

平成 21 年 6 月 2 日現在

研究種目：基盤研究 (B)

研究期間：2006～2008

課題番号：18300101

研究課題名 (和文) 複雑な環境における脳の意思決定モデルとロボット制御への応用

研究課題名 (英文) Computational model of human decision-making in complicated environments and its applications

研究代表者

石井 信 (ISHII SHIN)

京都大学・大学院情報学研究科・教授

研究者番号：90294280

研究成果の概要：

線形確率システムとして表現できないような複雑あるいは動的な環境における最適意思決定過程を模擬する機械学習モデルを、強化学習に注目して構築し、工学応用、特に多自由度ロボットに対する自律制御に適用した。高度で階層的な推論を必要とするタスクを題材とし、複雑な問題解決に関わる脳内モデルを構築し、非侵襲脳活動計測器を用いた認知科学実験により検証した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006 年度	6,400,000	0	6,400,000
2007 年度	5,000,000	1,500,000	6,500,000
2008 年度	3,700,000	1,110,000	4,810,000
総計	15,100,000	2,610,000	17,710,000

研究分野：強化学習・計算論的認知心理学

科研費の分科・細目：情報学・生体生命情報学

キーワード：部分観測・前頭前野・意思決定・サンプリング・多自由ロボット

1. 研究開始当初の背景

強化学習は、元々動物の条件付け学習のモデルとして登場し、実際に中脳のドーパミン細胞の活動との関連性が指摘されて (Schultz, *Science*, 1997) より、神経科学分野において、大脳基底核、前頭前野、帯状回など報酬処理に関わる脳内部位との関連が調べられている。一方で、制御分野において 60 年代から研究されてきたマルコフ決定過程の解法として認識され、環境に適応するための学習法として、機械学習や人工知能の分野において独自の発展を遂げている。このように強化学習は動物学習 (生物学) と機械学習 (工学) との接点であり、環境に動的に適応する生物や機械の学習原理として注目されている。申請者らのグループは、数年来、機械学習としての強化学習法、特に環境のダイナミクスを明に内部モデルとして保持することで効率

を高めた、モデル同定型強化学習法の理論的研究 (Ishii, *Neural Networks*, 2002) とその各種工学応用、および前頭前皮質の機能に着目して、脳内実現の可能性の検討を進め、成果を得てきた。一方で、神経科学の分野において従来の強化学習理論では不十分であることが指摘され始め (Fiorillo, *Science*, 2003)、工学の分野においても方策勾配法 (Konda, *SIAM J. Control*, 2003) などのより安定した学習法が開発されてきている。本研究は、こうした世界的な研究動向を踏まえて、従来の強化学習を超えた理論を生物学的および工学的に開発しようとするものである。

2. 研究の目的

一般的で複雑な動的環境における動物の最適意思決定過程を模擬する機械学習モデルを強化学習に注目して構築し、それを用いた

工学的応用、特に多自由度ロボットに対して変動する環境下での自律制御を行う。一方で高度で階層的な推論を必要とするタスクを題材として、複雑な問題解決に関わる階層的な脳内モデルを構築し、非侵襲脳活動計算器を用いた認知科学実験により検証する。特に環境モデルの学習と推論に関わると考えている前頭前野と、習熟した系列の保持と実行に関わると考えている大脳基底核の役割分担に注目する。このことで、従来の価値関数に主眼をおく強化学習において、工学的にも生物学的にも新しい展開を与えることを目指すものである。神経レベルから、脳の機能分担、計算理論、行動制御に至るまでの動物の意思決定過程のモデル化を行いつつ、その実用レベルでの工学応用を図る融合研究を推進する。

