DQN と生物学的制約を用いた人間らしい振る舞いのゲーム AI

Video-Game Agents with Human-Like Behavior Using DQN and Biological Constraints

森田　隆弘
Takahiro Morita
法政大学情報科学研究科
E-mail: takahiro.morita.4p@stu.hosei.ac.jp

Abstract

Video game agents that surpass humans always select the optimal behavior, which may make them look mechanical and uninteresting to human players and audience. Since score-oriented game agents have been almost achieved, a next goal should be to entertain human players and audience by realizing agents that reproduce human-like behavior. A previous method implemented such game agents by introducing biological constraints into Q-learning and A search. In this paper, we propose video game agents with more entertaining and more practical human-like behavior by applying biological constraints into the deep Q-network (DQN). Especially, to reduce the problem of the conspicuous mechanical behavior found in the previous method, we improve the method of the biological constraint "tired", and propose additional biological constraints "confusion" and "carelessness". We implemented our method in the video game "Infinite Mario Bros." and conducted two types of experiments that were subjective evaluation. One is to evaluate the behavior of game agents that improved "tired". The other is to evaluate the behavior of game agents that introduced "confusion" and "carelessness" The results of both experiments indicated that the agents implemented with our method were rated more human-like than those implemented with the previous method.*

1. はじめに

近年，ゲーム AI の研究が盛んにおこなわれている．今までにも 2017 年にプロ棋士に勝利した将棋 AI Ponanza，同年にプロとの 3 番勝負にて勝利を収めた囲碁 AI AlphaGO [1]，最強テトリス AI として知られる Cold Clear など様々なゲームにおいて，プレイヤスキルを重視したゲーム AI の実現が達成されてきた．

このような人間を凌駕したゲーム AI は，常に最適な行動を選び続けるがゆえに，対戦相手の人間プレイヤや観客などに機械的に見えてしまい，面白さを損なう場合がある．人間よりもはるかに強いゲーム AI の実現が達成されてきた今，次の目標として，人間の知性を再現したゲーム AI を実現することで人間プレイヤを楽しませるようなエンターテインメント性が求められている．昨今のビデオゲームでは人間らしい挙動をするゲーム AI を Non-Player Character (NPC)で実装することでプレイヤを楽しませることに力を入れている．また，研究面でも様々な研究が行われているが，多くは恣意的に人間らしい行動を定義して学習させなければならず，ゲーム AI の開発者に多くの作業負荷がかかるものであった．

この問題を解決するために，藤井ら [2]は，生物学的制約を加えることで人間のような振る舞いのゲーム AI を生成するための手法を提案した．この手法によって，ヒューリスティックの導入をする際の開発コストを削減でき，さらに様々なゲームジャンルにおいても汎用的に導入することが可能という結果が得られた．しかし，その実装は，機械学習の中でも単純な手法である Q 学習や A*探索を用いたものであった．

本論文では，生物学的制約をゲーム AI に多く使われている機械学習手法である Deep Q-Network (DQN) [3]に導入することで，より実用的で楽しませることができる人間らしい振る舞いをするゲーム AI を実現する．最初に予備実験として，DQN による実装によってどのような違いが生まれるかを検証した．その結果，DQN に先行研究の生物学的制約を用いたゲーム AI では，ある程度人間らしい振る舞いを感じることができたが，それ以上に機械的な振る舞いが目立った．この問題を解決するために，本論文では新たに「混乱」「油断」という 2 つの生物学的制約を提案し，さらに先行研究の生物学的手法である「疲れ」をより人間に近くなるように改良した．提案手法の検証を行うために 2 つの実験を行った．1 つ目は「混乱」と「油断」に対しての実験，2 つ目は改良した「疲れ」に対する実験である．それぞれについて 10 人の被験者による被験者実験で主観評価を行った結果，提案手法を用いたゲーム AI の方が先行研究の手法のみを用いたゲーム AI よりもより人間らしい振る舞いをすると評価された．

