

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 5 月 21 日現在

機関番号：14301

研究種目：基盤研究(B)

研究期間：2009 年度～2011 年度

課題番号：21300113

研究課題名（和文） あいまいで変動する環境におけるモジュール型意思決定モデルの研究

研究課題名（英文） A study of modular models of decision making in uncertain and non-stationary environments

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（和文）：

あいまいで変動し得るような複雑環境において意思決定を行うことのできる統計的学習モデルを強化学習に注目しながら開発を行った。オンライン的にモジュール追加を行いながら効率よく価値関数を学習する方法を開発し、セミパラメトリック統計に基づく価値関数の最適な学習法を導出した。応用として、ノンホロノミック制御モデルの自律制御学習に成功した。階層性を有する推論課題において前頭前野のモジュール構造によりあいまいさを解消することが分かった。またあいまいな環境での推論に関して MRI 信号からのデコードに成功した。

研究成果の概要（英文）：

We have developed statistical learning models, with a particular interest in reinforcement learning (RL), which can perform decision making in uncertain and even non-stationary environments. We have derived an RL method in which value function represented by a module structure can be online and efficiently approximated by adding new modules in an incremental fashion, and an optimal learning procedure of the value function based on the framework of semi-parametric statistics. As an application, we have succeeded in automatic control of non-holonomic systems by means of a policy-based RL method. In the human brain, we have found module-like structures which are activated when inferring a hierarchical inference task. Moreover, we have succeeded in decoding inference process based on the subject's behaviors and MRI scanned images.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	6,200,000	1,860,000	8,060,000
2010 年度	5,600,000	1,680,000	7,280,000
2011 年度	2,100,000	630,000	2,730,000
年度			
年度			
総計	13,900,000	4,170,000	18,070,000

研究分野：総合領域

科研費の分科・細目：情報学・

キーワード：強化学習・モジュールアーキテクチャ・計算論的神経科学・ロボット・非侵襲脳計測

1. 研究開始当初の背景

動物の報酬依存の意思（行動）決定過程は、

ふるくは条件付け学習モデル、近年は強化学習モデルにより説明されてきた。工学的立場からの強化学習法は、価値ベース法、方策ベース法、およびシステム同定法に分けられるが、いずれも統計学習アルゴリズムとしての性質が不明確なままであった。不確実環境においては直接観測できない変数（不観測変数）の推定は重要であり、環境のモデル同定に基づき推定を行うモデル同定強化学習は、上記のシステム同定法とは異なるアイデアであるが、これまでに価値ベース法および方策ベース法の拡張として研究が行われてきた。申請者らはこのモデル同定強化学習において、ベイズ推定によりモデル同定を行う手法を提案していた(Ishii, Yoshimoto, Yoshida, 2002)。また、脳においては、モデルの同定とそれに基づく状態推定が前頭前皮質の異なる部位で行われているとの仮説を提案しており、非侵襲脳活動計測を用いた認知科学実験により仮説を支持する結果を示してきた(Yoshida, Ishii, 2006)。一方で、しかし、モデル同定強化学習は、状態推定が収益最大化から切り離されているため、一般に不観測変数の次元が大きくなるにつれて適用が困難となる。高次元問題でも適用可能なアルゴリズムの実装、および、そうしたアルゴリズムが脳内で実現されているかの検証は重要な残課題であった。不確実で変動する環境に直面している脳の意思決定には、一度覚えた環境ダイナミクスを忘れることなく、それらを状況に応じて切り替えながら状態推定を行うモジュール型システム同定法が用いられていると考えるのが自然であろう。運動学習の分野ではモジュール型の制御学習アルゴリズムが提案されており、例えば小脳におけるマルチモジュール運動学習のアーキテクチャとして MOSAIC(Wolpert, Kawato, 1998)が提案されている。

本研究は、こうした背景に基づき、あいまいで変動する環境において、効率よく意思決定を行うモジュール型意思決定モデルとそのため学習アルゴリズムを、特に強化学習に注目して開発を行い、多自由度ロボットの制御などの工学的応用可能性を示しながら、類似のアルゴリズムが脳内に存在しえるのかを、行動実験と非侵襲脳活動計測実験により評価することを目標とした。

