好奇心のモデルにおけるアルゴリズム水準と実装水準の接続
Connecting Algorithmic and Implementation Levels in Curiosity Models

長島 一真[†]，森田 純哉[†]，竹内 勇剛[†]
Kazuma Nagashima, Junya Morita & Yugo Takeuchi
[†]1 静岡大学
Shizuoka University
nagashima.kazuma.16@shizuoka.ac.jp, j-morita@inf.shizuoka.ac.jp, takeuchi@inf.shizuoka.ac.jp

概要

Marr によれば個々の認知モデルは複数の水準からなる階層に位置づけられる．同一の対象に対して，複数の階層のモデルを統合することで，対象の総合的な理解が導かれる．そこで本研究では，好奇心を対象とした複数の階層のモデルを比較検討する研究を行った．下層の実装水準のモデルとして深層強化学習，中層のアルゴリズム水準のモデルとして ACT-R モデルを選択した．その結果，これらのモデル間で整合する特徴が現れた．

キーワード：認知モデル, ACT-R, 内発的動機づけ, 好奇心, 深層強化学習

1. 背景

人間の活動は，金銭などの外部から与えられる報酬と，意欲などの内的に生じる報酬によって駆動される．後者は内発的動機づけと呼ばれ，前者に対して，持続的で高い水準の学習に寄与するとされる．近年では，内発的動機づけの一種である好奇心に着目した機械学習の研究が注目されている [1]．また，人間の脳内で生じるプロセスを対象とした認知モデルの研究においても好奇心の数理的な原理が探らている．

こういった人間の認知機能に関わるモデルを整理する際に，Marr の 3 水準は有効である [8]．Marr は，モデルは数理によって表現される計算論水準（計算の目標は何か，なぜそれが適切か），認知アーキテクチャなどで表現されるアルゴリズム水準（計算理論をどのようにして実現するのか），そしてハードウェアなどの実装水準（アルゴリズム水準がどのようにして物理的に実現されるのか）に区別できると主張した．これらのモデルは，互いが互いの制約条件になるなど階層的[1]に関連している．しかし，各水準を表現するモデルの候補は広範囲にわたり，さらに各水準には独立した事項も含まれる，そのため，個々の認知モデルの研究において，階層間のつながりは必ずしも明確にされてこなかった．

こう言った問題意識に基づき，著者らはこれまで，好奇心の認知モデルにおいて，アルゴリズム水準 [14] と計算論水準を接続する研究 [15] を行ってきた．本研究では，その発展として，好奇心のモデルにおけるアルゴリズム水準と実装水準の接続を目指す．これらを接続することで物理的な信号を直接入力とする深層学習のモデルに透過性の高いアルゴリズム水準の説明を付与できると考えた．

2. 関連研究

本研究は，好奇心の認知モデルのアルゴリズム水準と実装水準の接続を目指すものである．この目的と関連した研究として，(1) 好奇心のモデルにおける計算論と実装水準の関係，(2) ACT-R の好奇心のモデルにおけるアルゴリズム水準と計算論水準間の接続を行った研究を紹介する．

2.1 好奇心のモデルにおける計算論と実装水準の関係

近年，好奇心を対象とした認知モデルにおいて，計算論水準と実装水準の接続が進展している．好奇心の計算論水準のモデルの代表例として，Friston によるフリーエナジー原理 [4] を挙げることができる．この理論は，生体の計算の目標を長期的な予測誤差（予測と観察の不一致）の減少と捉える，そして，好奇心のきっかけとなる驚きや興味,「楽しさ」などの感情が，予測誤差によって引き起こされ，学習を駆動するとされる，

好奇心の源泉となる「楽しさ」に関しては複数の研究者が，環境における新規なパターンの発見と関連することを述べている [3, 5, 6, 12]．このような考え

[1]計算論水準は上層，アルゴリズム水準は中層，実装水準は下層となっている．

方を，Schmidhuber[12] は，より形式的な数理表現に落とし込んだ．彼の理論において，パターンの発見はデータの中で反復される定型的なパターンを発見し，圧縮することと定義される．そして，データを圧縮すること，あるいは圧縮可能なデータを取得することを「楽しさ」の表現とした．さらに，そのような「楽しさ」の表現を，強化学習における環境からの報酬として定式化した．

