

氏 名	アヘメト オナト Ahmet Onat
学位(専攻分野)	博 士 (工 学)
学 位 記 番 号	工 博 第 1822 号
学位授与の日付	平 成 11 年 3 月 23 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研 究 科 ・ 専 攻	工 学 研 究 科 電 気 工 学 専 攻
学 位 論 文 題 目	Application of Recurrent Neural Networks to Reinforcement Learning under Incomplete Perception (リカレントニューラルネットワークの不完全知覚下での強化学習への応用)
	(主査)
論 文 調 査 委 員	教 授 上 田 院 亮 教 授 茨 木 俊 秀 教 授 荒 木 光 彦

論 文 内 容 の 要 旨

本論文は行動の評価を手がかりに自律的に望ましい行動を獲得する学習方式である強化学習において、環境の状態が直接観測可能でない場合に学習が困難になる「不完全知覚」の問題を取り上げ、学習エージェントのアーキテクチャとして内部にフィードバック結合を持つリカレントニューラルネットワークを使用することによりこの問題を解決する手法についての研究を述べたものである。

全体は9章に分けて構成されている。

第1章では研究の導入として、強化学習の意義と学習における不完全知覚問題の重要性を述べた後、本論文の構成を述べている。

第2章では行動の評価を手がかりに自律的に望ましい行動を学習する強化学習について、その概念、解法を構成する上で重要なモデルとなるMarkov決定過程とその解法である動的計画法、そして動的計画法の拡張として、Q-learning法などの強化学習アルゴリズムを紹介している。

第3章では強化学習において、環境の完全な状態が学習エージェントから直接観測できない不完全知覚の問題を取り上げ、動的計画法に基づくアルゴリズムでは学習が困難になることを紹介したのち、この問題に対する従来手法を観測能力を高めるアプローチとエージェント内部に状態を導入し、動的な挙動を可能にすることによって対処しようとするアプローチに分類して紹介している。

第4章では本論文で動的なエージェントを構成する手段として用いるリカレントニューラルネットワークについて、ネットワークのアーキテクチャと教師あり学習アルゴリズムについて代表的な手法を紹介している。

第5章ではリカレントニューラルネットワークにより構成された学習エージェントに強化学習アルゴリズムQ-learning法を教師あり学習アルゴリズムを介して結合する手法を提案している。

第6章では離散的な状態遷移を行う記号的环境において、第5章で提案された手法により不完全知覚下でも学習が可能になることを計算機実験により示している。ここではネットワークの構成法としてElman型のネットワークと野田によって提案されているX-modelを、また教師あり学習アルゴリズムとして誤差逆伝搬法とReal Time Recurrent Learning法を適用し、いずれの方法によっても学習が可能であることを示すとともに、X-modelやReal Time Recurrent Learning法が学習の繰り返し回数を低減する上では有効であること、計算時間の観点からは手数の多いReal Time Recurrent Learning法は負荷が高く、総合的にはX-modelと誤差逆伝搬法の組合せが有利であることを示した。

第7章では状態変数が連続変数である数値制御的な環境として、倒立振子の制御問題を取り上げ、第5章で提案された手法により、振子の角速度や台車の速度など状態変数の一部が観測できない場合にも学習により制御が可能であることを示し

ている。

第8章では不完全知覚の下でも確率的な行動決定を許容することにより次善の行動を学習可能にする手法として木村らによって提案された確率傾斜法を取り上げ、これをリカレントニューラルネットワークと組み合わせることにより次善の行動だけでなく、最適行動を学習可能にする手法を検討している。そして、行動・状態フィードバック型のネットワーク構成により学習能力の向上が可能であることを示すとともに、この構成特有の学習の不安定さを指摘し、それが生じる機構について分析している。

第9章では、本論文で得られた結果を取りまとめ、今後の研究の発展方向について展望している。

論文審査の結果の要旨

本論文は行動の評価を手がかりに自律的に望ましい行動を学習する強化学習において、環境の状態観測が不完全な場合に学習が困難になる「不完全知覚」の問題を取り上げ、学習エージェントとしてリカレントニューラルネットワークを使用する手法により、この問題を解決する研究について述べたものであり、得られた主な成果は以下の通りである。

1. リカレントニューラルネットワークと不完全知覚がない場合の強化学習アルゴリズムであるQ-learning法を教師あり学習アルゴリズムを介して結合する学習エージェントの構成法を提案した。
2. 1. で提案している学習エージェントを状態遷移が離散的な環境を持つ例題に適用し、その有効性を確認するとともに、リカレントニューラルネットワークのアーキテクチャとして野田のX-modelを、教師あり学習アルゴリズムとして誤差逆伝搬法とReal Time Recurrent Learning法を組み合わせることが最も有望な選択であることを各種手法の比較実験から示した。
3. 1. で提案している学習エージェントを状態が連続的な環境である倒立振子の制御問題に適用し、状態変数の一部が隠された不完全知覚下での学習が可能であることを数値実験により示した。
4. 確率的行動により不完全知覚の下でも次善の行動を学習可能にする確率傾斜法にリカレントニューラルネットワークを組み合わせる手法を提案し、これにより学習能力の向上が可能であることを示すとともに、この構成特有の学習の不安定さを指摘した。

以上を要するに、本論文は強化学習における不完全知覚の問題を解消する手法として、学習エージェントにリカレントニューラルネットワークを使用する方法を提案し、数値実験によりその有効性を示すとともに学習を効果的に行うためのネットワークの構成法や学習アルゴリズムの選択法などについての有用な知見を得ており、学術上、実際上寄与するところが少なくない。よって、本論文は博士（工学）の学術論文として価値のあるものと認める。また、平成11年2月23日、論文内容とそれに関連した試問を行った結果、合格と認めた。