

# 強化学習エージェントの学習精度と環境変動への適応に関する理論的検証： モデルベース学習とモデルフリー学習の比較

氏名：渡辺舜

指導教員：竹澤正哲

ヒトをはじめとする動物は、様々な課題や環境に応じて、柔軟かつ適応的に行動を選択する能力を持っている。この能力はモデルベース学習 (MB 学習) とモデルフリー学習 (MF 学習) と呼ばれる、2 種類の強化学習アルゴリズムを使うことによって実現されている。MB 学習は、学習者が直面している課題や環境の構造を表現したモデル (e.g., 環境構造を模した樹状図や地図など) を生成して学習を行うが、MF 学習はそのようなモデルを生成せず、試行錯誤を繰り返すことによって学習を行う。そのため、MB 学習はより高い精度で学習することが可能で (Daw, 2018; Keramati et al., 2011)、環境の変化に対しても迅速に適応することができるが (Christopoulos et al., 2017)、モデルを生成するために高い認知コストが必要であり (Christopoulos et al., 2017)、行動選択の際にはモデルを参照した熟慮を行うことから、行動を決定するまでに長い時間を要するとされる (Keramati et al., 2011)。一方、MF 学習は学習精度や環境の変化に対する適応という点では MB 学習に劣るが、モデルを必要としないため、低い認知コストで学習が可能であり、素早く行動を決定することができる (Keramati et al., 2011)。先行研究では、MB 学習と MF 学習は異なる方法で学習を行うために、それぞれが異なる利点と欠点を持っており、この 2 種類の学習を状況に応じて使い分けたり、2 つを組み合わせたアルゴリズムを用いることで、動物は適応的な行動を獲得するという理論のもと、動物が 2 種類の学習をどのように使い分け、あるいは組み合わせているのか、意思決定課題を用いた実験やコンピュータ・シミュレーションによって研究が行われてきた (Daw et al., 2005; Daw et al., 2011; Kool et al., 2016)。これらの先行研究は、MB 学習と MF 学習の間に、明確な違いが存在するという前提のもとに行われてきたが、実はそのような違いが本当に存在するのか、理論的な検証はほとんど行われていない。例えば、先行研究で用いられた意思決定課題は、MB 学習と MF 学習には明確な差があるという前提のもと、どちらの学習が使用されたかを学習者の行動から弁別するためのものである。しかし、この課題を MB 学習と MF 学習

のいずれかを用いて学習した場合に、両者の課題成績に差が見られるかは明らかにされていない。また、先行研究では実験やシミュレーションを行う際に、やや非現実的な前提が置かれていることもある。例えば Keramati et al. (2011) では、MB 学習は常に課題の環境構造と、その変化を完全に認識することができるという前提や、課題の環境が変化したとき、学習者にその変化に関する情報が即座に与えられ、MB 学習だけがその情報を用いることができる、といった前提が置かれている。しかし、現実世界では環境が変化したとき、実際に変化した環境を探索し、試行錯誤することで情報を獲得し、どのような変化が起きたかを推測しなければならないという状況も多いため、より現実に即した前提を置いたうえで、検証を行う必要がある。そこで本研究は、課題における成績という観点から、定量的な比較を行った際に、MB 学習と MF 学習の間に本当に差が見られるのか、また、課題の状況設定をより現実に即したものに変更した場合に、MB 学習と MF 学習の差がどのように変化するかという 2 つの点について、理論的な検証を行うことを目的とした。まず、MB 学習は MF 学習に比べ、より正確な学習が可能であり、環境変動に対しても迅速に適応できるとする主張について、先行研究で用いられた課題を用いたコンピュータ・シミュレーションによって検討した。次に、先行研究におけるやや非現実的と考えられる課題設定を、より現実に即したものに修正した上で、MB 学習と MF 学習の学習精度、環境変動への適応差がどのような形で見られるかについても検討した。

シミュレーションの結果、本研究で使用した意思決定課題においては、学習精度の高さ、環境変動に対する適応度のいずれに関しても、MB 学習と MF 学習の間に小さな成績差は見られたものの、その差を効果量の観点から見ると、ほとんど効果がないという結果が得られた。この結果は、MB 学習と MF 学習の間には、先行研究で想定されていたほどの明確な差は見られないことを示唆している。MB 学習と MF 学習の間には明確な違いが存在するという事は、多くの研究者にとって所与の前提となっていた。しかし、本研究の結果が示唆するように、もし 2 種類の学習の間に明確な違いが存在しないとすると、動物がいつどの場面で、いずれの学習アルゴリズムを使用すると予測されるのか、また動物が MB 学習と MF 学習の両方を用いる必要性はどこにあるのか、理論的な再考が必要となる可能性がある。