Sarsa(λ)アルゴリズムに生物学的制約を導入した人間らしく振る舞うゲーム AI

Game AI with Human-like Behavior Based on the Sarsa(λ) Algorithm and Biological Constraints

久保　貴寛
Takahiro Kubo
法政大学情報科学部コンピュータ科学科
E-mail: takahiro.kubo.5t@stu.hosei.ac.jp

Abstract

In video games, AI is often introduced to entertain players. However, most game AI has been designed to take optimized action, which makes people feel like that the behavior of game AI is mechanical. Fujii et al. achieved game AI with human-like behavior by introducing four biological constraints defined as human native properties, i.e., "fluctuation", "delay", "fatigue", and "the balance between practice and challenge". Also, Glavin et al. achieved more natural game AI that have a long perspective by introducing the Sarsa(λ) algorithm that only uses the action that the game AI actually does and by using the eligibility trace able to assign a current reward to past actions. This paper proposes a method that introduces biological constraints into the Sarsa(λ) algorithm and applies it to the fighting game called FightingICE. By contrast to Glavin's study with the problem that game AI is not susceptive to the change of parameters because of the used FPS game with implicit randomness, this paper adopts the fighting game that is a one-vs-one match game without random elements, making it suitable to the Sarsa(λ) algorithm. To show the effectiveness of this method, this paper presents the results of a comparative experiment based on a questionnaire.

1. はじめに

近年，ゲーム業界ではプレイヤのゲーム体験を向上させるものとして，人工知能(AI)を利用する機会が増えてきており，ゲーム AI が注目を集めている．その中でも，格闘ゲームや First Person Shooter (FPS)といった対戦ゲームでは，AI がプレイヤ同士での対戦をより良いものにするための練習相手や，１人で気軽にゲームを楽しみたい時の対戦相手として活躍する場面が多く，AI はプレイヤがゲームを楽しむために必要不可欠な要素となっている．しかし，AI を搭載している既存の NPC は最適化された行動を追及しているものが多く，このような AI はプレイヤに対して機械的な印象を与えてしまうため，対戦ゲームのような人間を相手とする機会の多いゲームでは役割を十分に果たせない．そのため，対戦ゲームではプレイヤのゲーム体験を向上させるために，より人間らしい振る舞いをする AI を開発する必要がある．

本研究では，Sarsa(λ) [1] アルゴリズムに藤井らの生物学的制約 [2]を導入することで，従来の Q 学習ベースのゲーム AI と比べてより人間らしい振る舞いを獲得する手法を提案する．実験環境としては，研究に適した格闘ゲームの１つである FightingICE を使用する．また，格闘ゲームでは，プレイヤが取った行動が即座に結果として与えられるものだけでなく，未来の状態への布石として扱われることもある．そこで適格度トレースを導入することで，長期的な視点での行動選択が出来るようになり，人間らしい振る舞いを獲得出来ると考えた．

本研究で提案する手法の有効性を示すために，提案手法によって作成した AI と既存の手法によって作成した AI の比較実験を行う．実験方法としては，作成した AI と FightingICE に付随していた AI (MctsAi23i)との対戦動画をそれぞれ用意し，被験者に視聴してもらった後，人間らしい振る舞いが見られたかの主観的な 5 段階評価を作成した AI それぞれに行ってもらう．また，被験者には，なぜそのような評価となったのかの記述もしてもらう．

2. 関連研究

藤井ら [2]は，人間プレイヤを楽しませるための人間らしい AI を獲得する手法として，人間の生物学的制約の条件下での強化学習を提案した．人間の生物学的制約とは，人間が生来有している性質から生まれる制約や欲求のことであり，人間の行動制御における制約や自己実現理論を基に，それらをそれぞれ「ゆらぎ」「遅れ」「疲れ」の身体的な制約と，生き延びるために必要な欲求である「訓練と挑戦のバランス」に分類し定義した．藤井らは，これらの制約を Q 学習に組み込むことで，AI の人間らしい振る舞いを獲得している．任天堂が開発したスーパーマリオワールドを基に作成された Infinite Mario Bros.というブラウザゲームを採用している．

Glavin ら [1]は，ゲーム AI がリアルタイムでの意思決定が必要な環境下でも，戦略を学習し，適応し続けるような自然で挑戦的な対戦相手となることを目標に，Sarsa(λ)アルゴリズムを導入した AI を開発した．この Sarsa(λ)アルゴリズムとは，参考となるモデルを必要とせ

