様式 C-19

科学研究費補助金研究成果報告書

平成２３年　５月２４日現在

機関番号：１７２０１
研究種目：若手研究（Ｂ）
研究期間：２００７～２０１０
課題番号：１９７００２１５
研究課題名（和文）　部分観測マルコフ決定過程理論に基づく高次脳機能の自動的かつ高速な実装
研究課題名（英文）　Automatic and rapid realization of higher brain functions by partially observable Markov decision processes
研究代表者
伊藤　秀昭（ITOH HIDEAKI）
佐賀大学・大学院工学系研究科・講師
研究者番号：２０３４５３７５

研究成果の概要（和文）：
本研究は、ゴール指向性推論・選択的注意・作業記憶の利用などの高次脳機能を包括的に実現するエージェントを設計することを目的としている。このような諸機能を設計者が作りこむのは容易ではないので、本研究では報酬最大化原理に基づきエージェントが環境にあわせて自動的に必要な機能を発現するように設計する。本研究では部分観測マルコフ決定過程(POMDP)理論を用いて、これを効率的に実現する手法の開発を行った。

研究成果の概要（英文）：
This study aims at making an agent that is equipped with various "higher brain functions" including the goal-directed reasoning, the selective attention, and the use of working memory. Since it is difficult to make such an agent by hand coding, I try to use the reward maximization principle in order to make an agent that can automatically realize the functions that are suitable for its surrounding environment. In this study, I have been developing a novel method for making such an agent efficiently, based on the theory of partially observable Markov decision processes (POMDPs).

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００７年度	600,000	0	600,000
２００８年度	500,000	150,000	650,000
２００９年度	500,000	150,000	650,000
２０１０年度	500,000	150,000	650,000
総　計	2,100,000	450,000	2,550,000

研究分野：　総合領域
科研費の分科・細目：　分科は情報学、細目は感性情報学・ソフトコンピューティング
キーワード：　POMDP、確率的最適制御、高次脳機能、推論、報酬最大化、適応制御、階層モデル、階層制御

１．研究開始当初の背景

実世界のような複雑な環境において、ヒトと同等以上に知的に行動する存在（エージェント）を創るにはどうすればよいだろうか。

ヒトは、環境からの入力に対してただ反射的に何らかの行動を出力するだけの存在ではない。選択的に何かに注意を向けたり、作業記憶に何かをとどめておいたり、推論したり、自らを省みて行動様式を改善したりというように、頭の中でいろいろなことをすることができ、これによって複雑な世界にうまく対応できている。このようないわゆる高次脳機能と呼ばれる機能を工学的に実現することが必要である。

しかし、このような機能をエージェントの

設計者があらかじめ作りこむのは容易ではない。実世界のような複雑な環境において、例えばいつ何に注意を向けるのがよいか、また、いつどのような推論をしたらよいかなどをあらかじめ設計者が決めておくのは困難である。

それに代わる有望な設計原理として報酬最大化原理（例えば強化学習など）がある。これは、エージェントを設計する際に、様々な機能を設計者がいちいち作りこんで実装するのではなく、環境がどのようになるのが望ましいかをエージェントに与えておき、その望ましい状態ができるだけ実現されるようにエージェントが環境にあわせて自動的に必要な機能を発現するように設計するというものである。これにより、個々の状況におけるエージェントの動作をあらかじめ設計しておくことができないような複雑な環境であっても、知的に行動するエージェントを作ることができる可能性がある。

ヒトの知的な行動自体も、この原理に基づいて実現されている可能性がある。近年の研究で、大脳基底核ー大脳新皮質系を始めとする脳の様々な部分が報酬最大化に重要な働きをしていることが分かってきている。例えばこの系に属する中脳ドーパミン細胞の活動は、強化学習における報酬誤差関数の振る舞いとかなり一致する。そしてこの系は注意や推論等の高次機能に深く関係することも示唆されている。

筆者らはこれらの理由から、高次脳機能を報酬最大化原理に基づいて実現することが有望であると考え、これまで研究を進めてきた。これまでは、推論、特にゴール指向性推論を研究した。そして、実際にゴール指向性推論機能を報酬最大化原理に基づいて自動的に発現させることができることを、計算機実験によって示すなどの研究を行ってきたところであった。

２．研究の目的

上記のような研究を発展させることにより、別の環境ではゴール指向性推論とは別の機能が最適なものとして発現し、また複数の機能を持つことが合理的な状況においてはそれらが共存して発現するようにもできるものと期待することができる。このような柔軟な機能が実現すれば、ヒトのような知的エージェントの創造へ大きく近づくはずである。