こういった計算論的な好奇心の定式化は，近年，盛んに研究される深層強化学習エージェントに取り入れられている [2, 10, 11]．その中でも Pathak は，ICM (Intrinsic Curiosity Module [11]) という手法を提案した．ICM は，画面上のピクセル情報から得られるエージェントの次状態との予測誤差を好奇心とし，それをエージェントに内部報酬とし与える手法である．この研究において，Pathak は ICM と深層強化学習 [10] の一種である A3C (Asynchronous Actor Critic [9]) のモデルとの統合を行った．A3C は，Advantage（複数ステップ先まで動かして更新），Actor-Critic[13], Asynchronous（非同期）の特徴を持つ．

深層強化学習は，エージェントを物理空間に近似した環境に接地し，その環境の条件下のタスクにおいて高いパフォーマンスを発揮する．そのため，筆者らは深層強化学習モデルを実装水準のモデルと位置付け，計算論水準と実装水準との接続についてはすでに高水準でなされていると考える．しかし，深層強化学習におけるモデルの実装は end-to-end に行われるため，モデルの内部の処理（アルゴリズム）が明確とは言い難い．

2.2 ACT-R の好奇心のモデルにおけるアルゴリズム水準と計算論水準間の接続

前節の深層強化学習エージェントの研究では，計算論水準と実装水準のモデルが直接的に接続される．しかし，その間に位置するアルゴリズム水準のモデルとの接続が不足しているため，エージェントの内部処理に透明性があるとは言い難い．これらの研究に対し，認知アーキテクチャを用いることで透明性のあるアルゴリズム水準のモデルとの接続が実現できる．

筆者らは認知アーキテクチャの 1 つである ACT-R を用いてアルゴリズム水準の好奇心のモデルの研究を行ってきた [14]．この研究では，ACT-R のシンボリックプロセス[2]を用いて内発的動機づけの一種である好奇心 [7] の認知モデルを構築した．著者らのモデルにおけるアルゴリズムの表現は，2.1 節で述べた数理的な好奇心の説明から発展するものである．数理的には，好奇心は外界の認識と経験から得られる予測との差分によって生じる．この予測からの差分が驚き（好奇心）を生じさせ，そのうちの一部は,「楽しさ」などの感情的反応を引き起こす．そしてその「楽しさ」は新しいパターンを発見することと説明される [6, 12]．

新たなパターンの発見をアルゴリズムとして記述するために，著者らの研究では，ACT-R のパターンマッチングを利用した．研究において開発したモデルは，迷路の継続課題を行うものである．課題中にパターンマッチングが発生するとモデルは課題を継続する動機（報酬）を獲得する．さらに，ACT-R における手続き学習（コンパイル）は，経験を重ねることよることによるパターンマッチの機会の減少（ルールと知識の圧縮）を導く．著者らのモデルはこの学習のプロセスを含めることで，モデルの「楽しさ」を感じる頻度の減衰により，飽きのプロセスを表現した．

1 章で述べたように，Marr によれば，計算論水準はアルゴリズム水準に目標を与え，アルゴリズム水準は計算論水準にその実現方法を与える．上記の著者らの研究は，計算論水準における予測誤差の減少という目標を，パターンマッチングとコンパイルという具体的な手続きによって実現する．ただし，ACT-R によって構築されたアルゴリズム水準のモデルは，課題ごとに多数のルールを持つため表現が複雑なものとなる．そのため，アルゴリズム水準のモデルを抽象化し，計算論的な表現へ変換する仕組みを構築する必要がある．

これを達成するために，筆者らは，ACT-R によって構築されたモデルが出力する行動から，ベイジアンネットワークによる数理的な因果関係のモデルを再構築する手法を考案した [15]．この手法により，好奇心の数理モデルから発展した ACT-R の好奇心のモデルを，数理的な表現に変換することが可能になった．著者らの過去の研究において，ベイジアンネットワークに変換された数理表現と予測誤差に関する好奇心の数理モデル [4] との対応が検討されているわけではない．しかし，アルゴリズムによって生成される複雑な振る舞いを簡潔な表現に置き換えることで，モデルの数理解析が可能になり，従来の数理モデルとの対応を議論する足がかりが構築されたといえる．

