

「脳を創る」

平成11年度採択研究代表者

銅谷 賢治

(国際電気通信基礎技術研究所 ATR脳情報研究所 計算神経生物学研究室長)

「行動系のメタ学習と情動コミュニケーション機構の解明」

1. 研究実施の概要

本研究の目標は、人間や動物の柔軟な学習を可能にしている、脳の「メタ学習」の機構を理論的、実験的手法により解明し、その理解をより人間的なロボットやコンピュータの設計論に結びつけて行くことである。そのため、

- 1) 学習系のメタパラメタ調節の理論と神経修飾物質系の機構
- 2) 課題や環境に適した内部表現の獲得における大脳皮質の機能
- 3) メタ学習機構としてのコミュニケーションと進化の役割

という3つの主要テーマを掲げ、計算理論／生理実験／脳活動計測／ロボット実験の連携によりその解明を進めた。

まず理論面では、環境の変化をベイズ推定の原理により検出し、それをもとに学習速度などのメタパラメタを自動調整するアルゴリズムを導出し、シミュレーションによる検証を進めた。また、変分ベイズ法によるMEGセンサ信号からの電流源推定法に関して、視覚刺激のシミュレーションデータにより評価と改善が進み論文投稿を行った。Sequential Monte Carlo法により人間やサルの行動データからの学習系の内部変数を推定する方式を特許出願し、NIPS2003にて発表した。

生理実験では、まず上乳頭体核による海馬の記憶制御機構に関して、c-Fosタンパクの発現実験により、新奇環境では上乳頭体核の活動が上昇すること、上乳頭体核の破壊により海馬歯状回の活動が低下することを明らかにした。また、サルの前頭前野ニューロンの活動記録により、報酬獲得時に、その直前の行動に応じて活動を変えるニューロンの存在について、Cerebral Cortexに論文を出版した。

脳活動計測実験では、異なる時間スケールでの報酬予測に関わる脳部位を比較する実験を進め、線条体および前頭葉で、予測の時間スケールに応じたトポグラフィックなマップの存在をNIPS2003にて発表した。さらに、セロトニンの前駆物質であるトリプトファンの欠乏と過剰投与下での報酬予測による行動学習実験を行い、その行動データとfMRIデータを解析中である。

ロボット実験では、ネズミ型ロボット群「サイバーロデント」のハードウェアとシミュレーターの整備を進めた。そのもとで、行動選択のランダムさを進化アルゴリズムにより

バッテリーレベルに応じて変化させる実験，報酬関数自体を進化により獲得させる実験を進めた。

今後プロジェクト終了に向けて，これまで開発したメタ学習アルゴリズムの統合，神経修飾物質系機能仮設の再評価と拡張を行い，またサイバーロデントを変化する環境もとで動的に自己保存する系として自立させることをめざす。

2. 研究実施内容

1) システム統合グループ

a) メタ学習と神経修飾物質系モデル

広島大グループとの共同により，セロトニンの前駆物質であるトリプトファンの欠乏および過剰摂取下での，長期と短期の報酬予測に基づく行動学習実験を行った。行動実験およびfMRIデータの解析を行い，論文投稿に向け準備中である。また，マルコフ決定課題の学習時の脳活動計測データの強化学習モデルによる解析で，前頭葉腹内側部，島皮質，線条体で，それぞれ腹側部はより短期の，背側部はより長期の報酬予測に関与することを明らかにし，NIPS2003で発表し，変分ベイズ法によるMEGセンサ信号からの電流源推定法に関して，視覚刺激のシミュレーションデータにより評価と改善が進み論文投稿を行った。Sequential Monte Carlo法により人間やサル の行動データからの学習系の内部変数を推定する方式を特許出願し，NIPS2003にて発表した。

b) 目的指向的内部表現の獲得

追跡眼球運動の脳活動計測により，指標の消失時には一次視覚野の活動は低下するのに対し，MST野とFEF野では活動が上昇することが明らかになり，両者が視標運動の予測に関与することを示した。また，重点サンプリング法により異なる自由度を持つ並列学習回路を複数の学習させる方式について特許出願した。

c) ネズミ型ロボット群「サイバーロデント」の学習と進化実験

ハードウェアとシミュレーターの整備を進め，一度充電すると一定時間不応期を持つバッテリー，複数ロボットの相互作用のシミュレーションが可能になった。行動選択のランダムさをバッテリーレベルに応じて変化させる実験を行い，進化アルゴリズムにより，満充電と電池切れの近くではよりgreedyな行動を取るU字形の制御カーブが得られることを確認した。さらに，報酬関数自体の起源を探るため，電池の視覚センサーで発見することに対する補助報酬と，電池を実際に獲得することに対する報酬とのバランスを，進化アルゴリズムにより最適化させる実験を進めた。