そこで本研究課題では、この実現、すなわち報酬最大化原理に基づいてゴール指向性推論だけではなく様々な高次脳機能を自動的に発現させることを目的とした。

本研究課題は、このように筆者らのこれまでの研究を発展させるものであるが、これまでの手法には問題があった。それは学習が遅い、すなわち必要な機能が発現するまでにエージェントが環境と非常に多く相互作用する必要がある、というものである。そのため、ゴール指向性推論だけを発現させるのにも非常に長い時間がかかってしまい、多様な機能を包括的に実現させるには至らなかった。この手法が高速でない理由としては、様々なものが考えられるが、一つの大きな理由として、エージェントが環境を直接には学習しないモデル無し強化学習手法であるためではないかと考えた。そこで、本研究では環境を学習しそれに基づいて必要な機能を発現させるという、モデルベーズドな報酬最大化手法を用いることにより高速な学習を目指すこととした。

具体的には、部分観測マルコフ決定過程(Partially Observable Markov Decision Process; 以下では略して POMDP と書く)の理論を用いることとした。POMDP は、環境を記述する確率モデルの一つであり、環境の状態をエージェントが完全には観測できない場合も考慮することができるという特徴がある。本研究では、エージェントは POMDP として環境をモデル化し、環境との相互作用を通じてこのモデルを学習するものとする。そして学習されたモデルにおいて報酬を最大化する行動を求め、実行するものとする。

このように、POMDP 理論を用いると環境の学習から行動まで一貫して設計することができる。また、確率モデルを用いるため、高次脳機能が必要となるような複雑な環境においても扱いやすいコンパクトなモデルで十分良い性能がでる可能性がある。さらに重要なことに、推論のような高次脳機能を扱う際にはエージェントにとって何が既知で何が未知であるかを考慮した報酬最大化が必要であるが、POMDP は上記の特徴から未知のものがある場合にも設計指針を与えてくれる。

これらの理由により POMDP 理論は様々な高次脳機能を実現するための枠組みとしてふさわしいと考えられる。そこで本研究では、POMDP を用いて環境モデルを学習し、そのモデルにおいて報酬を最大化する行動を求め、実行する、という方法でエージェントを作ることを目的とした。

３．研究の方法

本研究ではまず、ゴール指向性推論について、POMDP を用いることにより旧手法と比較してどの程度学習が高速化されるかを調べるものとした。まずは、筆者らが既に開発している POMDP の学習法や最適化法を用いて、

どの程度高速な学習が可能となるかついて調べることとした。

そして、各種の高次脳機能（選択的注意、作業記憶の利用、ゴール指向性推論以外の推論、内省など）について自動的な実装を試みることにした。その際、別々に実装させるのではなく、全ての機能を適切に組み合わせて使えるようにすることを目指す。例えば選択的注意の場合、経験に基づいて効率よく必要なものに注意を向けることを学習させるが、それとともに、その注意の効率よい向けかたについて推論する機能なども学習させる。これらがどの程度高速に学習できるかを調べるものとした。

POMDP はよい解法が開発されてきているが、問題によっては満足な解が得られないことがあるという可能性も考えられた。これは予想される困難のうちで最も大きなものであった。しかし、筆者は POMDP の解法について深く研究した経験があり、必要に応じて新たな手法を考えるものとした。また POMDP は現在研究が進んでいる分野なのでさらによい手法が開発される可能性も高く、良い手法が開発されればそれを用いることとし、それでもうまくいかない場合には POMDP ではない手法（非マルコフ的な枠組みを用いた手法）を用いることも考えるなど、柔軟に対応することとした。

また、環境の学習手法としては筆者が研究した経験のあるダイナミックベイジアンネットワークの online 学習法を想定したが、online 学習は計算量的に実行が難しい場合も予想された。その場合は batch 学習を採用するものとした。近年様々な batch 学習の研究が進んでおり、より高い性能が出せる可能性もあるので、必要に応じて比較・検討するものとした。

研究は、主として数式等を用いた理論的な解析と、計算機を用いた数値計算とによって進めるものとした。

４．研究成果

まず 1 年目は、「POMDP に基づくモデル有り学習」を可能とする既存の手法を実装し、有効性を調べる研究を行った。その結果、選択的注意および作業記憶の利用については簡単な問題を解くことに成功し、本研究のアプローチが有効であることを確認することができた。

例えば、選択的注意課題では、視覚センサと触覚センサがある場合に、どちらからの情報をより多く用いるかを自動的に学習させるという課題を行った。結果を図 1 に示すが、強化学習に比べて短い時間で最適解へ収束していることが分かる。