3. 研究の方法

下記3点に重点をおいて研究を進める。

(1) 強化学習のアルゴリズム開発

方策勾配法ベースの強化学習法に注目しながら、新しい強化学習アルゴリズムを開発する。

(2) 階層的部分観測環境における神経基盤

部分観測課題における最適意思決定過程の神経基盤を調べるため、fMRI をもちいて認知科学実験を行い、結果を分析する。

(3) 多自由度ロボットの強化学習による制御

方策勾配法強化学習法および中枢パターン生成器に対する強化学習法を、実際にロボットに適用し、結果を分析する。

4. 研究成果

(1) 強化学習のアルゴリズム開発

方策勾配法ベースの強化学習法に注目しながら新しい強化学習アルゴリズムを開発した。サンプリングに基づく部分観測マルコフ決定過程の解法を開発し、4人で行うマルチエージェントゲームである Hearts の効率良い自動学習に成功した (Fujita and Ishii, 2007)。予測器と制御器の対が複数あるようなモジュール型アーキテクチャに適用することで、変動する環境での適応的制御を可能とした (Osaga, et al., 2008) (Hiei, et al., 2008)。方策内部にダイナミクスを保持した部分観測強化学習法アルゴリズムを開発し、メモリを保持するものよりも効率が良いことを見出した (Taniguchi, et al., 2008)。また、強化学習のための価値関数の学習において、統計的に最適となる手法について理論的な結果を得た (Ueno, et al., 2008)。

(2) 階層的部分観測環境における神経基盤

部分観測課題における最適意思決定過程の神経基盤を調べるために、fMRI を用いた認知神

経科学実験を行い、前部前頭前野の関わりを明らかにした (Yoshida and Ishii, 2006)。階層性のある意思決定課題においては、前頭前野の異なる領域が関わるということが分かった (Yoshida, et al., submitted)。また、複数の入力モダリティがある、あるいは階層性があるような部分観測環境における意思決定過程の神経基盤を調べるための実験を行い、ヒトが能動サンプリングによる効率良い環境同定を行っていることがわかったので、モデルを用いた脳画像解析を継続実施中である。

(3) 多自由度ロボットの強化学習による制御

中枢パターン生成器に対する方策勾配法ベースの強化学習法を2足歩行ロボットシミュレータに適用し、安定した歩容が実現されることを示した (Nakamura, et al., 2007)。また、同様の手法をヘビ型ロボットに適用し、アクチュエータの故障などシステムの変化に追従できることを示した (Makino, et al., 2008)。方策オフ型方策勾配強化学習法を2足歩行準受動歩行ロボットに適用し、自律的な環境適応が可能であることを示した (Ueno, et al., 2006)。

さらに、物理学的に相補的である2つのモジュール制御器、すなわちゼロモーメントポイントに基づく軌道追従制御と準受動歩行強化学習、とになるハイブリッド型制御法を、2足歩行ロボットシミュレータに実装し、評価のための実験を継続実施中である。

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17件)

① N, Yano., T, Shibata., S, Ishii. (2009). Adaptive particle allocation for multi-focal visual attention based on particle filtering. *Artificial Life and Robotics*, 査読有, 13, 522-525.

② M, Shikauchi., S, Ishii., T, Shibata. (2008). Prediction of aperiodic target sequences by saccades, *Behavioural Brain Research*. 査読有, 189, 325-331.

③ T, Tamei., S, Ishii., T, Shibata. (2008). Virtual force/tactile sensors for interactive machines using user's biological signals. *Advanced Robotics*, 査読有, 22, 893-911.

④ J, Hirayama., M, Nakatomi., T, Takenouchi., S, ishii. (2008). Collaborative prediction by multiple Bayesian networks and its application to printer usage modeling. *Behaviormetrika*, 査読有, 35, 99-114.

⑤ S, Osaga., J, Hirayama., T, Takenouchi., S, Ishii. (2008). A probabilistic modeling of MOSAIC learning. *Artificial Life and Robotics*, 査読有, 12, 167-171.

⑥ K, Makino., Y, Nakamura., T, Shibata., S, Ishii. (2008). Adaptive control of a looper-like robot based on the CPG-actor-critic method. *Artificial Life and Robotics*, 査読有, 12, 129-132.

⑦ Y, Nakamura., T, Mori., M, Sato., S, Ishii. (2007). Reinforcement learning for a biped robot based on a CPG-actor-critic method. *Neural Networks*, 査読有, 20, 723-735.

