

## 審査の結果の要旨

氏名 増山 岳人

増山岳人氏の博士論文は「行動経験の変換不変性に基づく移動ロボットの行動学習」と題し、全 6 章より構成される。本論文では未知環境で運用可能な自律ロボットの設計論の構築を背景とし、移動ロボットを制御対象とした強化学習の学習速度を向上する新たな枠組みを提案している。

第 1 章では、自律ロボット研究に関する研究領域の 1 つである発達ロボティクスを取り上げ、明示的な教師信号が与えられない中でロボットが自律的にそのセンソリモータ系を構造化していくための方法論について述べている。特に強化学習研究における、短時間的な行動を 1 つのまとまり（スキル）として、累増的にスキルを獲得しスキル間の関係性を記述する、内発的動機づけを導入した階層型強化学習が 1 つの有力な方法論であると主張している。また、最適性を意識した設計論に基づく従来手法の、階層構造の拡張に伴う見かけ上の学習速度の低下という、自律ロボット制御の方法論としてのボトルネックを指摘している。この問題を解決するために、本論文では知識に基づいて探索空間を積極的に縮減するトップダウン処理を、階層型強化学習に実装する枠組みの構築を目的とすることが述べられている。

第 2 章では、関連研究との対比から、本論文の位置付けと従来研究の問題点を明らかにしている。また、本論文の狙いであるトップダウン処理の枠組みについて概説されており、これは過去の成功経験から抽出されたスキルを、新たな環境において実行したときに観測される経験の再現性にもとづいて実装されると述べている。

第 3 章では、効率的に探索空間を縮減するためには、経験の再現性を計るための尺度として、距離のような概念ではなく、長距離的な依存関係をも記述し得る形式を与えることが重要であるとの認識に基づいて議論が展開されている。具体的な経験の再現性の尺度として、本論文では音声認識研究において提案さ

れたアファイン変換不変量に着目している。これはスキル実行に伴って観測されるセンサデータ系列に関する、アファイン変換によってモデル化される情報を捨象した特徴量として利用することができる。そのため、移動ロボットが扱う環境の幾何形状を反映したセンサデータから、スキルという行動の実行に伴う情報を抽象化することで、適応的に経験間の類似性を計る尺度として利用可能であることが示されている。

第4章では、第3章で導入した変換不変量によって計られる経験の再現性に基づいて与えられる内発的動機づけを導入した階層型強化学習手法を提案している。提案手法は行動価値と、各状態におけるスキルの価値を表すスキル価値を並列に学習する階層構造をもつ。行動価値の学習には一般的な強化学習手法が用いられる。他方、スキル価値の学習においては上述の内発的動機づけが導入される。スキル価値は、行動価値に基づいて行われる行動選択の過程に対し一時的なバイアスを与える。このバイアスによってスキルの指定する行動の選択確率が操作され、経験の再現性に基づいた状態行動空間における探索空間の縮減が実現されると述べている。

また、本論文で提案されるトップダウン処理による探索空間の縮減効果の有効性を検証するためのシミュレーション実験が示されている。実験結果より、最適性を意識した一般的な強化学習が価値関数のマップを学習する等方的な探索を行う一方で、提案手法ではスキルによって状態行動空間上に構成されるパスに沿った、指向的な探索が実現されることが示されている。また、代表的な階層型強化学習手法との比較実験によって、提案手法の探索空間の縮減効果と、階層型強化学習による意思決定の効率化の違いが示されている。

第5章では、第4章で提案した手法の問題点として、累増的に獲得することが想定されるスキルの多様化及びスキル数増大に伴う学習効率低下の問題が扱われている。この問題を解決するために、適格度トレースと呼ばれる、強化学習における基本的なメカニズムのスキルの類似度に基づいた拡張が行われている。これによりロボットの行動の履歴に基づいて、スキル価値が斉次的に学習されることでスキルの多様化とスキル数増大に伴う学習速度低下の問題が解決されると述べている。

シミュレーション実験により、様々なスキルをロボットが所持している場合にも、学習速度の低下が抑えられ、より一般的な条件でも提案されたトップダウン処理が有効に利用可能であることが示されている。

第6章では、本論文の成果が総括されている。また、今後の発展として強化

学習研究としての技術的な発展性，実ロボット制御問題に適用する際に想定される課題，そして自律ロボットの設計論に関する展望が述べられている。

以上，本論文では従来の階層型強化学習における，最適性を意識した等方的な探索に対して，実時間で準最適な解を素早く獲得する，自律ロボットの即時性を意識した探索空間縮減の枠組みを構築している．シミュレーション実験により，提案されたトップダウン処理の有用性が確認されており，博士論文として十分なオリジナリティとコントリビューションがあると判断する。

よって本論文は博士（工学）の学位請求論文として合格と認められる。