

# 認知バイアスを用いた「ミスをする」AIの設計

高橋 翔太<sup>1,a)</sup> 松原 仁<sup>1</sup>

**概要:** 人間はタスクを実行する際に、しばしば最適解以外の行動、つまりミスを犯す。言い換えると、行動の中でミスをすることは人間らしい行為と言える。そのため、ある特定のタスクを実行するエージェントに認知バイアスによるミスを実装し、人間らしいAIの設計を試みる。本研究では、タスクを将棋に設定し、将棋で起こりうる認知バイアスを分析した。分析の結果から、直近効果が将棋に作用していることが分かった。今後は直近効果を実装した将棋AIを用いて、定性的な実験を行う。

## Design of AI taking mistakes with cognitive bias

SHOTA TAKAHASHI<sup>1,a)</sup> HITOSHI MATSUBARA<sup>1</sup>

**Abstract:** When performing task, Human often do suboptimal behavior, or mistakes. In other words, Taking mistakes in perform is a human-like behavior. Therefore, we design a human-like AI by implementing mistakes caused by cognitive biases in agents that perform a certain task. In this study, We set shogi as task, and we analyzed what the possible cognitive biases in shogi were. As a result, it was found that recency effect is working in shogi. In the future, we will conduct qualitative experiments using a Shogi AI that implements the recency effect.

### 1. はじめに

囲碁や将棋をはじめ、多くのゲームにおいてAIは人間の成績を上回っており、今後も人間を上回るゲームの種類や実力の差は広がっていくことが予見される。しかし、自分の実力を確かめて上達するためにプレイする人間や、ただ楽しく遊ぶためにゲームをプレイする人間にとってはただ強いだけのAIでは不十分であり、対戦して楽しいものである必要がある。人間と対戦しているような楽しさを実現するには人間らしい思考を持ったAIが必要であるが、現行の多くの対戦型のゲームAIは最適な行動を取ることに特化されており、人間らしい思考を経て行動を取っているとは言えない。そのため、本研究では人間が最適な行動を取ることができない原因の一つであるミスが人間らしい行動につながるという仮定の下、人間が起こすミスを分析し、AIに実装し、評価を行う。

また本研究では、ミスの種類の一つである認知バイアス

によるミスを実装する。認知バイアスは知識や能力、疲れなどに関係なく起こる現象であり、精神疾患などの疾患がない人たちの行動の中に広く現れる偏りや歪みである [1] ことから今回の研究に適していると考えられる。

### 2. 関連研究

#### 2.1 ゲームAIへの人間らしさの実装

面白さを主題としたゲームAI研究では様々なアプローチが成されている。

池田ら [2] は人間を楽しませるための『接待碁』において人間プレイヤーを楽しませるために必要な要素をまとめ、いくつかのアプローチを示した。その中で、不自然な着手を減らすことが人間らしさに繋がると述べられており、高度すぎる手や下手な手加減を感じる手は人間らしさを損なうと指摘している。

藤井ら [3] は、ゲームAIに「疲れ」「ゆらぎ」などの人間の行動原則を実装し、それらが人間の実際のプレイよりも「人間らしい」と評価されることを示した。人間に共通する行動原則を実装することにより、ゲーム内容に依らずに

<sup>1</sup> 公立はこだて未来大学  
Graduate School of Future University Hakodate

<sup>a)</sup> g2120023@fun.ac.jp

ゲーム AI に人間らしさを実装するアプローチを開拓した。

## 2.2 将棋におけるミスの分析

伊藤ら [4] は将棋において起こりうるヒューマンエラーを大別し、思考過程の段階ごとにヒューマンエラーを分類した。結果として、無意識下で起きるエラーの割合が多いと指摘した。認知バイアスは主に無意識下で作用するため、将棋では認知バイアスが起こる可能性がある。

## 3. 研究プロジェクト

研究は以下の 4 つの段階に分けて行う。

- (1) 人間が認知バイアスにより行動が歪められうるタスクを選定する
- (2) そのタスクにおいて、どの場面でどういったバイアスが働いているのかを分析する
- (3) タスクを実行する AI を作成し、加えて認知バイアスを考慮した行動を取る AI を作成する
- (4) 2 つの AI と人間のプレイの 3 種類を比較し、認知バイアスを考慮したものとそうでないもので人間らしさの差がどれくらい出るのかについてを評価する

現在、(3) の実装を行っている。

## 4. 認知バイアスで行動が歪められうるタスクの選定

前述のとおり将棋は認知バイアスが起こりうるタスクであるため、本研究ではタスクを将棋に選定した。将棋は運が絡まない完全情報ゲームであり、プレイヤーのミスが分かりやすく表れると考えられる。また、対局後に感想戦を行う文化があることから、経験者は自分のミスやその時の思考について言語化することに慣れているため、分析がしやすい。更に、将棋は日本での人気が高いため実験の被験者を募りやすく、AI 研究も発達していることから今回の研究の題材とするタスクにふさわしいと考えられる。