⑧ H, Fujita., S, Ishii. (2007). Model-based reinforcement learning for partially-observable games with sampling-based state estimation. *Neural Computation*, 査読有, 19(11), 3051-3087.

⑨ S, Ishii., W, Yoshida. (2006). Reinforcement learning: machine learning and natural learning. *New Generation Computing*, 査読有, 24, 325-350.

⑩ 銅谷 賢治, 石井 信. (2006). 学習ダイナミクスの制御と脳の物質機構. *システム/制御/情報*, 査読無, 50(8), 303-308.

⑪ 森 健, 中村 泰, 石井 信. (2006). 重点サンプリング法に基づく Natural Actor-Critic法による効果的なサンプルの再利用. *電子情報通信学会論文誌*, 査読有, J89-D(5), 954-966.

⑫ Y, Sakumura., S, Ishii. (2006). Stochastic resonance with a differential code in feedforward network with intra-layer random connections. *Neural Networks*, 査読有, 19, 469-476.

⑬ J, Hirayama., J, Yoshimoto., S, Ishii. (2006). Balancing plasticity and stability of on-line learning based on hierarchical Bayesian adaptation of forgetting factors. *Neurocomputing*, 査読有, 69, 1954-1961.

⑭ T, Bando., T, Shibata., K, Doya., S, Ishii. (2006). Switching particle filters for efficient visual tracking, *Journal of Robotics and Autonomous Systems*. 査読有, 54, 873-884.

⑮ Y, Igarashi., Y, Sakumura., S, Ishii. (2006). The role of short-term depression in sustained neural activity in the prefrontal cortex: A simulation study. *Neural Networks*, 査読有, 19(8), 1137-1152.

⑯ 坂東 誉司, 柴田 智広, 清水 幹郎, 石井 信. (2006). 適応的サンプリングによる階層モデル化された対象の効率的状態推定, *システム制御情報学会論文誌*, 査読有, 50(10),

369-377.

⑰ K, Hitomi., T, Shibata., Y, Nakamura., S, Ishii. (2006). Reinforcement learning for quasi-passive dynamic walking of an unstable biped robot. *Robotics and Autonomous System*, 査読有, 54(12), 982-988.

[学会発表] (計 18 件)

① K, Hitomi. (2008). Online multibody factorization based on Bayesian principal component analysis of Gaussian mixture models. In 15th International Conference on Neural Information Processing Lecture Notes in Computer Science 5506 (Auckland, 27 Nov., 2008), Part I, 678-686.

② A, Asahina. (2008). Interpreting dopamine activities in stochastic reward tasks. In 15th International Conference on Neural Information Processing (Auckland, 26 Nov., 2008).

③ Y, Hiei. (2008). Self-organized reinforcement learning based on policy gradient in nonstationary environment. In *Artificial Neural Networks - ICANN 2008, Lecture Notes in Computer Science 5163* (Prague, 3 Sep., 2008), Part I, 367-376.

④ Y, Taniguchi. (2008). A continuous internal-state controller for partially observable Markov decision processes. In *Artificial Neural Networks- ICANN 2008 Lecture Notes in Computer Science 5163* (Prague, 3 Sep., 2008), 397-406.

⑤ T, Ueno. (2008). A semiparametric statistical approach to model-free policy evaluation. In the 25th International Conference on Machine Learning (ICML) (Helsinki, 8 Jul., 2008), 1072-1078.

⑥ N, Yano. (2008). Adaptive particle allocation for multifocal visual attention based on particle filtering. In the 13th International Symposium on Artificial Life and Robotics (Beppu, 31 Jan., 2008), 13, 522-525.

⑦ N, Nanjo. (2008). An effective reinforcement learning with automatic construction of basis functions and sequential approximation. In the 13th International Symposium on Artificial Life and Robotics (Beppu, 31 Jan., 2008), 662-665.

⑧ Y, Taniguchi. (2008). Continuous internal-state controller for a partially observable linear dynamical system. In the 13th International Symposium on Artificial Life and Robotics (Beppu, 31 Jan., 2008), 674-677.

