

学位論文題名

A Study on Reinforcement Learning
for Multi-Agent Systems

(マルチエージェントシステム向き強化学習に関する研究)

学位論文内容の要旨

本論文では、ロボットの経路計画や動作計画問題等に対する柔軟なプランニング機構の構築を目的としている。このような問題に対する従来のアプローチでは、作業領域やロボットに関するモデルを予め設計者が記述し、そのモデルに基づいてロボットの動作を計画する。例えば、ロボットを目標位置に誘導する問題では、予め目標地点がどこなのか、どこに障害物があるのか、進行方向の直前に障害物があるときはどちら向きに回避するのかを与えた上で、ロボットの動作列を決定する。しかし、障害物が突発的に移動する場合など、ロボットが直面する状況を予めすべて予測し、その状況でロボットがとるべき行動を記述することは、問題が大規模になるにつれ非常に困難になる。換言すると、対象問題は時間とともに変化する環境におけるモデル化が困難な問題と考えることができる。したがって、ロボットあるいは計算機が自律的に動作し、学習によりそれらの状況を獲得することが望まれている。

このような要求を満たすための研究が、1990年代初頭より活発に行われはじめている。これらの研究の方向性として以下の二つが挙げられる。

(1) 作業空間の記述を行わずにすむように、作業環境に関する情報は未知として、試行錯誤を繰り返しながらそれらを獲得する要求仕様。

(2) 環境の変化に柔軟に対応する能力を獲得させることを目的とし、目標までの動作列をすべて決定し、それらを一度にロボットに適用するのではなく、一つずつ動作を行いながらその都度動作を決定し適用するような要求仕様。

このような要求に対するアプローチの一つとして、集中管理型のシステムと固定的なプログラムではなく、分散型のシステムと学習機構により実現することが考えられる。ここでは、分散型のシステムを複数の自律的に動作を行う動作主体(=エージェント)から構成されるマルチエージェントシステムとして統一的に扱う。このマルチエージェントシステムの利点として、例えば、多関節型ロボットアームに対しては、関節にエージェントを割り当てることにより、ロボットアームの一部分の故障といった問題に柔軟に対応することが可能であり、また高速な処理が可能になることが期待される。

ここで、このようにマルチエージェントシステムに対する学習方式が必要となるが、対象問題の性質から、学習においては明示的なモデルをもたない機構が必要となる。そのような学習法の一つに、強化学習と呼ばれる方式が提案されている。これは、行動の選択とその評価という短いサイクルをくり返し行うことにより、目標状態に至る行動列を学習するというものである。しかし、この学習方式は単一の学習主体用のものであり、マルチエージェントシステムに対する有効な学習方式は未だ提案されていない。

本論文では上述の点について中心的に議論を行い、それに基づいた柔軟なプランニン

グ機構の構築を目指した、マルチエージェントシステムと、そのシステム向きの学習法を提案することを目的としている。

本論文の構成は、以下の通りである。

第1章では、序論として、工学諸問題に対する柔軟なプランニングシステムの具備すべき条件から、制御対象に関するモデル構築の困難さについて考察を行なっている。その結果として、マルチエージェントシステムの必要性と、その学習機構としてマルチエージェントシステムを考慮した強化学習の必要性について言及している。そして、本論文の概要を述べている。

第2章では、まず、強化学習の研究の流れと代表的な強化学習アルゴリズムについて要約している。その後、状態、行動、評価といった強化学習の枠組みから、人工ニューラルネットワーク、分類子システムといった学習手法、及び山登り法、遺伝アルゴリズムといった探索手法の形式化を行っている。次に、探索と学習の違い、あるいはタスク処理とタスク学習の違いが処理中に利用する知識の違いであることを明らかにしている。この議論を拡張することにより、教師有り学習と教師なし学習、学習結果の一般化、単一エージェントによるタスク処理と複数エージェントによるタスク処理、マルチエージェントシステムと分散システム等、学習の様々な側面を統一的に記述している。この議論の最終的な結果として、マルチエージェントシステム向きの学習機構の具備すべき性質が教師なしのオンライン学習であることを導いている。

第3章では、マルチエージェントシステムにおける各エージェントの動作規範とその獲得を中心にすえ、システムにおける協調、競合といった様々な挙動の実現について議論を行っている。まず、本論文の目的の一つは環境の変化に柔軟に対するプランニングシステムの構築であるため、古典的プランニングの限界を明らかにすることによりマルチエージェントシステムの在り方を考察している。その結果として、システムは動作列を出力するのではなく短い動作をくり返し出力すること、行動の結果が予想と一致しなかったときの対応が必要であることが挙げられてる。次に、各エージェントの局所的な目標とシステム全体の目標との関係との関係により、エージェント間の相互作用は協調問題解決から交渉による問題解決という二つの側面が明らかになった。また、エージェントの学習すべき内容を、動作ルールと戦略等に分類を行っている。