## 5. 将棋で発生する認知バイアスについての予備実験

(2) の分析を行うにあたり、プレイヤーが自身で気づくミスと気づかないミスを顕出させ、それらを分類する必要があるため、予備実験を行った。前述の伊藤らの分析では自身で気づけないミスは対象にしていなかったが、認知バイアスによるミスは自身では気づけない場合もあるため、それも考慮した。

### 5.1 被験者

「将棋のルールを理解している」という条件で被験者を募り、16 名が参加した。被験者の棋力は、一部のルールを

表 1 知識不足に分類されたミス

その場の効きを見逃してしまった
ずっと守りが浅く、そこから攻められてしまった
前の手が苦しくて、焦って無理やり攻めようとした
攻めるスキがないときに何をすればいいかわからず、適当な無駄な手を打ってしまった
思いつく中で最善だった
打った手が思った通りの効果がなかった

表 2 判断不足に分類されたミス

相手の持ち駒が見えていなかった
「持ち駒を大切にしすぎる」という自分の癖が出た
打った後、次の相手が良い手を持っていることに気づいた

表 3 感情的な影響に分類されたミス

焦って間違えた手を指した
王手で焦って正常な判断ができなかった
なんで打ったのか分からない
考える余裕がなかった

理解していないレベルから将棋ウォーズ [5] 初段レベルまでの範囲であった。

### 5.2 実験内容

まず被験者に口頭で自身の棋力やブランクについて聞き込みした。被験者と将棋 AI の棋力に差が出てしまうとミスが顕出しづらくなってしまうため、聞き込みの内容から筆者が対戦する AI の棋力を設定した。対戦する将棋 AI には激指 13 を用いた。また、実験のリハーサルから早指しする AI に被験者が釣られてしまい十分に考えないまま指してしまうことが考えられたため、実験の説明の際に筆者から「AI は早指しするが、釣られずに好きなだけ考えてよい」と伝えた。

対局後に自分自身でミスだと自覚する箇所を指摘させ、「ミスを起こした理由」と「どの手が最善だと思うか」を質問した。その後、対局した AI とは別の将棋 AI に対局を評価させ、評価値が大きく下がった箇所を被験者のミスとして被験者に提示し、上記と同じ内容の質問をした。対局が 50 手未満で終わった場合は実力差が開きすぎているとみなし、難易度を調整したのちもう一度対局をさせた。

### 5.3 結果

結果、ミスがいくつかの傾向に分かれることが分かった。最も多かったミスは知識不足によるものであった。表 1 に主な回答を示す。

次に多かったミスは判断不足によるものだった。表 2 に主な回答を示す。

その他、対局の緊張により短絡的な手を指してしまうミスや、AI の好手のせいで焦ってしまい、精神的に追い込まれたことによる感情的なミスなどが見られた。表 3 に主な回答を示す。

また、AI の評価値が大きく下がった箇所について「どの手が最善だと思うか」を質問した結果、実力者の方が AI が最善とする手と同じ手を選ぶ傾向、つまりミスを実感している傾向にあった。それに対し初心者は指摘された手が何故悪い手だったのかが分からない場合が多く、基本戦術などの知識不足によるミスが多い傾向にあると考えられる。それに対し、AI がミスだと判断しなかった手を被験者がミスだと考える場合もあった。つまり、AI による指し手の評価基準とはかけ離れた評価基準を持っており、それによってミスが生まれるというパターンの存在も確認された。

#### 5.4 考察

知識不足によるミスは実力に応じて顕出する種類が変わることが分かった。例えば、初心者は「駒の動きを理解していなかった」などの初歩的な知識不足によるミスを起こすが、上級者は「相手の戦法への理解が足りていなかった」といったような高レベルなミスを犯し、駒の動きの理解などの初歩的なミスは絶対に起こさない。将棋 AI に実装する際は、難易度に応じて顕出されるミスの種類を変えていけばより人間らしい振る舞いが見込めるが、認知バイアスによるミスとは異なるため本研究では扱わない。

感情的な影響に分類されたミスは実験のデザインによって引き起こされたものだと考えられる。本実験の対局時間は最大 1 時間と設定していたため、被験者は一手ごとに平均で 1 分ほどの持ち時間で指す必要があった。そのため、より長い対局時間を用意すれば焦りによるミスは減らせると考えられる。また、筆者が事前に説明したにも拘わらず、AI の早指しに釣られたと考えられる回答も見られた。そのため、AI の手番で時間をわざと使い、考えているように見せるようなアプローチなどでミスを減らせる可能性がある。

