論文内容の要旨

博士論文題目

Hierarchical Decomposition and Min-max Strategy for Fast and Robust Reinforcement Learning in the Real Environment
（階層分割と Min-max 戦略による実環境での高速かつロバストな強化学習）

氏名
森本　淳

我々人間は、具体的なやり方を誰かに教えられなくても、試行錯誤の結果として新たな行動を獲得する能力を持つ。そのような学習システムの理解によって、人間が細かく指示を出すことなく、自ら目的を達成するための手段を発見するロボットやソフトウェアの開発が可能となる。

試行錯誤によってある評価を最大化するような行動を学習する枠組みは「強化学習」と呼ばれ、近年理論的解析と工学的応用研究が盛んに行われている。

これまで、強化学習の行動獲得課題や運動制御への応用は、計算機シミュレーション内での二次元迷路探索や倒立振子の振り上げのように、未知外乱を考慮する必要がない計算機内での低自由度の問題に関しては成功を収めて来た。しかし、人間は、未知外乱の存在する実環境中で自らの多自由度の体を制御している。これを従来の強化学習を用いて実現することは次の二点において困難である。

・多自由度の制御対象は、その自由度の組み合わせにより膨大な種類の状態を取ることができ、単純なランダム探索による強化学習は非常に困難である。特に、実環境での長時間の探索は、制御対象に高い耐久性を要求する。
・従来の強化学習によって獲得される制御則は特定の環境に対して最適化されるため、実環境における未知外乱や環境の変化に適応できない。

そこで、本論文では、強化学習に階層構造を導入することで高次元状態空間での学習を可能にするための手法と、環境モデルを用いた学習結果を実環境でも有効に活用するため、強化学習で獲得される行動則のロバスト性を向上する手法を提案する。

提案する階層型強化学習では、上位階層においては低次元離散空間で効率的に探索を行い、タスク達成のための離散的なサブゴール系列を学習する。下位階層においては、もともとの高次元状態空間中で、上位階層によって与えられたサブゴールに到達するための局所的な制御器を学習する。また、制御対象としては、３リンク２関節のロボットを考え、目標タスクとしては、起立運動課題を扱う。上位階層はQ学習を用いて、下位階層は連続系actor-critic法を用いて実現した。計算機シミュレーションにおいて、ロボットは750試行後に起立運動を獲得し、追加の170試行で実機においても起立運動を学習した。

さらに、本論文では、入力外乱やモデル誤差を考慮した強化学習法の提案を行う。強化学習では、シミュレーションによるオフライン学習や、行動のオンラインプラニングなど、環境や制御対象のダイナミクスモデルが重要な役割を果たす。しかし、実際の環境とモデルとの間の誤差のために、学習した制御器を実際の制御対象にそのまま利用すると、望みの性能が得られない可能性がある。そこで、H∞制御理論の考え方に基づき、外乱生成器が最悪外乱を出力し、行動生成器が最適制御を行う微分ゲームを考える。この問題は、外乱による報酬の変化と、外乱自体の大きさを考慮した評価関数のmin-max解を見つける問題として定式化できる。この知見を用いて、オンラインで評価関数の推定と最悪外乱、最適制御の計算を行う手法を示す。提案する学習法を単振り子の振り上げ課題に適用し、従来の強化学習によって獲得される制御則では対応できないようなモデル誤差に対してロバストな制御ができることを示す。

氏　名	森　本　　淳

（論文審査結果の要旨）

本論文は，階層分割とMin-max戦略による高速かつロバストな実環境での強化学習法について述べている．

試行錯誤によってある評価を最大化するような行動を学習する枠組みは「強化学習」と呼ばれ，近年理論的解析と工学的応用研究が盛んに行われている．しかし，強化学習を実環境における応用問題に適用するとき，高次元状態空間において必要な学習時間が指数関数的に増大することや，未知外乱やモデル誤差の存在が問題となる．

この問題を解決するために，本研究では階層型強化学習法，ロバスト強化学習法を提案している．本論文の成果は以下の２点に要約される．

・強化学習に階層構造を導入することが，タスクを達成するための行動獲得の高速化に有効であることが分った．
・制御理論の分野で開発されたH無限大制御理論の枠組みは，強化学習にも応用可能であり，未知外乱やモデル誤差を陽に考慮することが，強化学習によって獲得される行動則をロバストにすることが分った．

以上のように，本論文では階層構造を強化学習に導入することにより学習の高速化を行うことを提案し，その有効性を３リンク２関節の実ロボットを用いた動的起立運動の獲得により実証している．さらに，最悪外乱を考慮することで，強化学習によりロバストな行動則が得られることを実証している．

本研究は，強化学習の実用化を可能にするための重要な研究として評価でき，ロボティクス，機械学習の分野において，学術，実用の両面での貢献を認めることができる．なお，本論文の主要部分に相当する内容は，学会論文誌２件，解説論文２件，査読付国際会議４件として公表されている．

よって，本論文は博士（工学）の学位論文として価値あるものと認める．