

好奇心の認知モデルを用いた特性推定 Trait Estimation Using a Cognitive Model of Curiosity

長島 一真[†], 森田 純哉[†]

Kazuma Nagashima, Junya Morita

[†] 静岡大学

Shizuoka University

nagashima.kazuma.16@shizuoka.ac.jp

概要

認知モデルは、人間の属性を推測するためのツールとして使用される。認知モデルを使用すると、行動と認知特性の結びつきを実験的に仮定し、他者の特性帰属に伴うエラーを明確にできる。本研究では、ACT-Rモデルを使用して、「好奇心」に関する特性の帰属を検討した。先行研究において構築された複数の好奇心の認知モデルの振る舞いを可視化し、被験者の主観評価を行った。結果、モデル間での差異が見られ、モデルの振る舞いと特性の関連性が示唆された。

キーワード：属性帰属; 認知モデル; ACT-R

1. はじめに

人間は、他者の行動を、性格などの内的属性を想定することで解釈する傾向を有する。属性の推測には、人間による他の思考や意思決定と同様、根本的なエラー（バイアス）が含まれる (Gilbert & Malone, 1995)。ただし、これらのバイアスに関する研究は調査や実験をベースに行われ、行動の背後に存在する真値となる内部状態に対する操作が欠けていた。

この問題を解決するために、本研究では、認知モデルを人間による内部属性の推測過程を検討するツールとして利用することを提案する。認知モデルは、人間の内部状態に関する仮定から外的行動を導出するシミュレータである。認知モデルをこの領域の研究に応用することで、エラーを引き起こす要因を、行動の原因である内部メカニズムの観点から詳細化することが可能になる。

人が有する内部属性のなかで、モチベーションは、行動の原因として頻繁に帰属され、影響力を有するものである。Weinerによる帰属理論において、モチベーションは内部的かつコントロール可能なものに分類される (Weiner, 1985)。この次元への原因の帰属は、帰属された他者の事後における挑戦的行動の生起に大きく影響する。この意味でも行動からのモチベーションへの帰属過程を検討することは、理論的にも応用的にも重要である。その重要性から、モチベーションの自

律エージェントのモデルが提案されている (Ribeiro et al., 2017)。さらに、近年ではモチベーションの要因を分析しようという試みがある (Murayama & Jach, 2024)。よって、本研究ではモチベーションの認知モデルにより生成される行動を実験刺激とし、その帰属に影響する要因を検討する。

2. 好奇心の認知モデル

モチベーションの一つである好奇心は、教育やエンターテイメントなど、幅広い分野で個人の活動を促進するために重要な内的特性である。近年、好奇心に関する人間の神経科学に基づいた数理モデル (Friston, 2010) が提案され、それを基にした機械学習のエージェントの研究が注目されている (Aubret et al., 2019)。

上記の好奇心のエージェントは、内部処理が不透明な深層学習を用いて実装されている。このような研究に対して、著者らはこれまで、認知モデルの観点から、好奇心の研究を実施してきた (長島一真他, 2021)。認知モデルのアプローチに即すことで、人間が好奇心を発生させるプロセスをトレースすることが可能になる。

多様な認知モデルの手法があるなかで、著者らの研究では、認知アーキテクチャの1つであるACT-R (Adaptive Control of Thought-Rational (Anderson, 2007))を用いた。ACT-Rのモジュールとシンボリックプロセスを用いて、好奇心を「パターンの発見」に基づいて表現した。パターンの発見とは、人間が因果関係のパターンを発見し、組み合わせ、利用する能力のことである (Baron-Cohen, 2020)。パターンの発見は、好奇心の機械学習モデルで用いられる数理モデルの説明から発展するものである。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 (Friston, 2010)。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 (Koster, 2013;

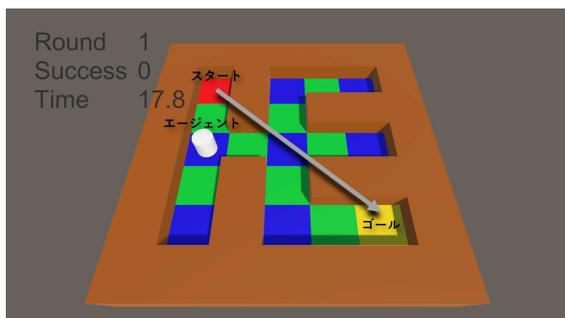


図1 シミュレータ

Schmidhuber, 2010).