2. 関連研究

人間らしい振る舞いをするゲーム AI の研究は今までにもされてきた．その 1 つに，人間のプレイヤの動きを模倣して人間らしい振る舞いを得る手法がある．Polceanu ら [4]は Unreal Tournament 2004 という 1 人称視点シューティングゲーム(FPS)で，リアルタイムで他プレイヤの動きを記録し，その記録したデータを断片的に再現していくことによって，人間らしい振る舞いのゲーム AI を実現

Supervisor: Prof. Hiroshi Hosobe

させた．Ortega ら [5]は Infinite Mario Bros.における人間プレイヤの進んだ経路について注目し，それをゲーム AI に模倣させるために，ハンドコーディング，教師あり学習に基づく直接的な手法，類似度測定の最大化に基づく間接的な手法の 3 つの手法を実装して比較した．その結果，類似度測定の最大化に基づく間接的な手法が最も優れているとされた．しかし，以上のゲーム AI は，人間らしい振る舞いを獲得するために人間プレイヤが実際にプレイしたデータを大量に必要とし，多くの開発コストがかかる．

開発コストの問題を解決するために，人間らしい振る舞いを自動獲得するゲーム AI の研究がされている．池田ら [6]は十分に強化された囲碁のゲーム AI に意図的に人間が起こしがちなミスをさせることによって，うまく手加減を行い人間プレイヤに楽しませる，いわゆる「接待碁 AI」の実現について研究した．しかし，この手法ではゲーム AI 開発者自身が人間らしいミスを定義しなくてはならず，開発における作業負荷が大きくなる．一方，藤井ら [2]の研究では，Q 学習や A*探索のゲーム AI に人間がプレイしている際に起こりうる，ゆらぎ，遅れ，疲れなどといった生物学的制約を用いて，人間らしい振る舞いを自動獲得させた．この研究では，それぞれの機械学習の手法に生物学的制約を用いたゲーム AI を Infinite Mario Bros.に適用し，その動きについて被験者実験による主観評価を行った．その結果，生物学的制約から人間らしいと考えうるゲーム AI を自律的に構成した．

3. 準備

3.1. Q 学習

Q 学習は 1992 年に Watkins ら [7]が機械学習手法として提案したものである．この手法はエージェント(学習者または意思決定者)がある状態の下でどのような行動をとるべきかという指標である価値推定関数$Q(s,a)$を更新していくことで学習を行う．エージェントは行動した結果に応じて報酬を受け取り，その報酬を用いて，$Q(s,a)$を更新する．更新の式は以下の通りである．

$$Q(s,a) = (1-\alpha)\,Q(s,a) + \alpha[R(s,a,s') + \gamma\, maxQ(s',a')]$$

ただし，αは学習率，γは割引率である．$(1-\alpha)$の項は今の Q 値を表し，αの項は学習で用いる値を表している．この式より，報酬が高いほど更新される Q 値が高くなることがわかる．

3.2. Deep Q-Network

Deep Q-Network (DQN) [3]は Q 学習にニューラルネットワークの考え方を含めた手法である．Q 学習では価値推定関数の表を更新していく仕組み上，連続的な状態を表そうとすると，膨大な数の状態数となり，学習を行うことが現実的に難しい．一方，DQN では Q 値の推定にニューラルネットワークを使用して，Q 値の近似関数を得ることで，複雑な状態の定義を行うことも可能となる．具体的には，最適行動関数をニューラルネットワークによる近似関数で求め，ある状態の時に行動毎の Q 値を推定できれば，最も Q 値の高い行動，すなわちとるべき最善の手が分かる．ある状態s_tを入力層，行動aを出力層のノードとなるニューラルネットワーク(図 1)を使用し，$Q(s_t,a_t)$を計算する．