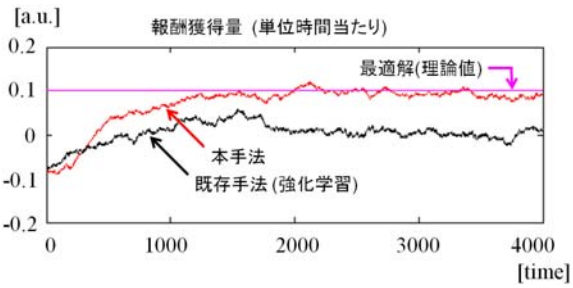

図 1　選択的注意課題の学習曲線

このように、比較的簡単な問題に対しては、本研究のアプローチが有効であることが分かったが、ゴール指向性推論を始めとして、本研究が実現目標として掲げているような複雑な機能を実際に実現するためには、既存手法では学習に無駄な部分が多いため、学習に時間がかかりすぎることも明らかとなった。特に、環境に同じような構造が繰り返し現れる場合に、既存手法では何度も学習し直す必要があり、非効率的であった。そこで、そのような場合にうまく対処することが可能な、より学習効率の高い手法が必要であることが分かった。

そこで、2 年目と 3 年目では、より高性能な手法の開発に取り組んだ。特に、階層性を取り入れたモデル推定法およびその最適制御法を研究し、理論的に優れた性質を持つ新たな手法の開発を行った。この手法は、環境モデルの online 学習と、学習されたモデルの最適制御とをどちらも階層的に行うもので、既存手法で扱うことが困難であった複雑な環境を高速に学習および制御できることを目指したものである。なお、本手法は online 学習だけではなく、batch 学習版を考えることもでき、場合によって使い分けることが可能である。

本手法については、まず簡単な問題を対象とした数値実験により既存手法との性能比較を行い、どのような場合に有効であるかを調べた。さらに、理論的な性能の解析も行い、どのような場合に本手法が有効であるかについて定理としてまとめることにも成功した。

4 年目は、この開発および理論的解析を進めるとともに、手法を一部の具体的な問題に適用する研究も行った。その結果、既存手法よりも優れた結果が得られることが多いことを示すことができた。

これらの結果については論文発表を準備中である。また、今後は、本手法をより多くの問題に適用して有効性を検討したいと考えている。

また、既存手法と新手法の比較検討の一環として、既存手法の問題点を調べるという研究も行った。その結果、ゴール指向性推論の

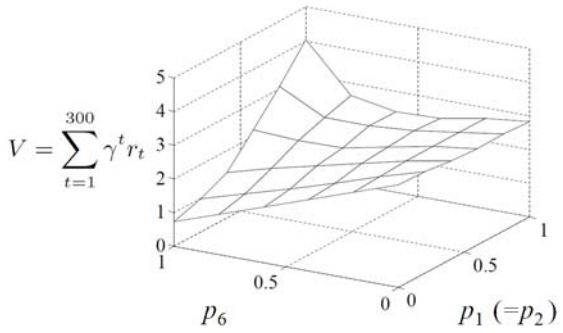

図２　最適化対象の関数の形状

学習における問題点の一つが、最適化とする対象となる関数の多峰性にあることが分かった（図2)。なお、図２で、p_1とp_6は学習パラメータ、縦軸が最大化したい値である。このような多峰性は、高次脳機能を必要とする多くの問題において共通に見られる、一般的な構造であると考えられ、既存手法がうまく行かない理由の一つを明らかにすることができた。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計１件）

① 毬山 利貞，伊藤 秀昭，中村 清彦，数理的モデルを用いたタマリンの道具使用行動の解析，霊長類研究，査読有，VOL. 26, NO. 1, pp. 13-33, 2010

〔学会発表〕（計４件）

① 伊藤 秀昭，福本 尚生，和久屋 寛，古川 達也，関係強化学習によるゴール指向性推論の学習，電子情報通信学会技術研究報告(AI, 人工知能と知識処理)，VOL. 110, NO. 301, pp. 1-6, 2010.11.19, 福岡
② Kenji Aoki, Hiroki Takahashi, Hideaki Itoh, Kiyohiko Nakamura, Comparison of Near-Threshold Characteristics of Flash Suppression and Forward Masking, International Conference on Neural Information Processing, 2009.12.3, Bangkok, Thailand
③ Toshisada Mariyama, Hideaki Itoh, Towards a Comparative Theory of the Primates' Tool-use Behavior, International Conference on Neural Information Processing, 2008.11.26, Auckland, New Zealand
④ 清川舞，伊藤秀昭，中村清彦，情動状態の思考による鎮静化現象の分析の試み，「脳と心のメカニズム」冬のワークショップ，2008.1.10，北海道

６．研究組織
(1)研究代表者
伊藤　秀昭（ITOH HIDEAKI）
佐賀大学・大学院工学系研究科・講師
研究者番号：２０３４５３７５

(2)研究分担者
該当無し

(3)連携研究者
該当無し