3. 方法

本研究では、著者らが開発した好奇心の認知モデルを採用したうえで、モデルの予測と参加者による属性帰属の関係を検討することを目的とする。なお、この目的に関して、先行研究 (Nagashima & Morita, 2024) では既に予備的な実験の結果を示している。しかし、先行研究では操作されるパラメータ (好奇心や探索環境の広さ) に関するパラメータが限定されていた。本研究ではより完全なパラメータ集合を対象に、好奇心と属性帰属の関係を検討する。

3.1 参加者

300名の参加者を、クラウドソーシングサービスのランサーズにて募集した。そのうち、アンケートの回答の収集において問題が生じた参加者を除外し、254名のデータを分析対象とした。

3.2 材料

モデルが実行したシミュレーション課題を示し、その上で本研究で設定した実験条件と対応する3体の認知モデルを示す。本節の最後に、認知モデルを参加者に提示する環境を示す。

3.2.1 シミュレーション課題

本研究が参加者に提示する材料において、モデルは好奇心に基づいて、設定された5×5のマップを探索する。モデルがスタート位置からゴール位置まで移動するか、または制限時間 (180秒) に達するまでのプロセスを1回のラウンドとし、同一のマップに対する複数のラウンドを、課題全体の制限時間 (1080秒) に達するか、モデルが止めると判断するまで繰り返す。

3.2.2 実験条件

上記の設定で刺激されるモデルの好奇心を検討するために、モデルが有するエージェントの特性の条件を思考水準 (環境探索の戦略) の観点から操作した。モデルが有する思考水準は、モデルの思考量 (パターンの発見の量) によって区別される。モデルの特性として思考水準の高いモデルは、複雑な戦略を有し、課題

を遂行すると仮定される。思考水準の低いモデルは、戦略を持たずに課題を遂行するモデルと仮定される。以下に、モデルの概要を示す。

1. Random モデル：ランダムに環境を探索する。
2. 確率的 DFS モデル：確率的な DFS (depth-first search), によって環境を探索する。
3. 確率的 DFS + IBL モデル：確率的 DFS と現在の課題の解決に過去の記憶を用いる IBL (instance-base learning) を組み合わせる。

これらのモデルの思考水準は、1から3になるほど思考量が増加し、思考水準が高いと仮定される。

上記の各条件について好奇心の度合い (報酬) を3段階に変化させ、10回モデル化し、その結果を参加者に対する提示刺激とした。ラウンド数、課題継続時間 (モデルの思考時間)、ゴール達成率である。

3.2.3 エージェント評価用ウェブサイト

エージェントの動作を動画にし、それに基づいて参加者が主観評価を行うウェブサイトを構築した。エージェントの動作を示す動画は、Unity で実装されたモデルのログからエージェントの行動を復元するシミュレータによって作成した。図1は、そのインターフェイスである。画面には、参加者が現在の状況を把握するために、現在のラウンド数 (Round)、ゴール達成数 (Success)、現在時刻 (Time) が表示されている。

上記の情報に基づき、参加者はエージェントの属性帰属に関するアンケートに回答した。アンケートは、エージェントの行動に基づいて特性を評価する Godspeed 質問紙 (Bartneck et al., 2023) の「知性の有無 (Perceived Intelligence)」の項目を使用した。この質問紙では、モデルの思考水準が反映されることを期待した。さらに、好奇心の種類を詳細に検討するために、好奇心の性質を調査するためのアンケートである 5DS (Kashdan et al., 2018) を用いた。5DS は5つの好奇心のタイプを分類する。その中から本課題に関係のある「Joyous Exploration」と「Deprivation Sensitivity」を用いた。前者は好奇心の喜びや肯定的な経験のために物事を探求に関係があるとされ、後者は問題を解決や知識のギャップを埋めるような不安感を拭う探求に関係があるとされる。3.2.2節で示したモデルの仮定から、思考水準の高いモデルは、楽しみを感じると仮定される、そのため、Joyous Exploration が高くなることを期待した。一方、思考水準の低いモデルでは、思考量の少ない手数が多さから Deprivation Sensitivity が高くなることを期待した。

表1は、本研究で使用した質問票である。表からわ

表1 質問紙

Joyous exploration	
1	挑戦的な状況を、成長し学ぶ機会としてとらえる。
2	常に、自分自身や世界についての考え方に挑戦するような経験を求めている。
3	何かについて深く考えなければならないような状況を探し求める。
4	馴染みのないテーマについて学ぶことを楽しむ。
5	新しい情報を学ぶことに魅力を感じる。
Deprivation sensitivity	
1	難しい概念的な問題の解決策を考えていると、夜も眠れなくなることもある。
2	答えがわからないと休めないため、ひとつの問題に何時間も費やすこともある。
3	問題の解法がわからないと苛立ちを感じ、それを解決するためにさらに努力する。
4	解決しなければならぬと感じた問題には、執拗に取り組む。
5	必要な情報をすべて持っていないことに苛立ちを覚える。
Perceived Intelligence	
1	無能な - 有能な
2	無知な - 物知りな
3	無責任な - 責任のある
4	知的でない - 知的な
5	無能な - 有能な

