

「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、強化学習のベイズ学習理論の立場からの定式化と、行動探索のランダムさ、学習速度などのメタパラメタの自動調整アルゴリズムの導出、その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、Sequential Monte Carlo法により、MEGセンサ信号からの電流源推定や、人間やサルの行動データからの学習系の内部変数の推定の新方式を開発した。

生理実験では、上乳頭体核から海馬への入力による記憶の制御機構の解明に向け、ラットの行動実験、マウスの分子生物学実験を進めた。また、サルの前頭前野ニューロンの活動記録により行動と報酬を関係づける応答がみられること、またドーパミンが各カラムごとの活動を制御することを明らかにした。

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明らかにした。さらに、隠れ状態の推定を必要とする行動学習課題の脳活動計測を開始した。

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」の開発を進め、「エサ」となるバッテリーパックの捕獲と収集、協調行動、「遺伝子」となるプログラムの赤外通信によるコピーなどの機能を実現した。シミュレータ上での進化アルゴリズムにより、強化学習の回路構造の選択とメタパラメタ適応の実験を行い、その有用性をハードウェアでの学習実験で確認した。

これらの研究活動に加え，Neural Networks誌の特集号“Computational Models of Neuromodulation”を，チーム研究代表者がゲストエディターとして企画し，20本の論文を集め2002年8月に出版した．今後さらに，メタ学習理論の開発，セロトニンなど神経修飾物質系と大脳基底核，海馬，前頭前野などの回路の学習における役割の解明を進め，それらをロボット実験で目に見える形で実証することをめざす．

## 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

メタパラメタの最適化を行う階層強化学習モデルを定式化し，その脳内での実現の可能性を示した．また，異なる時間スケールでの報酬予測機構に関する脳活動計測実験を行い，強化学習モデルを用いた解析により，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た．また，変分ベイズ法による脳磁気計(MEG)からの脳内電流源推定，Sequential Monte Carlo法による行動データからの学習系のパラメタ推定のアルゴリズムを開発し，シミュレーションによりその有効性を確認した．

#### b) 目的指向的内部表現の獲得

追跡眼球運動の行動心理実験により，ヒトは視標の運動を等速直線運動に限らず予測できることを明らかにし，さらに脳活動計測により，MST野が視標運動の予測に関与することを示した．また，階層並列的な学習系において，報酬とダイナミクス予測によるモジュール選択方式，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を定式化し，非線形制御課題でその有効性を確認した．

#### 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

ハードウェアと基本プログラムの整備が進み，「エサ」となるバッテリーパックの探索，捕獲と巣への回収，追跡行動，「遺伝子」となるプログラムの赤外通信によるコピーなどが実現した．また，シミュレータの整備が進み，その上でエサ捕獲行動の学習回路の構造選択とメタパラメタ適応の進化実験に成功した．さらにその結果をハードウェア実験にフィードバックし，シミュレーターと実機を融合した実験の有効性を確認した．

### 2) 学習理論グループ

強化学習におけるメタパラメタの自動調整アルゴリズムの開発に向け以下の取り組みを行った．

a) Sequential Monte Carlo法の枠組みから，学習データが逐次的に与えられる場合，適応的に学習を遂行する新たなアルゴリズムを提案し，非線形ダイナミカルシステムの再構成と予測の数値実験を行いその有効性を確認した．

b) 過去の行動価値関数の標本分散に注目し，それを極力小さく押さえるメタパラメタの調整方法を提案し，初期的数値実験を行った．

### 3) 適応システムグループ

#### a) 部分観測マルコフ決定過程の数理モデルと自由度の制御

オンライン型のベイズ推定法に基づく線形システム同定法を開発し、連続部分観測システムのための強化学習アルゴリズムを導出した。本手法はベイズ推定に基づくため、系の自由度の変化にオンライン的に追従することができる。本手法を用いてヒトの運動の自由度を評価した

#### b) 部分観測マルコフ決定過程の脳内学習モデル

部分観測マルコフ決定過程の数理モデルの脳内実現可能性を探るために、fMRI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。

#### c) 動的環境に適応するノルアドレナリン系モデル

ノルアドレナリン系が環境変化を検出し温度パラメタを自動制御することで、適切な環境適応を可能にする数理モデルを、線形システム同定とベイズ推定の枠組で導出した。

### 4) 情動記憶グループ

#### a) 上乳頭体核-海馬系における情動行動と記憶の制御機構

学習成立初期過程において、海馬と上乳頭体核との神経活動相関の脳波解析を試みた。探索行動時には両部位で同期したシーター波が記録され、両部位を含み広範囲にわたってc-Fosタンパクの発現が認められた。