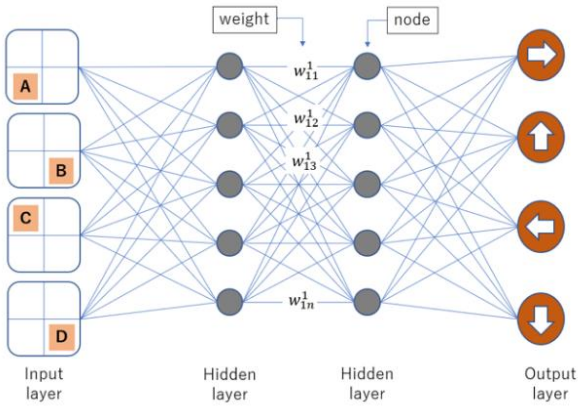

図 1　Deep Q-Network

3.3. 生物学的制約

学習を自律的に行ったゲーム AI は，操作が正確すぎる，敵やフィールドに対する反応速度が速すぎる，一定の行動を繰り返す，アクションゲームでいえば歩きやジャンプなどといった動きが一定で不自然であるなどといった，人間プレイヤには実行不可能な振る舞いが表出することが多い．このような問題を防ぐために，藤井ら [2]は，学習を自律的に行う機械学習手法と生物学的制約を融合させた．生物学的制約とは人間が生きる上で起こりうる身体的な制約や生き延びようとする心理的な制約のことである．生物学的制約によって，ゲーム AI における上記の振る舞いをなくし，ミスをさせることや弱くさせるといった調整方法に起こりうる「わざとらしさ」やそれに伴う「つまらなさ」を排除することが可能となる．

藤井らは Q 学習に以下の生物学的制約を定義し，導入した．

1. ゆらぎ(見間違い・操作ミスなど)：人間プレイヤは，必ず正確に見たものを判断できるわけはなく，また正確に操作できるわけでもないため，度々見間違いや操作ミスをなどが起こる．これを再現するためにゲーム AI が観測する情報に対してガウスノイズを付与した．
2. 遅れ(人間の反射神経)：人間プレイヤは観測した情報から実際に操作に移るまでに遅れが発生してしまう．これをゲーム AI が観測する情報を数フレーム前のものにすることで再現した．
3. 疲れ(キー操作によるもの)：人間プレイヤはコントローラを操作することによって疲れが生じ，パフォーマンスが落ちる．そこで，キー操作を変更する度に負の報酬を与えた．

3.4. Infinite Mario Bros.

Infinite Mario Bros. (図 2)は，任天堂から発売されたスーパーマリオワールドという 2D 横スクロールアクションを模したゲームである．このゲームの特徴として，ステージが事前に与えられたランダムな値によって無限に生成

される．Infinite Mario Bros.を対象としたゲーム AI についても多く研究されており，2009 年の Conference on Computational Intelligence and Games (CIG)ではこのゲームを対象とした Mario AI Competition [8]という大会が開催された．この大会運営より，ゲーム AI 実装用にゲーム内の敵の位置やマリオの周りのオブジェクトなどといった環境内のパラメータを変数によって取得できるように書き換えられたプログラムが配布されているため，強化学習を比較的容易に実装することが可能となる．本研究の予備実験ではこのプログラムを用いて実装を行った．

図 2　Infinite Mario Bros.のプレイ画面

4. 提案手法

本研究では，藤井らの研究の発展として，ゲーム AI の開発に多く使われている DQN に生物学的制約を用いた手法を提案する．この提案手法の目的は，人間らしい振る舞いと強さのバランスが最適であるゲーム AI の開発を可能にすることである．DQN の高い学習能力によって生物学制約を導入した際に起こるプレイヤスキルの低下を中和させることができると考えた．

DQN への生物学的制約の導入は，基本的には Q 学習に実装した方法と同じ方法で行う．DQN に状態を入力し，$Q(s_t, a_t)$を求める際や，行動した後の誤差関数を求める際に，nフレーム前の状態s_{t-n}を用いて計算することによって「ゆらぎ」と「遅れ」を再現した．藤井らの研究では，操作キーの変更した行動を行った後に得られる報酬に一定値の負の報酬を追加することで「疲れ」を再現していたが，本論文ではそれに加えて，操作キーの変更毎に微弱の「ゆらぎ」と「遅れ」のパラメータの上昇を行うことで改良すること提案する．

