

# 論文の内容の要旨

論文題目      **Effects of Information Trace to Cerebral Learning (情報トレースの脳の学習への効果)**

氏名            齋藤 大

## 1. 序論

脳のような複雑な情報処理機械内で行われている処理を理解するためには、計算、アルゴリズム、ハードウェアといったマクロからミクロなレベルでの解析が必要であると考えられている。そのため、ヒトや動物の行動といったマクロな挙動、神経細胞の活動といったミクロな挙動が、行動心理実験、神経イメージング実験、電気生理実験といった様々なスケールの手法で観測されてきた。そして、ミクロな挙動がどのようにマクロな挙動に影響を与えるかを解明し、異なったスケールで得られた実験結果の間をつなぐことが脳の数理モデル研究の一つの大きな役割である。

我々の脳は運動の制御、記憶の形成と想起、将来状態の予測、意思決定、言語の使用などといった多様な機能を兼ね備えている。それらの機能を実現するためには、学習が必要不可欠なものであり、そのメカニズムの解明は脳科学研究の主要な目的の一つである。多くの行動心理実験によって、ヒトや動物は外部の環境に応じて様々な行動や反応を獲得することができることが示されてきた [Mazur, 2002]。そして、学習は脳神経細胞間の情報伝達を担っているシナプスの結合荷重の変化によって実現されると考えられている [Dayan & Abbott, 2001]。言い換えれば、外界の情報がシナプ스에保存されているということである。しかしながら、脳内にはシナプス以外にも様々な形で、様々なタイムスケールで情報が保持されていることが知られている。それらの情報は短いものは数ミリ秒、長いものでは数ヶ月間保持され [Abraham, 2003]、シナプスの学習に多大な影響を与えていると考えられるが、その詳しい影響は未だ分かっていない。

本論文では、学習課題の試行にまたがって保持される情報の痕跡（トレース）が学習にどのような影響を与えるかを、数理モデルの解析を通して解明する。研究の対象として、報酬の時空間的信頼度割当問題 [Sutton, 1984], 摂餌課題におけるマッチング行動 [Herrnstein, 1961], 状態遷移確率 [Sutton & Barto, 1998] の学習を取り扱い、その中でトレースがどのような影響や機能的意味を持っているかを論じる。

## 2. 時空間的信頼度割当問題を解決するニューラルネットワークモデル

オペラント条件付け [Skinner, 1938] や強化学習 [Sutton, 1998] など報酬によってフィードバックが与えられる学習において、報酬は脳の学習において時空間的な不確実性を持っている。すなわち、報酬はそれを生み出す原因となるヒトや動物の反応や行動から遅れて与えられるため、その反応や行動、ひいては脳の神経活動の状態と遅延した報酬の関連に不確実性がある。学習のためには、報酬の獲得に寄与した過去の反応を同定する必要がある、時間的不確実性は一種の信頼度割当問題を引き起こす。遅延報酬問題として古くから知られるこの問題 [Hull, 1943] は、強化学習の分野における時間的信頼度割当問題の解決策である適格度トレース [Klopf, 1972] によって解消可能であると提案されている [Izhikevich, 2007]。報酬の空間的不確実性とは、報酬の獲得に貢献した脳部位がどこかという不確実性であり、空間的信頼度割当問題を引き起こす。効率的な学習のためには空間的信頼度割当問題も解決される必要がある。これら時空間的な不確実性が脳の学習に相互にどのような影響を与えるかは未だ分かっていない。

我々は、報酬のようなフィードバック信号の時空間的信頼度割当問題を解決するニューラルネットワークモデルを構築し、時空間的な不確実性が学習にどのような影響を与えるのかを調べた。我々のモデルは、単層の線形ニューラルネットワークを並列したものであり、同形の教師ネットワークの結合荷重を学習する。ただし、報酬は各々のネットワークの出力誤差の和というスカラ値であるため、個々のネットワークがどの程度の出力誤差を発生させたかを同定する必要がある。また、報酬は遅延して与えられる。我々は、時間的な不確実性に対して適格度トレースを導入し、空間的な不確実性に対してノードパータベーション学習 [Fiete et al., 2007] という、各ネットワークの出力に独立なノイズを印加する学習手法を用いた。統計力学的解析と数値シミュレーションによって、時空間的な不確実性は相互作用をもって学習に影響を与えること、適格度トレースが時空間的な不確実性を部分的に解消することを示した。

## 3. マッチング則を説明するベイズ意思決定モデル

報酬を与えるスケジュールの違いが、ヒトや動物の学習とその結果として生起する行動の違いをもたらすことが行動心理実験によって知られている [Mazur, 2004]。草食動物の摂餌環境を模した変動時間間隔スケジュールを用いた課題においては、ヒトや動物はマッチング行動を示すことが知られている [Herrnstein, 1961]。マッチング行動とは、ある選択肢を選んだ割合がその選択肢によって得た報酬の割合と合致するという、マッチング則に従った行動である。後年のより詳しい解析の結果、ヒトや動物はマッチング則よりやや一様な選択を行う傾向を示すことが分かり、これはアンダーマッチングと呼ばれている [Baum, 1974]。近年、マッチング行動における選択は直近の報酬履歴に強く依存