[2]ACT-R は記号操作を行うシンボリックプロセスを持つ．本研究では，記号操作を行いモデルを説明する過程をアルゴリズムとみなした．

3. モデルとシミュレーション

前節で示した著者らの先行研究 [15] より，アルゴリズム水準のモデルと計算論水準のモデルを接続する足がかりが構築された．これに対し，好奇心のアルゴリズム水準のモデルに対応する実装水準のモデルに関する検討はなされていない．Marr による 3 水準間を接続する好奇心のモデルを構築するためには，アルゴリズム水準と実装水準の接続が必要である．そのために，本研究では，ACT-R の好奇心のモデル [14] と同様の課題を行う好奇心の深層強化学習モデル（以降，ICM モデル）を実装する．

3.1 ICM モデル

ICM モデルは Pathak の研究 [11] に従って構築した．このモデルによって得られる結果と，先行研究 [14] において得られている 2 つの ACT-R モデルの結果を比較した．本研究で ICM モデルと比較する ACT-R モデルは，深さ優先探索と事例ベース学習を備えた思考水準の高いモデル（以降，DFS+IBL と表記） と，ランダムに振る舞う思考水準の低いモデル （以降，Random と表記） である．

ICM モデルが遂行する課題は，先行研究 [14] と同様に 30 のグリッド状の迷路（3 段階のサイズ × 10 のバリエーション）の探索である．この迷路を，ICM モデルは Actor-Critic における方策 π に従い，スタートからゴールまでの経路を探索する[3]．探索を通して，

$$r_t = r_t^i + r_t^e \tag{1}$$

を最大化する方策が学習される．式より，モデルに付与される報酬は，内部報酬 (r_i) と外部報酬 (r_e) の合計となる．よって，モデルは，異なる 2 種類の報酬のバランスを取りながら環境を探索することになる．

本研究において，外部報酬は，

$$r_e = \begin{cases} -1 & \text{if 移動に失敗} \\ 0 & \text{if 移動に成功} \\ 10 & \text{if ゴールに到達} \end{cases} \tag{2}$$

と定義される．1 回の移動においてモデルは，東西南北のうちのひとつの方角を選択し，迷路の曲角から別の曲角へ状態を遷移させることを試みる．経路の繋がらない方角を選択した場合，移動の失敗とみなす．また，移動の結果，ゴールに到達した場合は成功した報酬を得た後に，次のラウンドへと移行する．

[3]割引率 $gamma = 0.99$

内部報酬は先行研究 [11] に従い，

$$r_t^i = \frac{\eta}{2} \left\| \hat{\phi}\left(s_{t+1}\right) - \phi\left(s_{t+1}\right) \right\|_2^2 \tag{3}$$

として決定される．

状態 s は，深層強化学習において，ピクセルデータとして定義される．本研究では，迷路の状況（プレイヤー，壁，道）をグレースケール画像 (42×42) に変換したものとした．η は，好奇心の強さとみなせる．本研究では，ACT-R モデル [14] と対応する結果を得るために，試験的にこのパラメータの値を 0.1 から 0.9 まで 5 段階に変化させた．

モデルは同じ迷路のラウンドを繰り返す．すなわちエージェントがゴールに達した後に，エージェントの位置はスタートに戻される．また，エージェントが 100 回の移動（ステップ）の後にもゴールに到達しなかった場合，ゴールの達成によらず，新たなラウンドへ移行する．

モデルによる迷路の探索は，

$$th < r_i \times 500 + egs \tag{4}$$

に適合した際に終了する．th は閾値，egs はノイズを表す．すなわち，設定された閾値に対して，内部報酬が下回った際に探索が終了する[4]．また，この条件によらず，すべてのラウンドにおける行動の最大数が 3600 ステップに達した際にモデルは課題を終了するものとした．

3.2 結果と考察

図 1 には，ICM と ACT-R の対応を検討するめに，先行研究 [14] の ACT-R モデルの結果の一部（左と中央）と 3.1 節に示した ICM モデルの結果を示している．

図の縦方向には，各モデルの振る舞いを示す指標が並べられている．課題継続数，課題のゴール達成率，マップを広範囲に探索した際に増加するエントロピー，探索を通して学習される知識の量（プロダクション生成数）が示される．このうち，プロダクション生成数については ACT-R においてのみ計測可能な指標となっている．いずれのグラフも横軸はモデルに設定される内部報酬の重み（好奇心の強さ）となっている．