次のセクションで紹介する予備実験の結果から，DQN に先行研究で提案された 3 つの生物学的制約を組み合わせるだけでは，人間の反応速度を超えた最適な行動を行ってしまい，人間らしい動きが得られにくいことが分かった．本論文ではその問題を解決するために新たな生物学的制約である「混乱」と「油断」を追加する．人間プレイヤは処理しきれないほどの観測情報を得ると普段通りのコントローラ操作を行うことが難しくなり，最適とは言い難いぎこちない動きを行ってしまう．この生物学制約によって，ゲーム AI の，多くの敵に囲まれている状況でも動揺をあらわさずに敵を倒し進んでいくような機械的な動きを抑える．「混乱」は，プレイヤの周りの情報を受け取り，周りに敵を含むオブジェクトが一定数表示されていた際に，生物学的制約の「ゆらぎ」と「遅れ」のパラメータを増やすことで再現する．これを実装することで，ゲーム AI がより人間らしい行動を獲得できるようにする．「油断」とは，ゲームが順調な時に安全に進むような場面でも，気持ちが高ぶって人間がいつもとは違う行動やチャレンジするような行動をとってしまう生物学的制約である．これをゲーム AI に導入することで，プレイにおける緩急がつき，単調なプレイになることを防ぐ．

5. 予備実験

本研究の予備実験として，先行研究のゲーム AI を模した Q 学習と DQN のそれぞれに生物学的制約である「ゆらぎ」「遅れ」「疲れ」を導入したものとしてないものを Infinite Mario Bros.に適用し，人間らしい振る舞いについて比較する．本来は被験者実験による主観評価が望ましいが，予備実験では学習したゲーム AI の様子を見て著者自身による主観評価を行う．

主観評価を行うにあたって以下の点に注目した．

1. マリオの動きが不自然でないか
2. 人間の反射神経を超えたような動きではないか
3. 敵や穴の目の前で一瞬躊躇することや，安全に飛び越えようとすることがあるか
4. ダッシュやジャンプを行う際に不自然な様子はないか

表 1 は予備実験の結果をまとめたものである．最初に，生物学的制約を導入していないゲーム AI を比較する．生物学的制約を導入していない Q 学習では，特筆するような人間らしいと感じる振る舞いは見られなかった．また，生物学的制約を導入していない DQN では，Q 学習よりもさらに機械的な動きが顕著に表れ，敵や穴に対して一切の躊躇なく正確に最適な経路でスコアを上げていった．また，ほとんど止まらないように進んでいったため，ダッシュのスピードもほとんど変わらないといった様子であった．

表 1　予備実験の結果

手法	生物学的制約	主観評価
Q 学習	なし	特に人間らしい動きなし
Q 学習	あり	敵や穴の目の前でためらいが見えるなど，人間らしさが散見された
DQN	なし	常に最適な動きをしているため，人間らしさは一切ない
DQN	あり	全体的にやや人間らしさを感じる部分はあるが，機械的な動きの方が圧倒的に目立つ

次に，生物学的制約を導入しているゲーム AI と導入していないゲーム AI を比較する．Q 学習，DQN ではどちらも生物学的制約を導入したことによって，敵や穴の前での挙動が変化し，寸前でよける動きから少し余裕をもって安全に飛び越えるような動きへと変化した．ここでは，反応速度が速すぎることによる動きが抑制されていると感じられた．また多くの敵がいる場合に正確なコントローラの操作によって無理に突破するのではなく，進める状態になるまで待機するような動きも見られた．これらから人間らしい振る舞いがある程度獲得できたと考えられる．

最後に，生物学的制約を用いた Q 学習と DQN を比較する．Q 学習では，敵や穴の前にてためらいが見られ，また途中で止まって安全を重視するなどといった人間らしい振る舞いを確認できたが，最終的なスコアでは，あまり高くはなく，人間でいえば初心者程度のプレイに感じられた．一方，DQN では，最終的なスコアは十分に高く，人間らしい振る舞いもある程度確認できたが，それ以上に最適な行動によって動くことが多く感じられ，機械的な動きが目立っているように感じられた．

なお，この予備実験は著者自身の主観評価実験であり，どのゲーム AI がどの手法を導入しているかを知った上で実験を行っている．このため，実験結果の信憑性は比較的低く，改めて第 3 者の被験者による主観評価実験を行う必要がある．

