論 文 内 容 の 要 旨

博士論文題目

スケーラブルで安全かつ堅牢な強化学習のためのエントロピー正則化

(Entropy Regularization for Scalable, Safe and Robust Reinforcement Learning)

氏　　名

Zhu Lingwei

（論文内容の要旨）

本論文では、強化学習におけるエントロピー正則化の利用について検討する。エントロピーとしてシャノンエントロピーとカルバック・ライブラー距離を用いることで、価値ベースの強化学習において様々な有用な機能を得られることが示されている。具体的には、カルバック・ライブラー距離を用いることで、連続的に更新される方策間の距離が過度に大きくならないように拘束することができ、結果として価値関数や政策の近似誤差に対する頑健性の向上やサンプル効率の改善が期待できる。一方、シャノンエントロピーは、最適な政策を確率的に分布させ、探索を容易にする効果により、状態行動空間の効率的な探索が期待される。

本論文では、このようなエントロピー正則化に基づく強化学習理論をベースに、1)スケーラビリティ、2)安全性、3)堅牢性を特徴とする複数の強化学習アルゴリズムを導出し、数値シミュレーションによりその有効性を評価した。それぞれの詳細は以下の通りである。

1) スケーラビリティ：行動空間を分割し独立した複数の方策を設置し、それらを交互に学習するマルチエージェント型強化学習アルゴリズムを提案した。化学プラントシミュレータを用いた実験検証により、本アルゴリズムが合理的なサンプル数と計算量で、多数のバルブを有するプラント全体の自動運転化を達成できることを確認した。

2) 安全性：試行錯誤中に選択される危険な行動の選択回数を低減するために、タスク達成を志向する actor に加えて、制約充足を志向する advisor を持ち、両者の方策の違いをカルバック・ライブラー距離で制約する強化学習アルゴリ

ズムを提案した。ロボットアーム・ハンドモデルを用いた部品組み立てタスクにおいて、提案手法を用いることで、学習時に経験する危険行動の回数を比較手法に比べて減少させることができることを確認した。

3）堅牢性：強化学習アルゴリズムは、タスクごとにうまく調整する必要のあるパラメータが多数存在する。それらの設定を誤ると、学習性能の低下や、時には発散するなどの深刻な事態を招く可能性がある。このようなパラメータ調整の負担を低減するために、堅牢性を高めた強化学習アルゴリズムを提案した。方策の単調増加性を測る指標に基づき、エントロピー正則化に基づく下界近似により、実用的な近似指標とそれを利用する強化学習アルゴリズムを提案した。化学プラントモデルの課題において、提案手法がメタパラメータや行動空間の設計の違いに対して堅牢に学習できることを確認した。

氏　名	Zhu Lingwei

（論文審査結果の要旨）

本論文は、エントロピー正則化に基づく強化学習理論をベースに、1)スケーラビリティ、2)安全性、3)堅牢性を特徴とする複数の強化学習アルゴリズムを導出し、数値シミュレーションによりその有効性を評価したものである。それぞれの詳細は以下の通りである。

1) まず、強化学習のスケーラビリティ向上を目的として、行動空間を分割し独立した複数の方策を設置し、それらを交互に学習するマルチエージェント型強化学習アルゴリズムを提案した。化学プラントシミュレータを用いた実験検証により、本アルゴリズムが合理的なサンプル数と計算量で、多数のバルブを有するプラント全体の自動運転化を達成できることを示した。

2) 次に、強化学習の試行錯誤中の安全性向上を目的として、試行錯誤中に選択される危険な行動の選択回数を低減するために、タスク達成を志向する actor に加えて、制約充足を志向する advisor を持ち、両者の方策の違いをカルバック・ライブラー距離で制約する強化学習アルゴリズムを提案した。ロボットアーム・ハンドモデルを用いた部品組み立てタスクにおいて、提案手法を用いることで、学習時に経験する危険行動の回数を比較手法に比べて減少させることができることを示した。

3) 最後に、強化学習の堅牢性向上を目的として、パラメータ調整の負担を低減するために、堅牢性を高めた強化学習アルゴリズムを提案した。方策の単調増加性を測る指標に基づき、エントロピー正則化に基づく下界近似により、実用的な近似指標とそれを利用する強化学習アルゴリズムを提案した。化学プラントモデルの課題において、提案手法がメタパラメータや行動空間の設計の違いに対して学習を堅牢化できることを示した。

本論文はエントロピー正則化に基づく複数の強化学習アルゴリズムの開発と、ロボットおよび化学プラントシミュレータへの適用実験を行っていることに鑑み、新規性および有用性の観点から一定の学術的意義があるものと評価できる。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。