

## 主論文の要約

論文題目     **Reinforcement Learning in Humans: Rethinking Psychologically Plausible Model-Free and Model-Based Algorithms**  
(ヒトにおける強化学習：心理学的に妥当なモデルフリー・モデルベースアルゴリズムの再考)

所        属     環境学研究科    社会環境学専攻

氏        名     遠山    朝子

### 要約本文

本論文は、強化学習モデルで用いられるモデルフリーシステムとモデルベースシステムのそれぞれのアルゴリズムについて、ヒトの情報処理容量の制約を考慮した新たなアルゴリズムを検討し、それらを統合した計算論モデルを提案するものである。

第1章では、モデルフリー・モデルベースシステムの定義、及び、関連する先行研究についてまとめた。モデルフリーシステムは、経験によってのみ価値を更新する学習システムである。一方で、モデルベースシステムは、環境の内的モデルを利用して価値の計算を行う。強化学習の文脈において、“モデル”とは、ある行動をとった後にどの状態に遷移するか、という遷移構造についての内的モデルを指す。これら2つのシステムのバランスは、発達・認知・精神障害等を特徴づけるものとして注目されている。2つのシステムが選択に及ぼす影響を分離する課題として、Daw, Gershman, Seymour, Dayan, and Dolan (2011) が提案した2段階意思決定課題と、彼らが提案した計算論モデルを紹介する。このモデルの特徴は、モデルフリーな価値とモデルベースな価値を並列に更新する点にあり、この論文の中では、並列モデル (parallel model) と呼ばれる。ただし、このモデルの問題点は、計算負荷が高いことである。特にモデルベースの価値の計算には、すべての遷移情報と次のステップの価値が必要となる。認知資源に制約があるヒトや動物が、実際にこのような計算をしているのか、という点はこれまで検討されてこなかった。

そこで、第2章では、まず、モデルベースシステムのアルゴリズムについて、より計算負荷の小さいアルゴリズムを提案し、2段階意思決定課題から得られたデータを用いて検討を行った。この新しいアルゴリズムを用いたモデルは、適格度調整モデル (eligibility-adjustment model) と呼ばれるものであり、モデルフリーシ

システムで用いられる適格度トレースを利用した更新則の更新率を、遷移確率で重み付けするという仮定を導入している。モデル比較の結果、従来用いられている並列モデルに比べてデータへのフィットが改善することが認められた。

続く第3章では、モデルフリーシステムのアルゴリズムについて検討を行った。従来のモデルフリーシステムでは、選択した選択肢の価値のみ更新し、選択されなかった選択肢の価値は、それまでの価値が維持されるという仮定をおいている。しかしこれは、記憶の衰退を考慮すれば無理のある仮定である。そこで、本章では、選択されないものの価値が忘却されていくという過程 (forgetting process) を従来のモデルフリーシステムのアルゴリズムに導入した。また、選択されない選択肢の価値は一定のデフォルト価値 (default value) に回帰していく、という仮定も追加し、より柔軟に忘却過程を扱えるアルゴリズムを提案した。モデル比較の結果、忘却過程を導入したモデルフリーシステムの有用性が支持された。また、選択データの解析からも、選ばれないことによる価値の変化を仮定しなければ説明できない特徴が確認された。

第4章では、第2、3章で検討されたアルゴリズムのハイブリッドモデルを構築し、詳細な検討を行った。ハイブリッドモデルで推定されるモデルベースシステムの比重は、実験参加者が口頭報告した課題遂行中の遷移確率の利用の有無と最もよく対応していた。このことから、提案されたハイブリッドモデルがシステム間のバランスを評価するモデルとして優れていることが示された。

これまでの検討で、忘却過程を導入したモデルが2段階意思決定課題の選択行動の予測に大きく貢献することが示されたが、従来の研究では、忘却過程を入れないモデルが使用されてきた。そこで、第5章では、実際は忘却過程が存在するのに、忘却過程を仮定しない計算論モデルを適用した場合に、パラメータ推定に歪みが生じる可能性を検討した。シミュレーションの結果、モデルベースシステムの比重を評価するパラメータが、真の忘却過程のパラメータとして設定された値に依存して、システムティックに歪められることが確認された。

第6章では、決定的な遷移構造 (遷移確率 100%) の2段階意思決定課題を用いて、これまでの章で提案されたアルゴリズムの有用性を確認した。また、この課題の決定的な遷移構造を反映し、連続する行動の価値をまとめて更新する、という節約的な更新則が用いられている可能性が示された。さらに、最もデータへの当てはまりの良かったモデルを用いて精神障害の傾向との関係を検討したところ、従来用いられてきた並列モデルで報告されている結果とは異なる結論が得られることが示された。

第7章では、新しく提案されたアルゴリズムについて総括し、総合的な考察を行うと共に、今後の研究の展望について議論した。