かるように、5DSの各指標は5つの質問で構成されている。Godspeedから取られた知性の有無に関連する5組の反意語を含んでいる。すべての質問項目は5段階評価である。

3.3 手続き

ランサーズの募集に応募した参加者には、専用サイトで実験課題に関する説明ページが提示された。以下に、実験の手順を示す。

1. ランサーズの依頼画面にて参加登録
2. 課題説明の提示
3. 3つの思考水準分、課題画面にて以下の操作
 - (a) 3分間動画を視聴
 - (b) アンケート

参加者が説明を十分に理解したと自己判断した後、課題画面に移動した。この際、参加者に提示するモデルの行動は、各思考水準および各報酬に対して実行された10のシミュレーション結果から無作為に1つ選択された。各思考水準の表示順は、順序効果を考慮して、参加者ごとに無作為に決定された。参加者は動画を3分間視聴した後、アンケートに回答でき、「送信」ボタンを押すと次のモデルの課題画面に移動した。

3.4 分析

アンケートで得られた回答について、指標として5DSの2項目とGodspeedの1項目の平均得点を求め、アンケート項目の最大値(5)に対する割合に変換した。さらに、各項目の平均点を指標(応答変数)とし、一般線形混合モデル(Generalized Linear Mixed Model)を用いて尤度比検定を行った。説明変数には、質的変数である「モデル」、「報酬」とその「交互作用」を採用した。さらに、これらのモデルの内部属性に関して付随的に生じる「ラウンド継続数」、「継続時間」、「ゴール達成率」などの行動指標の影響を統制するために、これらを量的変数に含めた。さらに、変動効果(random effect)として10のシミュレーションの実行回の区別(刺激の種類)を加えた。

4. 結果

図2は、指標ごとに集計したアンケートの結果である。それぞれのバーの色は、モデルの思考水準を表し、x軸は報酬を表す。y軸は、アンケート項目に対して得られた評定得点をモデルの実行毎に平均し、それらを集計したものである($n = 10$)。

Joyous Explorationにおいて、モデルの主効果($\chi^2(2) = 17.07, p < 0.001$)およびモデルと報酬の交互作用($\chi^2(4) = 10.74, p < 0.05$)が有意となった。一方で、報酬の主効果($\chi^2(2) = 1.15, p = 0.56$)は有意とはならなかった。報酬6および18におけるモデルの単純主効果が有意となり、Turkey法による多重比較の結果、DFS+IBLとDFSがRandomよりも有意に高いことが確認された($p < 0.001$)。さらに、Randomにおける報酬の単純主効果も有意となり、報酬12が最も高い値となった($p < 0.05$)。DFS+IBLモデルにおいては、報酬18が最も有意に高い結果となった($p < 0.01$)。

Deprivation Sensitivityにおいては、モデルの主効果($\chi^2(2) = 16.61, p < 0.001$)、報酬の主効果($\chi^2(2) = 9.22, p < 0.05$)が有意となった。一方で、モデルと報酬の交互作用($\chi^2(4) = 8.56, p = 0.07$)は有意とはならなかった。さらに、各要因に対して多重比較を実施した。モデル間では、RandomがDFSよりも有意に高かったことが確認された($p < 0.001$)。報酬間では、報酬12が報酬18よりも有意に高かったことが確認された($p < 0.01$)。

Perceived Intelligenceにおいて、モデルの主効果($\chi^2(2) = 18.74, p < 0.001$)とモデルと報酬の交互作用($\chi^2(4) = 17.67, p < 0.01$)が有意だった。一方で報酬の主効果($\chi^2(2) = 1.03, p = 0.60$)は有意ではなかった。モデルの単純主効果では、すべての報酬でDFS+IBLモデルが最も有意に高かったことが示された($p < 0.001$)。報酬6では、DFSがRandomよりも有意に高かったことが確認された($0.001 < p$)。報酬の単純主効果では、報酬6の場合にDFS+IBLが最も高く($0.001 < p$)、DFSはRandomよりも高かった($0.001 < p$)。報酬12の場合にDFS+IBLは最も高かった($0.001 < p$)。報酬18の場合にDFS+IBLはDFSよりも高く($p < 0.01$)、Randomよりも高かった($p < 0.001$)。

上記の結果は、参加者が著者らの好奇心のモデルにおける思考水準の設定を推定できたことを示唆している。例えばJoyous Explorationでは、思考水準の高いDFS+IBLやDFSは報酬12で思考水準の低いRandomと差がないものの、他の報酬ではRandom

