論 文 内 容 の 要 旨

博士論文題目

Efficient Task-independent Reinforcement Learning Methods based on Policy Gradient

（方策勾配に基づく効率の良い課題非依存な強化学習法）

氏　　名

森村　哲郎

方策勾配強化学習法は，エージェントが環境と相互作用する際に得られる報酬の平均値を目的関数とし，この目的関数を局所最大化する方策（行動則）の獲得を目指した方策探索法で，方策パラメータを目的関数の勾配により逐次更新することで実現される．方策さえ適切にパラメータ化すればエージェントや環境に関する知識を必要とせず直ちに（部分）マルコフ決定過程に実装可能である．そのため方策勾配強化学習法は様々な分野への応用が期待されてはいるが，実用化に向けて解決すべき問題が少なくとも 3 つ挙げられる；1) 学習所要時間が膨大になり易い，2) メタパラメータの設定が困難，3) 適切な方策のパラメータ化が困難．そこで本論文では上記問題の解決を目指して，効率の良い方策勾配強化学習アルゴリズムを数理的に探った．問題 1) に対しては，特にプラトー（学習の停滞期間）に注目して，MDP の確率分布に対して各方策パラメータの敏感さの相違やその相関を考慮した自然勾配法の研究を行った．そこでは初めに，Kakade の提案した自然方策勾配(NPG)の逆行列演算を必要としない適応的な方法で推定する自然時間差分学習法（NTD アルゴリズム）を提案した．次に，自然方策勾配で必要とされるリーマン計量行列についても解析し，最適な方策への収束を遅くしている理由を学習すべきパラメータ空間の構造の性質から考察して新しい NPG を導出した．そして数値実験より，従来法に比べ特に状態数が多い場合でもプラトーに陥らず有効に働くことを示した．問題 2) に対しては，メタパラメータの中でもこれまで有効な調節法が提案されていない積算報酬の割引率に関する研究を行った．一般の方策勾配法により推定される方策パラメータに関する平均報酬の偏微分値は，状態の定常分布の偏微分の計算が困難であったため，その偏微分に関する項を無視したものであった．この影響（推定値の偏り）は割引率を 1 に近づければ減少するが，一方で分散は大きくなってしまう．そこで本論文では，逆方向マルコフ連鎖の性質を利用して定常分布の偏微分を推定する方法を導出し，割引率に依存しない新しい方策勾配法を提案した．問題 3) に対しては，方策が学習課題に十分な表現能をもつ時に 0 に収束する指標を導出することで方策の自動パラメタライズ法を考案した．数値実験により課題に応じた適切な隠れ素子数を持つパーセプトロンが獲得されることを確認した．

氏　名	森村 哲郎

（論文審査結果の要旨）

強化学習は，制御対象や環境に関する知識を多く必要としないため，明示的な教師信号が与えることが困難な様々な工学的問題に対して，潜在的に有効であると考えられている．しかしながら，実用化に向けては学習所要時間やメタパラメータの設定等の問題があった．これら問題に対する先行研究はあるが，それらは特定の課題を想定しており，学習課題の事前知識を利用したものであったため，汎用性に欠けていた．よって標準的な強化学習の枠組みに手を加えないアルゴリズムの改良が望まれている．本論文では，強化学習が抱える問題に対して課題に依存しない解決を目指し，効率の良い方策勾配強化学習アルゴリズムを数理的に探っている．本論文の主な成果は以下に要約される．

1. 学習所要時間の問題に対して，最適な方策への収束を遅くしている理由を学習すべきパラメータ空間の構造の性質から考察して新しい自然方策勾配（NPG）を導出した．従来のNPGと違って，提案NPGは方策の変化に応じた状態の定常分布の変化も考慮した勾配であることを示し，また数値実験より状態数が多い場合でもプラトーに陥らず有効に働くことを示した．
2. メタパラメータの中でも有効な調節法がなく，方策勾配推定量の偏り・分散のトレードオフ問題と直結する積算報酬の割引率の設定を必要としない方策勾配法を提案した．これは逆方向マルコフ連鎖の性質を定式化して定常分布の偏微分の推定を可能としたことにより実現した．割引率の設定が困難な課題に適用して提案法の有用性を示した．
3. 適切な方策の設定（パラメータ化）問題に対して，方策が学習課題に十分な表現能をもつ時0に収束する指標を導出した．また本指標を用いたパーセプトロンの隠れ層の素子数の自動調節アルゴリズムを提案した．数値実験により適切な隠れ素子数を持つパーセプトロンが獲得できることを確認した．

以上より本研究は，方策勾配強化学習法を適用する際に障害となっていた問題に対する解決法を与えているので，強化学習法の工学的な問題への適用可能性を広げることに貢献している．　よって，博士（工学）の学位論文として価値あるものと認める．