#### b) アデノシン受容体とドレブリンによる記憶形成の制御機構

アデノシンA1受容体過剰発現マウスを作製し、そのうち4ラインのマウス海馬で、受容体タンパクの過剰発現が認められた。まだドレブリン過剰発現マウスのうち、発現量の少ないラインでは長期増強の度合いは対照群と差がなかった。

#### c) 皮質-視床間発振神経活動の役割

膜電位変動の光学的測定法で、皮質-視床の3 Hzオシレーションの可視化に成功し、GABA受容体活動により300-400 ms持続する過分極応答の検出に成功した。

### 5) 学習修飾グループ

前頭連合野はルールの切り替えに柔軟に対応するなどの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。今年度は、報酬期待や期待の誤差などの情動的要素が、行動方略の変化や学習にどのような影響を与えるかを明らかにする事を目的として、サル前頭前皮質において、慢性単一ニューロン記録法等を用いた実験を行ない、以下の結果を得た。

a) 報酬獲得時に見られる神経活動が直前の行動に依存して活動を変える。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。

b) 感覚運動変換過程における感覚情報保持に関わるニューロン活動は、報酬期待に強い影響を受ける。このことにより、運動準備過程よりも、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。

c) 光学測定法を用い、前頭前皮質スライスにおけるコラム活動が、ドーパミンD1受容体作動薬によって、活動の幅を変えることなく増強される。このことから、ドー

パミンのコラムに対する影響は、その機能を変えることなく、活動の増強に寄与するのではないかと考えられる。

#### 6) 精神薬理グループ

将来の報酬予測など、セロトニンが重要な関与をしていると考えられる脳機能が脳のどの領域で行われているかについて検討するとともに、セロトニンの神経伝達が低下していると考えられるうつ病患者での課題遂行時の脳活動測定を進行中である。

- a) 健常者においてセロトニンが関与していると考えられる脳機能の賦活課題を作成しfMRIおよびMEGを用いて検討しこれまでに以下の結果が得られた。1) 将来の報酬予測および意思決定には帯状回前部、小脳、基底核および右前頭前野の活動が重要である。2) 衝動性の抑制には右前頭前野の活動が重要である。3) 快刺激の予期には左前頭前野の活動が関連し、不快刺激の予期と右前頭前野の活動が関連している。さらに、不快刺激の予期には扁桃体および帯状回前部が重要な役割を果たしている。4) 不快な単語の認知には、左扁桃体とその周辺領域および右海馬傍回が関連している。
- b) セロトニン神経系の異常が示唆されているうつ病患者の認知課題遂行中の脳活動をfMRIおよびMEGを用いて測定し健常者との比較検討を行いこれまでに以下の結果が得られた。1) うつ病患者では健常者と比較して、左前頭前野および帯状回前部の反応性が低下している。2) うつ病患者では健常者と比較して急性ストレス負荷後のMMNの反応が低下している。

### 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 主幹研究員
- 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電気電子工学科 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 群馬大学医学部行動分析学部門 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 広島大学医学部神経精神医学講座 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（原著論文）発表

###### ・システム統合グループ（銅谷賢治）

- Doya, K., Samejima, K., Katagiri, K., Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Computation*, 14(6), 1347-1369.
- Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, 15(4-6), 495-506.
- Doya, K., Dayan, P., Hasselmo M. E. (2002). Computational models of neuromodulation. *Neural Networks*, 15(4-6), 475-477.
- 銅谷賢治 (2002). 強化学習の計算論. *医学のあゆみ*, 202(3), 175-179.
- 銅谷賢治 (2002). 行動学習系のメタパラメタ制御と神経修飾物質. 別冊・数理科学「脳情報数理科学の発展」, 31-36.

###### ・適応システムグループ（石井信）

- Ishii, S., Yoshida, W., Yoshimoto, J. (2002). Control of exploitation-exploration meta-parameter in reinforcement learning. *Neural Networks*, 15(4-6), 665-687.
- 石井信 (2002). 強化学習と2足歩行学習. 別冊・数理科学, 118-124.
- 大羽成征, 石井信, 佐藤雅昭 (2002). 変分法的ベイズ推定による混合主成分分析. *電子情報通信学会論文誌*, J85-D-II(6), 1055-1065.

###### ・精神薬理グループ（山脇成人）

- Okada, G., Okamoto, Y., Morinobu, S., Yamawaki, S., Yokota, N. (2003)
- Attenuated Left Prefrontal Activation during a Verbal Fluency Task in
- Patients with Depression. *Neuropsychobiology*, 47(1), 21-26.
- 山脇成人, 岡本泰昌 (2002). 強化学習と精神医学. *医学のあゆみ*, 202(3), 33-36.

##### (2) 特許出願

H14年度特許出願件数：1件（研究期間累積件数：5件）