- ⑨K, Takeda. (2008). Active sampling based on Gaussian process for reinforcement learning. In the thirteenth International Symposium on Artificial Life and Robotics (Beppu, 31 Jan., 2008), 678-681.
- ⑩ Y, Taniguchi. (2007). Reinforcement learning for cooperative actions in a partially observable multi-agent system. In Artificial Neural Networks-ICANN 2007 Lecture Notes in Computer Science 4668 (Porto, 9 Sep., 2007), Part I, 229-238.
- ⑪ S, Maeda. (2007). Convergence analysis of the EM algorithm and joint minimization of free energy. In IEEE Machine Learning for Signal Processing Workshop (27 Aug., 2007), 318-323.
- ⑫ S, Osaga. (2007). A probabilistic model of MOSAIC. In Proceedings of the IEEE Symposium on Foundations of Computational Intelligence (Hawaii, 1 Apr., 2007), 41-46.
- ⑬ M, Ihara. (2007). Estimation of source-filter model via acoustical feature extraction by GA-like algorithm. In the 12th International Symposium on Artificial Life and Robotics (Oita, 26 Jan., 2007), GS12-4.
- ⑭ T, Date. (2007). On-line variational PCA for adaptive visual tracking. In the 12th International Symposium on Artificial Life and Robotics (Oita, 26 Jan., 2007), 458-461.
- ⑮ S, Osaga. (2007). A probabilistic modeling of MOSAIC learning. In the 12th International Symposium on Artificial Life and Robotics (Oita, 25 Jan., 2007), 12, 167-171.
- ⑯ K, Makino. (2007). Adaptive control of a looper-like robot based on the CPG-actor-critic method. In the 12th International Symposium on Artificial Life and Robotics (Oita, 25 Jan., 2007), 12, 129-132.
- ⑰ T, Ueno. (2006). Fast and Stable Learning of Quasi-Passive Dynamic Walking by an Unstable Biped Robot based on Off-Policy Natural Actor-Critic. In Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (Beijing, 13 Oct., 2006), 5226-5231.
- ⑱ H, Fujita. (2006). Feature extraction for decision-theoretic planning in partially observable environments. In Artificial Neural Networks- ICANN2006 Lecture Notes in Computer Science 4131 (Athens, 14 Sep., 2006), Part I, 820-829.

[図書] (計2件)

- ① Doya, K., & Ishii, S. (2007). A probability primer. In Bayesian Brain (eds., K. Doya., S. Ishii., A. Pouget., and R. P. N. Rao), 323, MIT Press, 2007.
- ② 石井 信. (2007). 統計数理は隠された未来をあらわにする, 分担執筆 (2章, pp. 33-55), 東京電機大学出版.

[産業財産権]

○取得状況 (計1件)

産業財産権の名称 : 適応型制御器、適応型制御方法および適応型制御プログラム

発明者 : 吉本 潤一郎 石井 信

権利者 : 独立行政法人科学技術振興機構
国立大学法人奈良先端科学技術大学院大学

産業財産権の種類・番号 :

特許 第3960286号

取得年月日 : 平成19年5月25日

国内・外国の別 : 国内

[その他]

■ 論理生命学 (石井) 研究室 HP

<http://hawaii.sys.i.kyoto-u.ac.jp/home>

■ 石井信 HP

<http://hawaii.naist.jp/~ishii/index-ja.html>

6. 研究組織

(1) 研究代表者

石井 信 (ISHII SHIN)

京都大学・大学院情報学研究科・教授

研究者番号 : 90294280

(2) 研究分担者

柴田智広 (SHIBATA TOMOHIRO)

奈良先端科学技術大学院大学・情報科学研究科・准教授

研究者番号 : 40359873

吉田和子 (YOSHIDA WAKO)

University College London・Researcher

研究者番号 : 30379599

中村泰 (NAKAMURA YUTAKA)

大阪大学・大学院工学研究科・助教

研究者番号 : 70403334

森健 (MORI TAKESHI)

京都大学・大学院情報学研究科・研究員

研究者番号 : 00457144

※2008年度より研究協力者