判断不足によるミスについて、「自分の進めたい動きのみ考えていて相手の動きを気にしていなかった」というミスについては認知バイアスの一つである確認バイアスが作用していると考えられる。確認バイアスとは自分に有利な情報を優先し、自分に不利な情報を軽視して判断に用いてしまうといったものである。他にも、「目の前の手のことしか考えておらず、別の弱点を見落としていた」というミスには、最近注目した情報を優先して判断してしまう現象である直近効果も作用していると考えられる。

例として、図 1 に直近効果が作用したと考えられる盤面を示す。図 1 は先手（黒）を手番として黄色で表している。この盤面で被験者は「▲7 五金」と指したが、対局後にこの手はミスであったと指摘しており、その理由に「7 筋の攻めに目が行ってしまい、防御の方に目が行っていなかった」と回答している。これは直近の自身の指し手に注目して他の手の価値を甘く見たことによるミスだとわかる。直

近効果によるミスは対局の序盤以降に広く分布しており、将棋において広く影響が出ているバイアスといえる。



図 1 直近効果が作用していると考えられる盤面

## 6. 認知バイアスを実装した将棋 AI の作成

本研究では認知バイアスのうち『直近効果』と呼ばれるものを実装する [6]。直近効果は時系列の情報の中で、最近得た情報に大きな価値を置いて判断をしてしまうバイアスである。逆に、最初に得た情報に大きな価値を置いてしまう『初頭効果』もあるが、将棋の場合は初期配置が一定であり最初の記憶は判断に影響がないため考慮しない。

### 6.1 実装方法

将棋において、直近の情報は直前の相手の指し手である。それぞれの指し手は複雑な戦略の一部であり、直近の指し手に偏って着目するのは不合理であるが、直近効果を再現するために敢えて直近の手に過度な評価をするように実装を行う。それぞれの指し手には駒の位置座標や駒の種類、他の駒との位置関係など様々な情報が含まれている。直近効果は複数の情報を提示するとバイアスが強固になる [6] ため、着目する指し手の情報を増やすと顕著に影響が確認できると考えられるが、指し手を見てどれほどの情報を得るかは実力によって違う。本研究の目標は多くのプレイヤーに共通して発現するバイアスを実装することであるため、その中でも最もシンプルな駒の位置座標だけに着目してバイアスを実装する。

将棋 AI の既存の評価値  $V$  に、一つ前の相手の手の座標から距離が遠ざかるにつれて減少する評価値を加えた評価値  $V'$  を用いる。与式は、 $V(M)^{(n)}$  は  $n$  手目の指し手  $M$  の評価値、 $M_B^{(n)}$  は  $n$  手目の指し手の駒を動かす前の座標、 $M_A^{(n-1)}$  は  $n-1$  手目の指し手の駒を動かした後の座標、 $C$  は距離ごとに評価値が下がる割合のパラメータとして以下の式で表される。

$$M_B^{(n)} = (x_1, y_1) \quad M_A^{(n-1)} = (x_2, y_2)$$

$$V'(M)^{(n)} = V(M)^{(n)} - C\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

## 7. 実験

前章で作成した将棋 AI を用いて実験を行う予定である。予備実験の結果から実力の違いで顕出されるミスが違うことが分かったため、レーティングなどを用いて実力別に実験する予定である。

## 8. 今後の展望

今後は (4) の実験を行う。また、本研究はタスクとして将棋を用いたが、今後他のタスクへ拡張できるように汎用的な考察を行う。

### 参考文献

- [1] 鈴木 宏昭, “認知バイアス 心に潜むふしぎな働き”, 講談社, 2020
- [2] 池田 心, “モンテカルロ碁における多様な戦略の演出と形勢の制御 ～接待碁 AI に向けて”, 情報処理学会 ゲームプログラミングワークショップ 2012 論文集, pp.47-54, 2012
- [3] 藤井 叙人, “生物の基本原則の導入によるビデオゲーム COM プレイヤの「人間らしい」振る舞いの自動獲得”, 情報処理学会 エンタテインメントコンピューティングシンポジウム 2013 論文誌, 2013-EC-27 巻, 16 号, pp.1-6, 2013
- [4] 伊藤 毅志, “ゲームにおけるヒューマンエラー – 将棋における考察 –”, 情報処理学会 ゲームプログラミングワークショップ 2014 論文集, pp.196-201, 2014
- [5] 将棋ウォーズ, <https://shogiwars.heroz.jp/>
- [6] B.E. Turvey, J.L. Freeman, in Encyclopedia of Human Behavior (Second Edition), Academic Press, 2012