿勢制御に対する提案手法の有効性を検討している．その結果，「環境に関する知識の組み込み」については，初等的な例ながらも学習初期の不安定な振る舞いを改善し,安全な強化学習の１つの実装例としての効果を実証した．「ドメイン知識の利用」 については，まず，double integratorに対する計算例で，1stepあたりの計算時間は増えるが，学習に要するepisode数が大幅に減少する,すなわちサンプル効率が向上することを示した.さらに,より応用的な例として,宇宙機の3 軸姿勢制御則の自律獲得を試みた結果,既存手法よりも安定かつ高効率で学習を行えることが示され，さらには非 episode 型，すなわちonlineでの 3 軸姿勢制御則の獲得にも成功した．また，航空機の耐故障制御を想定した例では，推力のみによるphugoidモードの制御において，提案手法は既存手法に比べて静定時間およびピッチ角変動が大幅に減少するという高い性能を示した.

現在提案されている強化学習の手法の多くが確率論的環境を想定し，1ステップに1回しか価値関数および/または方策の更新を行わないことを鑑みれば，決定論的環境を前提とすることで，従来のmodel-based強化学習よりも高速に，かつ制御周期の中で時間が許す限りplanningを行う本論文の提案手法は,宇宙機や航空機の耐故障制御のみならず，強化学習一般の実問題への適用に大きく貢献するものである.