2. 研究の目的

あいまいで変動する環境において、効率よく意思決定を行うモジュール型意思決定モデルとそのため学習アルゴリズムを、特に強

化学習とベイズ法などの統計的推定に着目しながら、開発する。ヒト上肢を模した多関節ロボットの運動制御の自律獲得課題に適用することで工学的検証を行う。複数のモダリティや時間スケールを混在させた意思決定タスクを用いたヒト行動・認知実験により神経科学的検証を行う。これらを統合した融合研究を実施することで、ヒトと機械における、高次元意思決定過程の類似性と相違性とを解明する。

3. 研究の方法

モジュール型意思決定アルゴリズムの開発と評価 (前田新一、石井信)

統計的学習法に基づくマルチモジュール型意思決定アルゴリズムを強化学習に注目しながら導出する。マルチモジュール型強化学習法を、複数モジュールによるアーキテクチャの統計学習として定式化し、オンライン学習によってアルゴリズム化する。特にサンプル数が増えるにしたがって、モジュールを追加することで環境適応を行うことができる手法を開発する。開発された手法の高次元の制御問題への応用可能性を評価する。価値関数ベースの強化学習法について、統計学習の立場からの理論構築を進める。また、不確実性のある環境においても適合することができる意思決定アルゴリズムである方策勾配学習法を、非ホロノミック制御問題に適用することで工学的応用可能性を示す。

意思決定モデルの脳内実装可能性の評価 (石井信)

環境における不観測変数が多次元であり、複雑数の階層を成している場合に、その推論を行うタスクを被験者に課し、行動実験および核磁気共鳴図(MRI)を用いた非侵襲脳活動計測実験により、脳内神経基盤を探る。この際に、被験者の脳内推論状況をMRI信号と行動データから統計モデルを用いてデコードすることを試みる。

多関節ロボットの制御 (中村泰、石井信)

本研究に用いる多関節ロボットは、ヒトの上肢を模したもので、背骨、鎖骨、肩胛骨、上腕骨、尺骨、橈骨と手から構成された骨格が大胸筋、上腕二頭筋など26本の代表的な筋肉により駆動される。構造の複雑さにより、状態をセンサによって直接計測することは困難であるため、システム同定と状態推定を包含した統計的手法により制御を行う。また、脳信号からロボットハンドを制御するための基礎技術を開発する。そのために、(1)脳活動から少数のマーカ位置を再構成する手法、(2)少数のマーカ位置から多自由度の手指形状を再構成する手法、および、(3)手指

形状を模することができるロボットハンドの設計を進める。

4. 研究成果

モジュール型意思決定アルゴリズムの開発と評価:

あいまいで変動する環境において、効率よく意思決定を行うモジュール型意思決定モデルとそのための学習アルゴリズムを、特に強化学習と統計的学習に注目しながら、開発を進めた。また、強化学習の統計的学習による理論化を進めた。主たる成果は以下の通りである。

マルチモジュール型強化学習法、特にサンプル数が増えるにしたがってオンライン的に基底モジュールを追加することで価値関数を学習し、それにより変動する環境において適応的意思決定を行うことができる手法を開発し、従来法よりも学習効率が良いことを示した(Mori & Ishii, *IEEE Transactions on Systems, Man, and Cybernetics*, 2012)。状況によって探索と搾取を切り替えることで、あいまいな環境においても意思決定を行うことができるマルチモジュール型システム同定強化学習法を策定し、ヒト被験者の行動実験において90%以上の行動を再現できることを示した(Adomi, et al., 2010)。価値ベースの強化学習法のセミパラメトリック統計に基づく一般化を行い、バイアスがない線形近似により価値関数が表現できる場合について最適(最小分散)な学習法を導出した(Ueno, et al., *Journal of Machine Learning Research*, 2011)。さらに価値関数近似にバイアスがある場合について、既存の学習法の漸近分散の理論的解明を行った(Ueno, et al., *Neural Networks*, to appear)。

また、不確実性のある環境においても適合可能な意思決定アルゴリズムである方策ベース学習法(GPOMDP)を非ホロノミック制御問題である猫型ロボットの空中ひねり運動制御課題に適用し、制御則が自律的に獲得可能であることを示した(Nakano, Maeda, & Ishii, submitted)。

意思決定モデルの脳内実装可能性の評価:

モジュール型意思決定モデルの脳内実装の可能性を評価するため、ヒト行動実験および核磁気共鳴図(MRI)などを用いたヒト非侵襲脳活動計測実験を実施した。後者に関しては、従来の相関解析に基づく検討のみならず、脳内表現のデコードを用いることで定量的評価を進めた。主たる成果は以下の通りである。階層性のある推論課題(階層型並べ替え課題)を被験者に課し、行動実験および非侵襲脳活動計測実験を実施した。階層ベイズ型モデ