2) 学習理論グループ

a) 強化学習の並列モデルメタ学習による動的環境への適応

強化学習の目標は，エージェントが各状態において最適な行動を選択し累積報酬和を最大化することであり，通常は価値関数(割引報酬和)の期待値を推定することにより行われる。しかし，現実には期待値のみに頼る学習では必ずしも安定した学習性能は保証されず，期待値に加え分散の値も推定し，これをメタ学習の評価基準に利用し

て学習速度 α を逐次選択することで、動的環境下における汎化性能の向上させる方式を導出した。

b) 強化学習における進化型メタ学習

単一のメタパラメータを設定して学習を行う代わりに、複数の値を用いて並列的に学習と最適化を行う進化型モデルを提案し、メタパラメータの自動調整を試みた。

3) 適応システムグループ

a) 部分観測マルコフ決定過程に関する数理モデルの脳内実現可能性を探るために核磁気共鳴図を用いた認知実験を行った。その結果、強化学習タスクと腹外側前頭前野、帯状回との関連が示唆された。この結果は、現在論文投稿中である。また、不観測状態の推定に関して前部前頭前野との関連が示唆されたので、論文執筆中である。

b) 連続系の自動制御のために、方策勾配法によるactor-critic法を2足歩行学習に適用し、自動制御に成功した。この成果は国際会議AAAIに採択された。さらに、自然方策勾配法に拡張した手法は、現在論文投稿中である。

c) ノルアドレナリンが環境変化を検出することで、適切な環境適応を可能とする数理モデルを、ベイズ推定の枠組みで導出した。この成果は、国際会議BICSに受理され、また論文投稿中である。

4) 情動記憶グループ

a) 上乳頭体神経活動による海馬ゲート機能制御に関する研究

探索行動時に海馬体と同期した脳波活動を示す視床下部・上乳頭体核が、海馬機能をどのように制御しているかを明らかにする目的で、オープンフィールドテストにおけるラットの海馬および上乳頭体神経活動を、c Fosタンパク発現を指標として免疫組織化学法により解析した。まず、オープンフィールドに置かれたラットの上乳頭体核では、c Fos陽性細胞数は著しく増加し、上乳頭体核の神経活動が新規環境下で上昇することが確認された。このとき上乳頭体核内のc Fos陽性細胞の分布密度を比べたところ、内側部ほうが外側部に分布密度が高いことが明らかとなった。この結果は、新規環境探索時に、上乳頭体核は脳幹網様体からの入力を受けて活性化されることを示唆している。次に、上乳頭体核と海馬神経活動の関係を明らかにするために、イボテン酸を局所注入して上乳頭体核を破壊したラットをオープンフィールド内に置き、海馬におけるc Fos陽性細胞数を計測した。上乳頭体破壊ラットでは、海馬歯状回のc Fos陽性細胞数は、擬似手術を受けたラット海馬歯状回で観察されたc Fos陽性細胞数に比べて有意に少なかった。これらの結果は、上乳頭体核神経活動が海馬神経活動を促進していることを示しており、新規環境の情報取り込み時に海馬歯状回のゲートを開くとともに海馬CA2/CA3a領域の神経活動を高めて、海馬内への情報入力と情報の海馬内伝播を促進していることが示唆された。

b) 遺伝子改変マウスの作成

アデノシンA1受容体過剰発現マウスの行動観察実験を開始した。現在、4ラインのうち2ラインで、十字高架式迷路、T字迷路、水迷路を用いた行動観察を行った。その

結果、学習行動には変化がないが、恐怖反応が増強している可能性が示唆された。しかし、おなじ週令の動物の例数が少ないため、統計学的な検定はできていない。また、神経細胞特異的アクチン結合タンパクドレブリンAの選択的ノックアウトが完成した。現在実験に使用する個体を繁殖中である。

5) 学習修飾グループ

前頭連合野はルールの切り替えに応じて行動を変化させるなど、柔軟な行動方略の変化や学習に関わる領野であり、その機能にはドーパミンなどの神経修飾物質が深く関与していると考えられている。