6. 実験

6.1. 実験 1：混乱と油断に関する実験

実験 1 ではゲーム AI に生物学的制約を含まないもの，含むもの(4 フレームの遅延、疲れあり)，追加の生物学的制約として「混乱」を含むもの(4 フレームの遅延，疲れあり，混乱時 5 フレームの遅延，ゆらぎのパラメータ増加)，「油断」を含むもの(油断時 6 フレームの遅延，ε-greedy 法のためのランダム選択確率を 0.1 に増加)をそれぞれ Q 学習と DQN で実装して，被験者実験によって人間らしい行動パターンを評価する．本研究のゲーム AI ではコイン，ブロック，アイテムは無視するものとしている．実験の被験者として 20〜25 歳でスーパーマリオシリーズを合計で 10 時間以上プレイしたことがある人を集めた．それぞれの学習回数は 50,000 ゲーム，学習率αは 0.2，減少率γは 0.9，ε-greedy 法のためのランダム選択確率は 0.07 である．この被験者実験はオンラインで行われた．

実験 1 で用意したゲーム AI または人間のプレイ画面は表 2 のとおりである．4 つの Q 学習を用いたゲーム AI のプレイ映像，4 つの DQN を用いたゲーム AI のプレイ映像，1 つの人間がプレイしたプレイ映像である．それぞれのプレイ映像は，穴や敵が表れているところを 20 秒前後にトリミングしている．さらに混乱，油断を実装しているゲーム AI との比較をすることを考慮して，その生物学的制約が働いている場面を選んで映像を準備した．

被験者実験の流れは以下のとおりである．最初に著者がプレイしている様子を 1 分程見せ，人間がプレイしているマリオのイメージを被験者に持たせる．この時に安全に行動し，コイン，ブロック，アイテムを無視して動くようなプレイを見せる．次に被験者に 1 つずつランダムで選んだあらかじめ録画された 20 秒前後の 8 つのゲーム AI または人間のプレイ画面を見てもらい，7 段階で 2 つの質問「このゲーム AI は人間らしい動きをしているか(1 が最も機械的な動きを，7 が最も人間らしい動きをしているとする)」，「このゲーム AI はどのくらいのゲームスキルを持っているか(1 が初心者，7 が上級者とする)」に答えてもらう．さらに，それぞれのプレイに対しての自由記述を行わせた(例として「穴の前でためらうような動作があった」「簡単に敵を倒していて不自然だった」など).

表 2　実験 1 で用意したプレイ映像

ラベル	人または AI	生物学的制約	プレイ時間
[Q, None]	Q 学習	なし	19 秒
[Q, Imp]	Q 学習	あり	20 秒
[Q, Imp, Con]	Q 学習	あり(＋混乱)	23 秒
[Q, Imp, Car]	Q 学習	あり(＋油断)	25 秒
[DQN, None]	DQN	なし	18 秒
[DQN, Imp]	DQN	あり	21 秒
[DQN, Imp, Con]	DQN	あり(＋混乱)	23 秒
[DQN, Imp, Car]	DQN	あり(＋油断)	30 秒
[Player]	人	-	20 秒

表 3　実験 1 の結果

ラベル	人間らしさ	ゲームスキル
[Q, None]	1.9	2.7
[Q, Imp]	3.8	3.0
[Q, Imp, Con]	3.3	2.7
[Q, Imp, Car]	2.5	2.1
[DQN, None]	1.3	5.5
[DQN, Imp]	3.0	5.0
[DQN, Imp. Con]	4.2	4.9
[DQN, Imp. Car]	3.5	3.2
[Player]	5.2	4.3