ICM と 2 種類の ACT-R を比較すれば，ICM は DFS+IBL ではなく，Random と類似した振る舞いを

[4]ACT-R の内部報酬と ICM モデルでは内部報酬のスケールが異なるため暫定的に固定値 500 を乗算した．th は 5 とした

[5]n の大きさに依存する標準誤差ではなく標準偏差に基づく値を示すことで，データのばらつきの程度を示している．

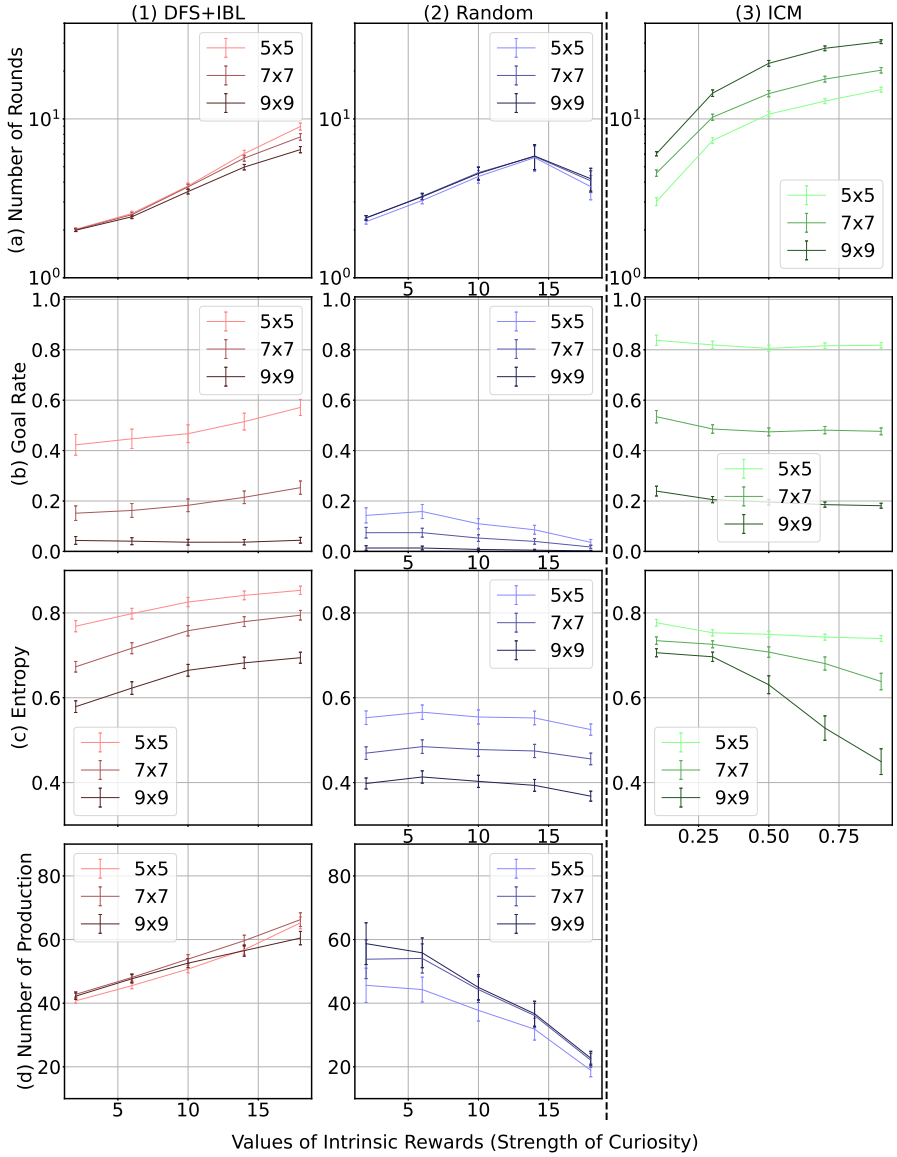

図 1　シミュレーション結果．それぞれのグラフのエラーバーは各マップにおいて得られた標準偏差 (ACT-R: $n = 1,000$, ICM: $n = 100$) の平均 ($n = 10$) に 1/10 を乗じた値を示す[5]．横軸は，好奇心の強度に基づく報酬を表す．