平成15年度は、報酬期待などの情動的要素や、行動の結果、報酬が得られたかどうか、行動方略の変化や学習にどのような影響を与えるかに焦点を当て、行動課題遂行中のサルの前頭前皮質から、ニューロン活動の記録を行なった。その結果1) 報酬獲得時に見られる神経活動が、直前の行動とその種類に依存して活動を変えることを見出した。これは、学習に必要な行動の評価に関わるのではないかと考えられる。2) 次に、視空間性の記憶を変換し行動を導く課題を用い、視空間情報を再現するニューロン活動が報酬期待の強い影響を受けることを見出した。このことから、運動準備よりも、視覚刺激保持に関わるワーキングメモリがより強く報酬期待の影響を受けるものと考えられる。3) さらに、方略の変換に関わる神経過程の解明を目指し、報酬獲得のためのルールが課題遂行中に強制的に交替する行動課題遂行中のサルのニューロン活動を記録した。その結果、報酬獲得時に見られる神経活動が、以前に使用した、あるいは次に使用するルールに応じて変化することを見出した。これは、方略の変換に関わる神経過程の一つと考えられる。

16年度は、研究計画の最終年度であり、これまでの研究成果を取りまとめ、国際誌への発表に重点を置く。さらに、サルで得られたこれまでの成果を通してfMRI等を用いた人を被験体とした研究へと発展させてゆきたい。

6) 精神薬理グループ

学習、情動などの高次脳機能に関与する神経修飾物質、特にセロトニンに焦点をあて実験的研究を行った。これにより、セロトニン系の学習と行動における機能仮説の検証を行った。

平成15年度は、健常者およびうつ病患者を対象に将来の報酬予測に基づく意思決定課題やGo-NoGo課題遂行中のfMRIを測定し、将来の報酬予測および反応の抑制に関する脳機能を検討した。健常者では、将来の報酬予測が右前頭前野(右腹外側前頭前野・右背外側前頭前野)・帯状回前部・右頭頂葉・視床・基底核(淡蒼球)・左小脳を含むネットワークと関連していることが示された。また、右前頭前野・右頭頂葉が反応の抑制に関連していることが示された。一方で、うつ病患者では、これらの機能に関連した脳領域の活動が低下していることを明らかにした。うつ病はセロトニンの活性低下と関連していると考えられており、これらの機能とセロトニンとの関与を示唆する結果と言える。

また、セロトニンの行動決定における役割を理解するため、セロトニンの前駆物質であるトリプトファン摂取量を調節して脳内セロトニン活性を変化させ、それによる行動と脳

活動の変化との関連を調べた。トリプトファン欠乏ではすべての被験者で有意に血漿トリプトファン濃度が低下し、トリプトファン負荷ではすべての被験者で有意に血漿トリプトファン濃度が上昇しており、これらの変化に連動して中枢神経系セロトニン機能も変化していると考えられた。心理尺度のProfile of Mood States (POMS) の変化を調べたところ、トリプトファン欠乏時は負荷時と比べてPOMSの下位項目のうち倦怠感の得点が有意に高かった。部分的な結果ではあるが、中枢神経系セロトニン変化が気分に影響を与えていると考えられた。