実験結果は表 3 のとおりである．最初に，Q 学習と DQN を導入したゲーム AI の比較を行う．[Q, None], [DQN, None]ではともに人間らしいかどうかの質問で 1.9 と 1.3 で，かなりネガティブな評価となった．この結果から，人間らしい行動は一切見られず，機械的な動きが目立ったと考えられる．[Q, Imp], [DQN, Imp]との比較では，[Q, Imp]では 3.8 で人間らしい動きがみられたとポジティブな評価が多かったことに対して，[DQN, Imp]は 3.0 となっており，[Q, Imp]よりも「敵やブロックなどをうまくよけすぎている」などのネガティブな評価が多かった．しかし，[Q, Imp]のプレイやスキルの対しての評価は高くはなかった．[Q, Imp, Con]と[DQN, Imp, Con]との比較では，[DQN, Imp, Con] が[Q, Imp, Con]よりも人間らしい動きやプレイヤスキルの面でも評価が上回っており，さらにプレイスキルも[DQN, Imp, Con]がどの Q 学習のゲーム AI よりも高く評価された．よって，[DQN, Imp, Con]がうまくプレイするスキルを持ちつつ人間らしい動きも得ることができたことがわかる．[Q, Imp, Con]と[DQN, Imp,

Con]の評価も同様に DQN の方がどちらの項目も Q 学習よりも上回っている．

次に，Q 学習どうし，DQN 同士での比較を行う．Q 学習では[Q, Imp], [Q, None]よりも人間らしい動きを得られるという結果が得られているが，[Q, Imp]と[Q, Imp, Con]とを比較した際，[Q, Imp, Con]の方がネガティブな結果が得られ，「不自然な動きに感じた」などの意見もあった．それに対して DQN では生物学的制約を増やしていくと 1.7, 2,7, 4.3 と人間らしい動きに対するポジティブな評価が増えていった．

最後に，[Player]と[DQN, Imp, Con]，[DQN, Imp, Car]との比較を行う．人間らしい動きの評価では，[DQN, Imp, Con]は[Player]よりもプレイスキルの評価としては，0.6 上回っており，逆に人間らしい動きに対する評価は 1.0 下回っていた．したがって，[Player]と[DQN, Imp. Con]の間にいまだにギャップがあることがわかる．また，[Player]では「立ち止まって次の手を考えているような動きがあった」という人間らしい行動としてポジティブな意見があったのに対して，[DQN, Imp. Con]では「あまり立ち止まらずに行動をし続けており，少し不自然であった」などのネガティブな意見があった．また，[DQN, Imp, Car]ではプレイヤスキル，人間らしさのどちらの面でも[Player]の評価を下回っている．

6.2. 実験 2：疲れに関する実験

先行研究の藤井らの研究では 6.1 節で示されている実験のみ行われていたが，この実験だけでは「疲れ」が目的通りに実装されているかどうか判断することは難しい問題がある．それを解決するために本論文では先行研究で行われていた実験に加えて，実験 2 として「疲れ」のみの挙動を確認する被験者実験も行った．学習したゲーム AI には実験 1 と同じものに加えて，「疲れ」を改良したゲーム AI も用意した．実験 2 では 3 つの DQN を用いたゲーム AI と人間のプレイ映像を開始して 1 分前後の部分と開始して 30 分前後の部分の 2 つを 20 秒前後の長さにトリミングして用意し，その 2 つを比較してどの程度プレイスキルに変化があるか，人間らしい行動に変化があるかを自由記述で評価させた．この時，被験者には 2 つのプレイ映像の違いについては説明していない．実験 2 で用いられたゲーム AI の映像を表 4 に示す．

表 4　実験で用いられるプレイ映像

ラベル	人または AI	生物学的制約	プレイ時間(前半)	プレイ時間(後半)
[DQN, None]	DQN	なし	20 秒	19 秒
[DQN, Imp]	DQN	あり	21 秒	23 秒
[DQN, Imp, Tir]	DQN	あり(＋改良した疲れ)	24 秒	23 秒
[Player]	人	-	18 秒	21 秒

実験結果の自由記述をまとめたものを表 5 に示す．[DQN, None]，[DQN, Imp]では過半数の被験者が変化を感じなかった．この結果から先行研究の「疲れ」の実装方法であるキー操作を行う毎に負の報酬を与えるだけでは，あまり効果が得ることができないことが考えられる．次に，提案手法として改良した「疲れ」を実装した[DQN, Imp, IPF]についての評価を見る．[DQN, Imp, IPF]では 1 つ目に見せたプレイ映像よりも 2 つ目に見せたプレイ映像の方の動きが悪いように感じた被験者が多かった．またこの結果が[Player]と似たような結果になっていることから，提案手法の方が先行研究の手法よりも「疲れ」の実装として適していると考えられる．しかしながら，明らかにぎこちない動きをしているというネガティブな評価もいくつかあるため，徐々に増加させる「遅れ」「ゆらぎ」のパラメータについて調整する必要が考えられる．