ルは被験者行動を良く説明でき、また、上位階層および下位階層の推論は前頭前野の異なる脳領域に位置するモジュールにおいて処理されている可能性があることが示された

(Yoshida, et al., *NeuroImage*, 2010)。

環境における不観測変数が多次元となる場合であってその推論を行うタスク(不完全迷路探索課題)を被験者に課し、行動実験および非侵襲脳活動計測実験を実施した。探索と搾取を切り替えるモデルは被験者行動を良く説明できることが分かったので、これに基づき、探索・搾取切り替えに関わる脳内情報処理基盤を論じた(Shikauchi, et al., 2010)。またこの際に、行動決定のための推論過程が、MRI信号からデコード可能であること、具体的には、推論の各次元に対し、被験者により最高で75%程度のデコード性能があることが明らかになった。この成果は、現在、実験数を増やしつつ論文としてまとめているところである。

多関節ロボットの制御:

強化学習などの統計的学習を、ヒト上肢を模した多関節ロボットなどの多自由度ロボットに適用することで、工学応用を図った。また、脳信号から多自由度のロボットハンドを動かす手法の開発を進めた。主たる成果は以下の通りである。

ヒト上肢を模した多関節ロボットについて、自由度間の同期をアトラクタとして表現し、それを選択することで制御する手法を開発した(Sugahara, et al., *Journal of Robotics and Mechatronics*, 2010)。人の教示に基づく見まね学習を用いた制御法を開発(Nishioka, et al., 2009)し、また、過去に蓄積したデータからの評価に基づく制御法を開発した(Okadome, et al., 2012)、高自由度制御対象であってもシナジーを用いて制御可能であることを示した。また、脳信号からロボットハンドを制御することを目的として、まず、脳活動から少数のマーカ位置を再構成する手法をベイズ的正準相関解析法に基づき開発した。さらに、少数のマーカ位置から多自由度の手指形状を再構成する手法を線形回帰法に基づき開発した。これらを組み合わせることで、脳活動から多自由度の手指形状を再構成した。マーカ位置によっては相関係数 0.65 程度となる再構成が可能であることが分かった。

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 10 件)

- ① T. Ueno, S. Maeda, S. Ishii: Asymptotic analysis of value prediction in well-specified

- and misspecified models. *Neural Networks*, (to appear)
- ② M. Ishihara, S. Maeda, K. Ikeda, S. Ishii: Low-dimensional feature representation for instrumental identification. *SICE Journal of Control, Measurement, and System Integration*, (to appear).
 - ③ T. Ueno, S. Maeda, M. Kawanabe, S. Ishii: Generalized TD Learning. *Journal of Machine Learning Research*, **12**(6), 1977-2020, (2011).
 - ④ T. Mori, S. Ishii: Robust reinforcement learning in sequential value function approximation. *IEEE Transactions on Systems, Man and Cybernetics*, **41**(5), 1407-1416, (2011).
 - ⑤ J. Yoshimoto, M. Sato, S. Ishii: Bayesian normalized gaussian network and hierarchical model selection method. *Intelligent Automation and Soft Computing*, **17**(1), 71-94, (2011).
 - ⑥ A. Sugahara, Y. Nakamura, I. Fukuyori, Y. Matsumoto, H. Ishiguro: Generating circular motion of a human-like robotic arm using attractor selection model. *Journal of Robotics and Mechatronics*, **22**(3), 315-321, (2010).
 - ⑦ S. Hotta, S. Oba, S. Ishii: Visual attention model involving feature-based inhibition of return. *Journal of Artificial Life and Robotics*, **15**(2), 129-132, (2010).
 - ⑧ W. Yoshida, H. Funakoshi, S. Ishii: Hierarchical rule switching in prefrontal cortex. *NeuroImage*, **50**(1), 314-322, (2010).
 - ⑨ K. Shibata, N. Yamagishi, S. Ishii, M. Kawato: Boosting perceptual learning by fake feedback. *Vision Research*, **49**(21), 2574-2585, (2009).
 - ⑩ N. Yano, T. Shibata, S. Ishii: Adaptive particle allocation for multifocal visual attention based on particle filtering. *Journal of Artificial Life and Robotics*, **13**, 522-535, (2009).