3. 研究実施体制

システム統合グループ

- ① 研究分担グループ長名：銅谷賢治 国際電気通信基礎技術研究所 室長
- ② 研究項目：神経修飾物質系のダイナミクスモデルとロボット実験

学習理論グループ

- ① 研究分担グループ長名：松本 隆 早稲田大学理工学部 教授
- ② 研究項目：メタ学習の理論と応用

適応システムグループ

- ① 研究分担グループ長名：石井 信 奈良先端科学技術大学院大学大学院情報科学研究科 教授
- ② 研究項目：強化学習系における探索と最適化の理論

情動記憶グループ

- ① 研究分担グループ長名：関野祐子 群馬大学大学院医学系研究科 助教授
- ② 研究項目：海馬の記憶と学習における情動性制御のメカニズム

学習修飾グループ

- ① 研究分担グループ長名：澤口俊之 北海道大学医学研究科 教授
- ② 研究項目：前頭連合野における報酬系におけるドーパミン受容体の役割

精神薬理グループ

- ① 研究分担グループ長名：山脇成人 広島大学大学院医歯薬学総合研究科精神神経医学 教授
- ② 研究項目：学習と行動におけるセロトニンの役割

4. 主な研究成果の発表（論文発表および特許出願）

(1) 論文発表

・システム統合グループ（銅谷賢治）

- Capi, G., Uchibe, E., Doya, K. (2003). Selection of neural architecture and the environment complexity. *Dynamic Systems Approach for Embodiment and Sociality from Ecological Psychology to Robotics*, 6, 311-317.
- Daniel M Wolpert, D. M., Doya, K., Kawato, M. (2003). A unifying computational framework for motor control and social interaction.. *Philosophical Transactions of the Royal Society*, 358, 593-602.
- Haruno, M., Kuroda, T., Doya, K. , Toyama, K. , Kimura, M. , Samejima, K. , Imamizu, H. , Kawato, M. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. *Journal of Neuroscience*, 24(7), 1660-1665.
- Samejima, K., Doya, K., Kawato, M. (2003). Inter-module credit assignment in modular reinforcement learning. *Neural Networks*, 16(7), 985-994.
- Schweighofer, N., Doya, K., Kuroda, S. (2004). Cerebellar aminergic neuromodulation: towards a functional understanding. *Brain research reviews*, 44, 103-116.
- Schweighofer, N., Doya, K., Fukai, H. , Chiron, Jean V., Furukawa, T., Kawato, M. (2004). Chaos may enhance information transmission in the inferior olive. *Proceedings of the National Academy of Sciences, USA*, 101(13), 4655-4660.
- 内部英治, 銅谷賢治 (2004). 複数報酬のもとでの階層強化学習. *日本ロボット学会誌*, 22(1), 120-129.
- 杉本徳和, 鮫島和行, 銅谷賢治, 川人光男 (2004). 複数の状態予測と報酬予測モデルによる強化学習と行動目標の推定. *電子情報通信学会論文誌*, J87-D-II (2), 683-694.
- 佐藤雅昭, 吉岡琢, 梶原茂樹, 外山敬介 (2003). MEG電源の階層ベイズ推定. 第18回日本生体磁気学会大会, 16(1), 148-149.
- 中村泰, 佐藤雅昭, 石井信 (2004). 神経振動子ネットワークを用いたリズム運動に対する強化学習法. *電子情報通信学会論文誌*, J87-D-II (3), 893-902.

・学習理論グループ（松本隆）

- 杉淳二郎, 栗原貴之, 松本隆 (2003). ダイナミカルシステムの階層ベイズ的最小

埋め込み次元推定. 情報処理学会論文誌, 44(12), 3098-3111.

- 中田洋平, 栗原貴之, 用水邦明, 和田健作, 松本隆 (2003). ベイズ的非線形ダイナミカルシステムの再構成と予測: Hybrid Monte Carlo. 電子情報通信学会論文誌, J86-D-II (12), 1143-1155.

・適応システムグループ (石井信)

- Oba, S., Sato, M., Ishii, S. (2003). Variational Bayes method for mixture of principal component analyzers. *Systems and Computers in Japan*, 34(11), 55-56.
- 吉本潤一郎, 石井信, 佐藤雅昭 (2003). 連続力学システムの自動制御のためのオンラインEM強化学習法. システム制御情報学会論文誌, 16(5), 209-217.
- 吉本潤一郎, 石井信, 佐藤雅昭 (2003). 変分法的ベイズ推定法に基づく正規化ガウス関数ネットワークと階層的モデル選択法. 計測自動制御学会論文集, 39(5), 503-512.
- 前田新一, 石井信 (2004). 学習によるproduct codeの設計. 電子情報通信学会論文誌, J87-A(3), 382-390.

・情動記憶グループ (関野祐子)

- Takahashi, H., Sekino, Y., Tanaka, S., Mizui, T., Kishi, S., Shirao, T. (2003). Drebrin-dependent actin clustering in dendritic filopodia governs synaptic targeting of postsynaptic density-95 and dendritic spine morphogenesis. *Journal of Neuroscience*, 23(16), 6586-6595.
- Beraudi, A., Traversa, U., Villani, L., Sekino, Y., Nagy, J.I., Poli, A. (2003). Distribution and expression of A1 adenosine receptors, adenosine deaminase and adenosine deaminase-binding protein (CD26) in goldfish brain. *Neurochemistry International*, 42(6), 455-464.

・学習修飾グループ (澤口俊之)

- Tsujimoto, S., Sawaguchi, T. (2004). Properties of delay-period neuronal activity in the primate prefrontal cortex during memory- and sensory-guided saccade tasks. *European Journal of Neuroscience*, 19, 447-457.
- Tsujimoto, S., Sawaguchi, T. (2004). Neuronal representation of response-outcome in the primate prefrontal cortex. *Cerebral Cortex*, 14, 47-55.

・精神薬理グループ（山脇成人）

- Shirao, N., Okamoto, Y., Okada, G., Okamoto, Yu., Yamawaki, S. (2003).
Temporomesial activation in young females associated with unpleasant words
concerning body image. *Neuropsychobiology*, 48, 136-142.

(2) 特許出願

H15年度特許出願件数：2件（研究期間累積件数：7件）