

審査の結果の要旨

論文提出者氏名 伊藤 秀昭

ヒトの脳に学んで高性能な機械を実現するために、脳のアлゴリズムを研究し真似ることは有望である。特に推論やプランニング等といった脳の持つ様々な高次機能は、工学的に見ても極めて重要な機能である。これらを包括的に実現しうる学習理論の枠組みとして、強化学習が存在する。強化学習は報酬最大化の理論であり、一方高次脳機能は生命にとって報酬をより多く得るために存在する可能性があるからである。

本論文は、“A reinforcement learning model of the basal ganglia system — towards realization of higher brain functions —”（和文題目「大脳基底核の強化学習モデル— 高次脳機能の実現に向けて—」）と題し、9章より成る。

第1章では、大脳基底核の強化学習モデルという研究の動機と目的を記述している。脳においては大脳新皮質—基底核系が強化学習に関与しているという仮説が提案され、研究されてきている。しかし実際には簡単な順序学習等で議論されておりまだ不明な部分が多い。そこで本論文ではサルの大脳基底核から得られた神経活動のデータと強化学習モデルとを詳細に対比する。

第2章では、背景として、大脳基底核についての解剖学的及び生理学的知見について説明している。大脳基底核は報酬あるいは行動制御に関連した処理が行われているという見方が受け入れられつつあるが、詳細については現在研究が進められている最中である。

第3章では、強化学習理論とそれに基づいて提案された既存の大脳基底核のモデルについて説明している。いくつかのモデルが既に提案されているが、それらの優劣や統一については未だ明らかではない。

第4章では実験タスクについて説明している。本論文ではタスクとして彦坂興秀教授（順天堂大学）の考案による1DR/ADRタスクを用いている。1DR/ADRタスクはmemory-guided saccade taskであるが、報酬条

件が制御されている。報酬条件以外は全く同じであるため、報酬条件の相違の影響を直接的に調べることができる。本論文ではこの実験での現象が強化学習モデルによってどのように説明できるかを考えている。また、将来的に多電極同時記録法を用いた場合の解析方法も提案している。

第5章では、基底核黒質緻密部のドーパミン (DA) ニューロンの発火活動と、強化学習モデルの1つである TD (temporal difference) モデルとの比較を行っている。DA ニューロンの発火活動については Schultz らによって報酬の予測誤差をコードしているという仮説が提案されており、TD モデルの TD 誤差に対応すると予想されていた (DA=TD 仮説)。ただし Redgrave らのようにこれを疑問視する意見も存在した。そこで 1DR/ADR タスクにおいてこの予想がどの程度正しいかを検討した。特に報酬条件が変化した際の DA ニューロン活動の変化、および非報酬トライアルの連続による変化についてシミュレーションおよび理論的解析によって検討し、DA ニューロンの発火活動特性と TD 誤差特性がほぼ一致し、さらにサルが報酬履歴の情報を不確かさを含む形で保持しているとする、よく再現できることを示した。これらの結果により DA=TD 仮説の妥当性を支持した。

第6章では、基底核の尾状核 (caudate; CD) ニューロンの発火活動と、強化学習モデルの1つである actor/critic モデルとの比較を行っている。CD ニューロンについては、Houk らにより、actor/critic モデルによって説明できるという仮説 (CD=actor/critic 仮説) が提案されていた。ただし実際の発火活動との比較はあまりなされていなかった。そこで 1DR/ADR タスクにおいてできるだけシンプルな actor/critic モデルにより CD ニューロンの発火活動がどの程度説明できるかを検討した。pre-cue, post-cue, saccade という各タスク期間における CD ニューロンの活動パターンおよびその変化について検討し、シミュレーションおよび理論的解析によりそれぞれの現象が actor/critic モデルのニューロンによってよく再現されることを示した。これによって actor/critic モデルの妥当性を支持した。さらに、cue 刺激提示後は報酬条件についての情報が脳基底核系から消失していることを示唆する結果も得た。

第7章では報酬に関連した大脳基底核系の処理の特徴についてより詳しく調べている。特に1DR/ADRタスクでのエラーサッケードの性質、眼球運動のパラメータとCDニューロンの発火頻度との相関、CDニューロンの発火活動に含まれる情報、およびそれらの時間変化を調べ、CDでは行動を制御するために報酬情報とcue情報が処理されており、それらがcue提示の前後で前者から後者へと変化することを示唆した。この変化はcue提示という外的イベントにより引き起こされたものと考えられるが、もし内的に（外的イベントに依存せず）同様のことが起こりうるとすれば、CDにおいてゴール指向性推論が行われる可能性を示している。

第8章では、これらの結果を踏まえ、actor/criticモデルに基づいたゴール指向性推論モデルを提案している。まず、サブゴールの設定をactorの行動と考え、actor/criticモデルに組み入れた。さらに”悪い行動の拒否”と”二重学習”という二つの改良を行なった。これらにより、actor/criticモデルによりゴール指向性推論が実現され、学習性能も向上しうることを示した。また、このような学習により階層的な行動制御が自動的に獲得され、さらに学習を重ねるとそれが消滅することを示した。

第9章では、各章の結果を総合して結論を次のようにまとめた。本論文は、大脳基底核のドーパミンニューロンと尾状核ニューロンの神経活動について、それぞれTDモデルとactor/criticモデルにより基本的にはよく説明できることを示した。さらに、大脳基底核系における報酬関連処理の特徴を明らかにしゴール指向性推論という高次脳機能への関与の可能性を示した。最後にこれらに基づきゴール指向性推論モデルを提案した。

以上を要するに、本論文は大脳基底核における神経活動を、実験データ解析および数理モデル解析により調べてその特徴を明らかにするとともに、新たなモデルを提案したものである。これは生体情報工学、そして数理工学上貢献するところが大きい。

よって本論文は博士（工学）の学位請求論文として合格と認められる。