第4章では、前章までの議論に基づいて、動作ルールの獲得を行うための強化学習を用いた基本的なエージェントを定義し、マルチエージェントシステムを提案している。このシステムを、ロボットアームの動作計画問題と、自律移動ロボットのナビゲーション問題に応用し、計算機シミュレーションによりマルチエージェントシステムに適用された強化学習の収束性、学習結果の内挿性、突発的な環境の変化への適応等を実験的に確認されている。またプランニングシステムの立場から、移動障害物や同一空間にとのロボットが存在する場合についてもシミュレーションを行い、マルチエージェントシステム向き強化学習が有効に動作していることを確認している。

第5章では、前章で定義したエージェントの限界を明らかにし、それを解消するためのマルチエージェントシステムの拡張を行っている。前章のエージェントでは、大局的な目標を実現するためには、局所的な目標を無視しなければならないようなデッドロック問題と、初期状態から目標状態までには長い遷移が必要とする長期的な動作列の学習問題に対して、有効な動作ルールを獲得するまでに長い時間を必要とするときが存在する。この問題に対して、以下の3種類の拡張が行われている。エージェント内部を複数学習モジュールに拡張、獲得された知識の一般化を行う機構の導入、そして通信機能の付加である。これらを、前章と同じ問題に適用し結果を比較することにより、拡張されたエージェントはいずれの場合も良好な結果が得られていることが確認されている。

第6章では、本論文の全体の総括として、提案したマルチエージェントシステム向き強化学習機構は、環境の変化も扱うことが可能な柔軟なプランニング機構に応用可能であることを述べている。

学位論文審査の要旨

主 査 教 授 嘉 数 侑 昇
副 査 教 授 島 公 脩
副 査 教 授 新 保 勝
副 査 教 授 宮 本 衛 市

学 位 論 文 題 名

A Study on Reinforcement Learning for Mult-Agent Systems

(マルチエージェントシステム向き強化学習に関する研究)

動的環境下で自律的に振る舞うシステムへの関心の高まりから、柔軟でボトムアップ的な行動計画生成、および適応的な行動獲得を行うための機構に関心が寄せられている。例えば、ロボットの動作計画問題等に対する研究の多くは、作業領域やロボットに関するモデルを予め設計者が記述し、ロボットが直面する状況を予めすべて予測しその状況でロボットがとるべき行動を記述することに基づいて、ロボットの動作を計画している。そのため、障害物が突発的に移動する場合、すべての状況を予想することは、問題が大規模になるにつれ非常に困難になる。この問題は言い換えると、時間とともに変化する環境におけるモデル化が困難な問題であり、ロボットあるいは計算機が自律的に動作し、学習によりそれらの状況を獲得するアプローチの発展が期待されている。

前者のボトムアップ的行動機構に関しては、複数の自律的に動作するエージェントが状況に応じてエージェント間の構造を変化させることによって対応するというマルチエージェントシステムが注目を集めている。一方、適応的な行動獲得に関しては、対象に関する明示的なモデルを持たずに学習可能な、強化学習に関する研究が行われている。しかし、これは本来単一エージェントのためのものであり、マルチエージェントシステムに適用することは困難であった。そのため、この両者の特性を活かした機構の実現がなされていなかった。

本論文はこの課題への取り組みを行い、マルチエージェント向き強化学習の提案とエージェントの機能構成の拡張による柔軟なプランニング機構の実現を図り、その有効性に関する研究結果をまとめたものである。その主要な成果は、次の3点に要約される。

1. マルチエージェントシステム向き強化学習として具備すべき条件を議論している。特にエージェント個々の局所的適応性と全体の適応の調和が学習戦略上の要であることを指摘し、学習内容を行動ルールと戦略に分類すべきであることを明らかにしている。
2. 強化学習を取り入れたマルチエージェントシステムの応用を図り、その学習収束性、内挿性、適応性について議論し、本手法の有効性を示すとともに、エージェントの機能構成に起因する限界を指摘している。
3. 強化学習戦略を有効に作用させるための拡張型学習エージェント機構として、複数学習モ

ジュール，獲得知識の一般化機構，通信機構を導入したシステムの提案を行い，本提案構成システムにより，マルチエージェントの学習に特有の問題である長期動作系列の学習問題の解決法を示している。

以上のように本論文は，当該分野固有の問題特性を明らかにし，要求される学習戦略機構の実現およびこの機構を有効活用するためのエージェント機構の構成について新知見を得ており，精密工学，知識工学，ロボット工学の進歩に貢献するところ大なるものである。

よって著者は，北海道大学博士（工学）の学位を授与される資格あるものと認める。