論 文 内 容 の 要 旨

博士論文題目

Practical Model-free Reinforcement Learning in Complex Robot Systems with High Dimensional States

(高次元状態を有する複雑なロボットシステムにおける実用的なモデルフリー強化学習)

氏　　名　　YUNDUAN CUI

(論文内容の要旨)

As a promising learning paradigm in recent years, reinforcement learning learns good policies by interacting with an unknown environment and thus is suitable to the scenario of controlling robots to explore in challenging tasks. On the other hand, both the main two groups of reinforcement learning algorithms, the value function approach and the policy search, are still impractical in model-free learning of complex robot systems due to several limitations. The value function approach learns value function over all states and actions without any prior knowledge but suffers from both the unstable learning process with insufficient real world samples and intractable computational complexity in high dimensional systems. The policy search efficiently finds an optimal solution in a local area while being sensitive to the initialization of a well parameterized policy based on some knowledge of the task and model.

The motivation of this thesis is to explore practical model-free reinforcement learning algorithm to control complex robot systems. Our main idea is to take advantages of both the value function approach and the policy search. We propose a new approach that focuses on learning the global value function from the local sample space defined by the current policy. The Kullback-Leibler divergence is employed to limit the over large policy update in order to generate samples in continuous and local areas. Other machine learning methods are then applied on the local samples to locally approximate the value function. This framework solves the high sampling cost and

intractable computational complexity withoug any prior knowledge of the model or task.

Two algorithms are proposed based on this framework: Local Update Dynamic Policy Programming (LUDPP) and Kernel Dynamic Policy Programming (KDPP). We first investigate the learning performance of the proposed methods in a range of simulation tasks including pendulum swing up and multiple DOF manipulator reaching, the proposed algorithms significantly outperform the conventional algorithms in high dimensional cases. Both LUDPP and KDPP are then successfully applied to control a Pneumatic Artificial Muscle (PAM) driven robotic hand, a high-dimensional system in finger position control and unscrew bottle cap task respectively while given limited samples and with ordinary computing resources. All results indicate the practicability of the proposed framework in controlling complex robot systems.

氏　名	崔 允端

（論文審査結果の要旨）

本論文では多自由度を持つ汎用的ロボットシステムの普及を目指して、ロボットが環境との相互作用を通じ、目的に応じた運動制御方策の獲得を可能にする研究を行っている。一般にロボット自体や相互作用する物体・環境も含めたシステムのモデル化は非常に複雑であるため、従来のモデルベース制御の適用は難しい。そこで、強化学習と呼ばれるデータ駆動型・モデルフリー制御が注目されている。しかしながら、高次元状態・行動空間に対して十分な制御性能を導くだけのサンプルデータを収集することは、時間的制約や故障リスクの観点から容易ではない。そのため、現在までに強化学習の実ロボットへの適用事例は極めて少ない。

本論文では、この問題の解決策を検討している。具体的には、大域的な最適解を探索する価値関数ベース強化学習と、事前知識によってパラメータ化された方策のパラメータを局所的に探索する方策ベース強化学習を組み合わせ、局所的に収集されたサンプルデータから大域的な価値関数を推定し方策を改善する新しいアプローチを提案している。このアプローチでは、パラメータ化方策の事前設計の必要がなく、次元の呪いも低減できると考えられるため、高次元ロボットシステムに対しても実用的なサンプル数での学習達成が期待できる。

上記の着想に基づいて、本論文では

1) 動的方策計画法 (Dynamic Policy Programming) に基づく強化学習アルゴリズム,

2) 局所更新動的方策計画法(Locally Update Dynamic Policy Programming),

3) カーネル動的方策計画法(Kernel Dynamic Policy Programming)

の３つのアルゴリズムを提案している。1) については、Azar らによって提案されたDPP 理論に基づいて、方策オン型の強化学習アルゴリズムを提案している。2) および3)については、最近傍法およびカーネル法を価値関数の関数近似器に応用することで、1)で提案したアルゴリズムの計算量の大幅な改善を実現している。2)および 3)は、様々な従来手法との比較実験により提案手法の有効性を示している。さらに、空気圧人口筋駆動・触覚センサ付きの多自由度ロボットハンドシステムの運動制御課題にも適用され、少ないサンプルデータ数および

計算量での学習を達成している。

これらの結果から、本論文は高次元状態空間システムに有効な強化学習アルゴリズム提案および高次元ロボットへの希少な適用事例として、新規性および有効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。