生成したことがわかる．いずれのモデルも内部報酬の重みが大きくなるに従って，課題継続数が増加する．しかし，ゴール達成率とエントロピーは，DFS+IBL が内部報報酬の重みに比例して増加させるのに対し，Random と ICM モデルは，内部報酬の重みが大きくなるに従い減少する．すなわち，ICM モデルは ACT-R モデルにおける浅い思考水準と類似した振る舞いを生成するモデルとみなすことができる．

4.　まとめと今後の展望

本報告では ACT-R のアルゴリズム水準のモデルと ICM を用いた実装水準のモデル間の接続を試みた．構築した ICM モデル対し，ACT-R モデルと同様の課題のシミュレーションを実施した結果，ICM モデルと思考水準の低い ACT-R モデルで類似した振る舞いが得られた．この結果から，本研究はアルゴリズム水準のモデルと実装水準のモデルの比較検討を行い，ICM モデルに関するアルゴリズム的な説明を付与したと考える．

しかし，モデル間の接続は不十分である。本報告において実装した ICM モデルには，シミュレーションに関わる内部パラメータを，ACT-R のシミュレーションにフィッティングできていないという問題がある．例えば図 1 の課題継続数は ACT-R モデルが多くても 10 回前後に対し，ICM モデルの課題継続数は遥

かに多い．この原因は，ICM のモデルの課題継続を判断する計算において，内部報酬の合計に試験的な値を乗算したためだと思われる．加えて 1 ラウンドのステップ数の上限を 100 と設定したが，値が適切であるかの検討が必要である．このような値を調整することで，ACT-R モデルと ICM モデルのさらなる検討が可能になると思われる．

文献

[1] Arthur Aubret, Laëtitia Matignon, and Salima Hassas. A survey on intrinsic motivation in reinforcement learning. *arXiv preprint arXiv:1908.06976*, 2019.

[2] Yuri Burda, Harri Edwards, Deepak Pathak, Amos Storkey, Trevor Darrell, and Alexei A Efros. Large-scale study of curiosity-driven learning. *arXiv preprint arXiv:1808.04355*, 2018.

[3] Roger Caillois. *Les Jeux et les Hommes: Le Masque et la Vertige*. Gallimard, Paris, 1958. （邦訳：多田道太郎，塚崎幹夫 訳：遊びと人間，講談社（1990））.

[4] Karl Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, Vol. 11, No. 2, pp. 127–138, 2010.

[5] Johan Huizinga. *Homo Ludens Versuch einer Bestimmung des Spielelementest der Kultur*. Pantheon, Amsterdam, 1939. （邦訳：高橋英夫 訳：ホモ・ルーデンス，中央公論新社（1973））.

[6] Raph Koster. *Theory of Fun for Game Design*. O'Reilly Media, Sebastopol, 2013. （邦訳：酒井皇治 訳：「おもしろい」のゲームデザイン―楽しいゲームを作る理論，オライリージャパン（2005））.

[7] Thomas W. Malone. Toward a theory of intrinsically motivating instruction. *Cognitive Science*, Vol. 5, No. 4, pp. 333–369, 1981.

[8] David Marr. Vision: A computational investigation into the human representation and processing of visual information. 1982.

[9] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pp. 1928–1937. PMLR, 2016.

[10] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, Vol. 518, No. 7540, pp. 529–533, 2015.

[11] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning*, pp. 2778–2787. PMLR, 2017.

[12] Jürgen Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, Vol. 2, No. 3, pp. 230–247, 2010.

[13] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, 2 1998. （邦訳：三上貞芳，皆川雅章 訳：強化学習，森北出版（2000））.

[14] 長島一真, 森田純哉, 竹内勇剛. ACT-R による内発的動機づけのモデル化. 人工知能学会論文誌, Vol. 36, No. 5, pp. AG21–E_1-13, 2021.

[15] 長島一真, 森田純哉, 竹内勇剛. 知的好奇心の計算論モデルとアルゴリズムモデルの接続：ベイジアンネットワークを用いた ACT-R モデルの分析. No. OS05-4, pp. 825–831. 日本認知科学会第 38 回大会, 2021.