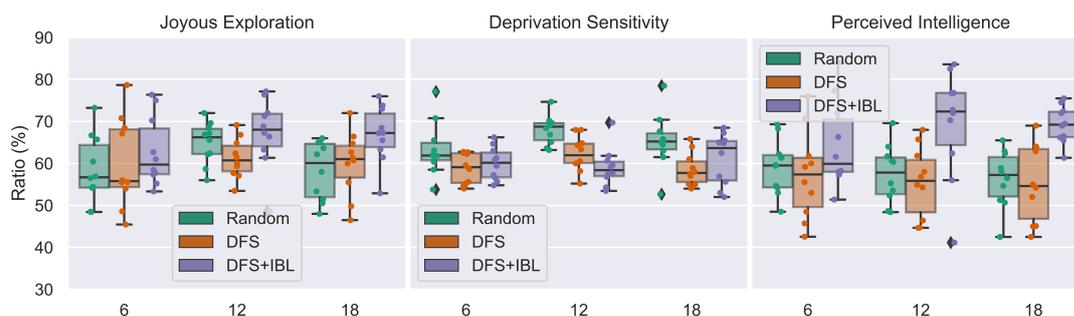


図2 アンケート結果

よりも高い結果が得られた。また、Deprivation Sensitivity では、モデルと報酬の交互作用はなかったものの、Random が最も高い結果となった。

5. まとめ

本研究では、ACT-R の好奇心のモデルを用いて、人間がエージェントの振る舞いからその特性を推定できるかを検討した。この目的のために、それぞれ異なる特性を持つ ACT-R のモデルの動作を復元するシミュレータを実装した。それを用いたオンライン実験では、参加者がエージェントの振る舞いからその特性をどの程度推定できるかを調査した。その結果は思考水準が最も高い DFS+IBL が楽しさを追求する好奇心が最も高く知的な振る舞いをするという結果になった。また、最も思考水準の低い Random モデルは、困難な状況を払拭する好奇心が最も高いという結果になった。以上は、エージェントが持つ実際の特性を参加者が評価可能であることを示唆する結果である。

本研究の意義は、認知モデルを用いて内部モデルを要因として特性の推定を行ったことである。認知モデルは内部プロセスに焦点を当て実装されるため、機械学習エージェントに比べて、人間が認知モデルの特性を推定する際のプロセスをトレース可能であるという利点がある。そのプロセスを解析することで、人間の特性の推定に応用可能であると考えられる。

今後、このアプローチをさらに進めていく必要がある。参加者に提示する情報や環境の改善が必要である。本研究では長時間にわたるモデルの行動を提示したが、それらの行動をクラウドソーシングによって集められた参加者が熟視することは想定しづらいよって、今後人間がモデルを推定できるようになるために、モデルが取り組む課題の改善やモデルの内面に関わる情報の提示や環境の検討を進めていく必要がある。

文献

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe*, New York: Oxford University Press.

Aubret, A., Matignon, L., & Hassas, S. (2019). "A survey on intrinsic motivation in reinforcement learning," *arXiv preprint arXiv:1908.06976*.

Baron-Cohen, S. (2020). *The pattern seekers: How autism drives human invention*, New York: Basic Books.

Bartneck, C., Cochrane, T., Nokes, R., Chase, G., Chen, X., Cochrane, T., Mitrovic, A., O'Sullivan, A., Wang, W., & Adams, B. (2023). "Godspeed questionnaire series: Translations and usage."

Friston, K. (2010). "The free-energy principle: A unified brain theory?" *Nature Reviews Neuroscience*, 11(2), 127–138.

Gilbert, D. T., & Malone, P. S. (1995). "The correspondence bias," *Psychological Bulletin*, 117(1), 21.

Kashdan, T. B., Stikma, M. C., Disabato, D. J., McKnight, P. E., Bekier, J., Kaji, J., & Lazarus, R. (2018). "The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people," *Journal of Research in Personality*, 73, 130–149.

Koster, R. (2013). *Theory of Fun for Game Design*, Sebastopol: O'Reilly Media.

Murayama, K., & Jach, H. (2024). "A critique of motivation constructs to explain higher-order behavior: We should unpack the black box," *Behavioral and Brain Sciences*, 1–53.

長島一真・森田純哉・竹内勇剛 (2021). 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, 36(5), AG21-E-1-13.

Nagashima, K., & Morita, J. (2024). "Trait Inference on Cognitive Model of Curiosity: Relationship between Perceived Intelligence and Levels of Processing," in *Proceedings of the 22th International Conference on Cognitive Modelling*, Advance online publication.

Ribeiro, E., Ribeiro, R., & de Matos, D. M. (2017). "A curiosity model for artificial agents," in *Artificial Life Conference Proceedings*, 561–568.

Schmidhuber, J. (2010). "Formal theory of creativity, fun, and intrinsic motivation (1990–2010)," *IEEE Transactions on Autonomous Mental Development*, 2(3), 230–247.

Weiner, B. (1985). "An attributional theory of achievement motivation and emotion," *Psychological Review*, 92(4), 548.