[学会発表] (計 13 件)

- ① S. Ishii: Machine learning methods for brain machine interface. *Statistics for Biomedical and Social Mathematical Sciences*, Tokyo, (2012).
- ② 石井 信: ネットワーク型ブレインマシンインターフェースに向けて. *Neuroscience 2011*, Yokohama, (2011).
- ③ 石井 信: ブレインリーディングからブレインマシンインターフェースへ. 平成 23 年度数学・数理科学と諸科学・産業との連携研究ワークショップ「ネットワー

ク型知識に関する機械学習的アプローチ」, Tokyo, (2012).

- ④ Y. Okadome, Y. Nakamura, H. Ishiguro: A control method for a redundant robot using stored instances. *International Symposium on Artificial Life and Robotics*, Beppu, (2012).
- ⑤ 石井 信: ネットワーク社会のブレインマシンインターフェース. (社) 電子情報通信学会 通信ソサイエティ総会, Sapporo, (2011).
- ⑥ J. Hirayama, A. Hyvärinen, S. Ishii: Sparse and low-rank estimation of time-varying Markov networks with alternating direction method of multipliers. *International Conference on Neural Information Processing*, Sydney, (2010).
- ⑦ M. Adomi, Y. Shikauchi, S. Ishii: Hidden Markov model for human decision process in a partially observable environment. *International Conference on Artificial Neural Networks*, Thessaloniki, (2010).
- ⑧ Y. Shikauchi, M. Adomi, S. Ishii: Separation of exploration and exploitation in maze navigation task. *Neuro2010*, Kobe, (2010).
- ⑨ S. Hotta, S. Oba, S. Ishii: Visual attention model involving feature-based inhibition of return. *International Symposium on Artificial Life and Robotics*, Beppu, (2009).
- ⑩ S. Nishioka, S. Maeda, S. Ishii: Machine learning approach to 9-DOF arm robot control. *International Symposium on Artificial Life and Robotics*, Beppu, (2009).
- ⑪ T. Mori, S. Ishii: Robust approximation in decomposed reinforcement learning. *International Conference on Neural Information Processing*, Bangkok, (2009).
- ⑫ T. Ueno, M. Kawanabe, S. Maeda, S. Ishii: Optimal online learning procedures for model-free policy evaluation. *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, Ljubljana, (2009).
- ⑬ T. Mori, S. Ishii: An additive reinforcement learning. *International Conference on Artificial Neural Networks*, Limassol, (2009).

[図書] (計 2 件)

- ① 石井 信: 科学, **80**(12), 分担執筆 (page 1188). 岩波書店, (2010)
- ② 石井 信: 価値と学習・よくわかる認知科学 分担執筆 (I V-7 節). ミネルヴァ書房, (2010).

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 2 件)

名称：推定装置

発明者：石井 信, 柴田智広, 坂東誉司, 深谷直樹, 清水幹郎

権利者：(株) デンソー, 奈良先端科学技術大学院大学

種類：US

番号：7,813,544 B2

取得年月日：2010年10月12日

国内外の別：外国

名称：三次元形状復元装置

発明者：柴田智広, 石井 信, 坂東誉司, 深谷直樹, 清水幹郎

権利者：(株) デンソー, 奈良先端科学技術大学院大学

種類：US

番号：8,107,735 B2

取得年月日：2012年1月31日

国内外の別：外国

[その他]

ホームページ等

■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.sys.i.kyoto-u.ac.jp/~ishii/>

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

中村 泰 (NAKAMURA YUTAKA)

大阪大学・基礎工学研究科・助教

研究者番号：70403334

前田 新一 (MAEDA SHINICHI)

京都大学・大学院情報学研究科・助教

研究者番号：20379530

(3)連携研究者

森 健 (MORI TAKESHI) 平成 21 年度

大阪大学・基礎工学研究科・研究員

(現・エジンバラ大学 研究員)

大塩 立華 (OSHIO RITZ) 平成 21 年度

京都大学・大学院情報学研究科・研究員

(現・自然科学研究機構 生理学研究所 研究員)

員)

鹿内 友美 (SHIKAUCHI YUMI) 平成 22 年度

京都大学・大学院情報学研究科・技術補佐員

森本 智志 (MORIMOTO SATOSHI) 平成 23 年度

京都大学・大学院情報学研究科・技術補佐員