表 5　実験 2 の結果

ラベル	評価
[DQN, None]	• 変化を感じなかった(7 人) • 1 つ目の方が上手に感じた(2 人)　等
[DQN, Imp]	• 変化を感じなかった(6 人) • 2 つ目の方がぎこちない動きをしていた(2 人)　等
[DQN, Imp, IPF]	• 変化を感じなかった(3 人) • 2 つ目が明らかにぎこちない動きをしていた(3 人) • 何となく 2 つ目の動きが悪く感じた(3 人)　等
[Player]	• 変化を感じなかった(4 人) • 1 つ目の動きが悪く感じた(1 人) • 何となく 2 つ目の動きが悪く感じた(4 人)　等

7. 議論

7.1. 混乱の実装によるゲーム AI への影響

最初に，混乱の実装によるゲーム AI への影響について議論を行う．[DQN, Imp. Con]は[DQN, Imp]よりも少ないプレイスキルの評価の低下に対して，人間らしい行動の高い評価を獲得できた結果から，提案手法によって高いプレイスキルのままで先行研究の手法よりも人間らしい挙動を獲得できることが示された．また，[DQN, None], [DQN, Imp]の自由記述で，敵が密集している場面などで敵を簡単に倒しているような人間らしい行動の評価としてネガティブな意見も多くみられたことに対して，[DQN, Imp. Con]ではそのような意見がなかったことから，敵やブロックなどのオブジェクトの密集した場面での，機械的な不自然な動きが起こるゲーム AI の問題が解決されたと考えられる．

[DQN, Imp, Car]はプレイヤスキル，人間らしさのどちらの面でも[Player]の評価を下回った．しかし，これは[DQN, Imp, Car]の映像の場面が油断している場面の映像として故意的に切り取っているからである可能性もあり，人間が油断しているようなプレイをしている映像と比較する必要があると考えられる．

7.2. DQN に導入した生物学的制約は十分か

次に，DQN に導入した生物学的制約は十分であったかについて議論する．[DQN, Imp. Con]と[Player]の比較ではいまだにプレイヤとの評価のギャップが存在する．DQN は生物学的制約を用いていない状態では非常に高いプレイスキルがあり，さらに生物学的制約を加えた後でも，その中で最適な動きを効率的に学習していくため，さら

に生物学的制約を加えてもプレイスキルをあまり下げず人間らしさを得られると考えられる．

本研究で加えた生物学的手法である「混乱」「油断」は，敵が密集している場合やうまく進み続けた場合の条件下での問題を解決するものであった．このような特定の条件下での生物学的制約はまだ数多く存在しており，それらを実装していくことで，機械的に見える問題を解決できると考える．

7.3. 提案手法はどのような場面で役に立つか

この提案手法で得られたゲーム AI がどのような場面で役に立つのかについて議論を行う．近年対戦ゲームなどで用いられる NPC などでは，大量の対戦データを用いて学習されたゲーム AI が使われているが，実装には大量のコストがかかる．この問題に対して提案手法を用いることで，戦って楽しいような人間らしい行動をするゲーム AI を自動獲得することが可能となる．他の利用方法として，近年，ゲームのプレイ映像を見て楽しむような趣味の人たちが多くいる．その人たち向けのエンターテインメントとして，見て楽しいゲーム AI を作ることができると考えられる．

7.4. 実験の方法

最後に，実験の方法について議論を行う．本論文では，被験者による主観評価実験によって人間らしさを検証した．しかし，主観評価実験では被験者によって感じ方がそれぞれ違うため被験者によって結果が大きく変わる可能性があり，結果の説得力は比較的低い問題が生じる．この問題を解決するために，人間らしさを定量的に定義して，実験を行うことが求められる．しかし，人間らしさは非常にあいまいであるため定量的に見ることは難しい．どの先行研究でも多くの場合，本研究と同じように被験者による主観評価実験のみを採用している．人間らしさが定量的に定義されれば，評価の曖昧性がなくなり，人間らしいエージェント生成の分野で大きな貢献をすることだろう．