していることが判明し、その依存性は指数関数よりややロングテールな二重指数や双曲線関数でよく回帰されることが判明した [Sugrue et al., 2004; Lau & Glimcher, 2005; Corrado et al., 2005]. これは、ヒトや動物の学習が、報酬のトレースに依存していることを示唆する結果である。マッチング則を説明する様々な数理モデルが提案されてきたが [Loewenstein & Seung, 2006; Soltani & Wang, 2006; Simen & Cohen, 2009; Sakai & Fukai, 2008; Katahira et al., 2010], アンダーマッチングと二重指数型の報酬履歴依存性を生み出す計算原理については未だ分かっていない。

我々はいくつかの単純なベイズ意思決定モデルを構築し、どのような計算原理がアンダーマッチングや報酬履歴依存性を生み出すかを解明した。数値シミュレーションを用いて、各々のモデルの反応特性を実験結果と照らし合わせた結果、環境（報酬の確率）の不確実性を考慮したモデルにおいてのみ、アンダーマッチングと二重指数型の報酬履歴依存性が示された。我々は、環境の不確実性を前提とした計算がアンダーマッチングや二重指数型の報酬履歴依存性を生み出す要因であると提案する。また、この結果から、報酬のトレースが環境の不確実性を前提とした計算を実現するアルゴリズムである可能性が示唆された。

#### 4. 遷移確率を学習するニューラルネットワークモデル

環境の状態は時々刻々と変化するものであり、その環境の将来状態を経験に基づいて予測することは、ヒトや動物の持っている重要で基本的な能力の一つである。脳内には将来状態を予測する情報が符号化されていることが多数の電気生理実験によって示されており [Schultz et al., 1997; Eskandar & Assad, 1999], また、近年の脳機能イメージング研究によって、将来状態の予測に必要な不可欠だと考えられる、状態間の遷移確率も符号化されていることが判明した [Glascher et al., 2010]. 脳内で行われている将来予測の神経基盤を解明することはその情報処理を理解する上で非常に重要なことであり、遷移確率を活用した神経回路モデルがいくつか提案されている [Rao, 2004; Wacongne et al., 2012]. しかしながら、予測に必要な不可欠な遷移確率そのものを学習する神経基盤については未だ分かっていない。

我々は、ヘブ則 [Hebb, 1949] と活動依存的な荷重減衰 [Kempster et al., 1999; Abbott & Nelson, 2000] という脳内での存在が示されている神経基盤に基づいた学習則を提案した。解析計算と数値シミュレーションによって、ニューラルネットワークモデルが遷移確率を学習し、将来の状態を予測できることを示した。さらに、このモデルはランダムドット刺激の方向弁別課題において、サル of 心理測定関数と外側頭頂間野ニューロン活動を再現することを確認した。また、状態が変化する場合だけではなく、環境そのものが変化した場合における学習についても解析を行った。環境が変化した直後、将来状態の予測誤差は大きく上昇するが、状態の適格度トレースを導入することで誤差の上昇が低減されることが判明した。また、適格度トレースは、学習速度をほとんど悪化させずに学習が定常状態に達した後の残留誤差を減少させることがわかった。

#### 5. 結論

この学位論文では、いくつかの脳の学習モデルを解析計算と数値シミュレーションによって調べ、試行をまたがって保持される情報のトレースがどのように脳の学習に影響を与えるかについて解明した。

第一に、報酬に時空間的な不確実性がある場合に、入力とノイズのトレースを用いたニューラルネットワークモデルを解析した。その結果、時空間的信頼度割当問題はトレースによって解消可能であることが示された。第二に、ヒトや動物のマッチング行動は、環境の不確実性を考慮した計算によって生み出されていることを提案した。近年の解析により、ヒトや動物のマッチング行動中の選択は、報酬のトレースに二重指数的に依存していることが判明している。すなわち、環境の不確実性を考慮した計算が、報酬のトレースを用いたアルゴリズムによって実現されている可能性を示唆した。第三に、状態の遷移確率を学習し、将来状態を予測する生物学的に妥当なニューラルネットワークモデルの学習則を提案した。入力のトレースを導入することで、学習速度を悪化させることなく、予測誤差を減少させることが可能であることが示された。さらに、環境そのものが変化する場合に、予測誤差の悪化を軽減することも判明した。これらの結果から、脳の学習において試行をまたがって保持される情報のトレースは、報酬の時空間的な不確実性、環境の不確実性に対して学習性能を向上させる機能を持つことが示された。