8. まとめと今後の課題

本論文では，DQN に生物学的制約と組み合わせた際に，人間らしい動きをしたゲーム AI の実現ができるかどうかについて注目した．本研究の目的は，人間らしい動きを残しつつ，ある程度のゲームスキルを実装するゲーム AI を自動で生成することである．予備実験の結果から，DQN に生物学的制約である「ゆらぎ」「遅れ」「疲れ」を導入したゲーム AI は人間らしい振る舞いをある程度確認できたが，それ以上に機械的な挙動が目立った．その問題を解決するために私たちは新たな生物学的制約である「混乱」「油断」を実装し，先行研究の「疲れ」の改良も行った．その結果，主観評価実験によってより人間のプレイに近いような実験結果が得られることを示した．

今後の課題として，2 点について考える必要がある．

1. 新たなの生物学的制約の検討
2. Infinite Mario Bros.以外のゲームへの適用

生物学的制約を導入したゲーム AI は，その制約の中で最適な行動を行うよう学習する．このため，高度な学習が可能な DQN に，人間に起こりうる生物学的制約を考えうる限り実装することで，より完全な人間らしい振る舞いを高いゲームスキルで獲得できるのではないかと考えている．実装すべき生物学的制約の検討しているものの例として「焦り」が考えられる．「焦り」とは，人間は劣勢であったり，制限時間がわずかになったりした際に，できるだけ良い結果にしようと急ぐことで本来の動きよりも前のめりになってしまう心理状態である．このような新たな生物学制約をさらに追加することで，人間らしい振る舞いが獲得できるかを検証する．

本研究の提案手法は藤井らの研究同様，どのようなゲームジャンルにも適用できることを目的の 1 つとしている．よって，Infinite Mario Bros.以外にも適用し，十分に人間らしい振る舞いを行うことができるかを検証する必要がある．問題としては，アクションゲームやシューティングゲームなどのように逐次行動が可視化可能ではなく，シミュレーションゲームなどのように動きの少ないゲームでは人間らしい行動かどうかを判断することは極めて難しいため，さらなる検討が必要である．

文　献

[1] D. Silver, H. Aja, C. J. Maddison, A. Guez, L. Sifre, G. v. d. Driessche and J. Schrittwieser, "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature*, vol. 529, pp. 484-489, 2016.

[2] N. Fujii, Y. Sato, H. Wakama, K. Kazai and H. Katayose, "Evaluating Human-like Behaviors of Video-Game Agents Autonomously Acquired with Biological Constraints," *Proc. ACE, LNCS*, vol. 8253, pp. 61-76, 2013.

[3] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, l. Antonoglou and H. King, "Human-Level Control through Deep Reinforcement Learning," *Nature*, vol. 518, pp. 529-533, 2015.

[4] M. Polceanu, "Mirrorbot: Using Human-Inspires Mirroring Behavior to Pass a Turing Test," *Proc. IEEE CIG*, pp. 1-8, 2013.

[5] J. Ortega, N. Shaker, J. Togelius and G. N. Yannakakis, "Imitating Human Playing Styles in Super Mario Bros," *Entertainment Computing*, vol. 4, no. 2, pp. 93-104, 2013.

[6] K. Ikeda and S. Viennot, "Production of Various Strategies and Position Control for Monte-Carlo Go—Entertaining Human Players," *Proc. IEEE CIG*, pp. 1-8, 2013.

[7] C. J. Watkins and P. Dayan, "Q-Learning," *Machine Learning*, vol. 8, pp. 279-292, 1992.

[8] J. Togelius, S. Karakovskiy and R. Baumgarten, "The 2009 Mario AI Competition," *Proc. IEEE CEC*, pp. 1-8, 2010.

[9] 片上大輔, 鳥海不二夫, 大澤博隆, 狩野芳伸, 稲葉通将, 大槻恭士, "人狼知能研究のすすめ," 知能と情報, vol. 30, no. 5, pp. 236-244, 2018.