Supervisor: Prof. Hiroshi Hosobe

ず，AI 自身の経験から学習を行える強化学習の 1 つである．このアルゴリズムは，ある状態での行動の価値を Q 値という数値で表現し，この数値を特定の行動方策に基づいて更新していくことで学習を行っている．また，このアルゴリズムはオンポリシー手法を採用しており，実際に実行した行動を基に学習を行える．加えて，Glavin らは，このアルゴリズムに適格度トレースを導入し，現在の報酬を過去の行動にも適切な値で与えられるようにすることで，連続的な行動の学習も行った．提案手法を検証する環境としては，優先すべき事柄が変化しやすい FPS ゲームの Unreal Tournament 2004 を採用している．そのため，Glavin らは環境への適応を容易にするために，それぞれ Danger, Replenish, Explore という 3 つの異なるモードの開発を行った．それぞれのモードは独立して学習しており，状況に応じてモードを切り替える手法を採用している．実験では，Sarsa(λ)アルゴリズムで必要となる割引率と減衰率というパラメータを変化させながら，デスマッチサーバーにて対戦を行わせた．結果としては，良好なパフォーマンスを披露出来たものの，パラメータの変化に対して鈍感であるといった課題も見つかった．Glavin らはこの原因として，学習アルゴリズムが高レベルな設計になっていることと，ゲーム内の暗黙的なランダム性が問題であると考えた．

3. 準備

3.1. 生物学的制約

本研究で用いる生物学的制約は藤井ら [2]が定義したものを扱うものとする．生物学的制約の具体的な導入方法は以下の通りである．

- ゆらぎ：人間はゲーム情報を把握する際に，実際のゲーム情報と認知する情報の間に誤差が生じてしまう．そこで，ゲーム AI が観測するゲーム情報に対し，ガウスノイズを付与することで誤差を表現する．
- 遅れ：人間はゲーム情報を認識してから動作を起こすまでの間に時間差が生じてしまう．そこで，数百ミリ秒過去の情報を使用することで時間差を表現する．
- 疲れ：人間は短時間に連続でコントローラーのキー操作を行うと疲れが生じ，行動選択に悪影響を及ぼしてしまう．そこで，キー操作変更を行う際に負の報酬を与えることで表現する．
- 訓練と挑戦のバランス：人間は同じ行動を繰り返すことで訓練を行う．しかし，人間は訓練の中で失敗を繰り返した時，新たな動きに挑戦しようと考える傾向がある．そこで，訓練時に失敗を繰り返した場合，挑戦の傾向を強めることでバランスを表現する．

3.2. Q 学習

以下の数式は Q 学習 [3]における更新式であり，時刻tにおける状態s_tで選択した行動a_tの Q 値を，次の時刻$t+1$での状態s_{t+1}，行動a_{t+1}，報酬r_{t+1}を基に更新する

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \gamma \max Q(s_{t+1}) - Q(s_t,a_t)]$$

ただし，α, γはそれぞれ学習率と割引率のパラメータである．また，AI が行動を選択する手法としては，ε-greedy 法を採用する．生物学的制約の 1 つである訓練と挑戦のバランスは負の報酬を繰り返した場合，εの値を増加させることで表現する．

Q 学習では状態s_tにおける行動a_tの Q 値を，次の時刻$t+1$の状態s_{t+1}で最適な行動を必ず選択すると仮定して更新を行っている．そのため，実際には AI が最適でない行動を選択し悪い結果となってしまった場合でも，高い数値を獲得する可能性がある．

3.3. Sarsa(λ)アルゴリズム

以下の数式は Sarsa(λ)アルゴリズム [1]の更新式である．

$$e(s,a) = \begin{cases} e(s,a) + 1 & (s = s_t \text{かつ} a = a_t) \\ \gamma\lambda e(s,a) & (s \neq s_t \text{または} a \neq a_t) \end{cases}$$

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r_{t+1} + \gamma Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)]e(s,a)$$

Sarsa(λ)アルゴリズムでは Q 学習と異なり，状態sにおける行動aの Q 値を，次の時刻$t+1$の状態s_{t+1}で実際に選択した行動a_{t+1}を用いて更新を行っている．そのため，方策に一貫した学習を行え，より現実的な動きを模倣出来る．

本研究では Sarsa(λ)アルゴリズムに適格度トレース$e(s,a)$も導入している．適格度トレース$e(s,a)$は現在の報酬をそれまで実行してきた状態sと行動aのペアにも与えられる．具体的には，全ての状態sと行動aのペアに対して重みづけを行い，時刻tで実際に選択した状態s_tと行動a_tのペアに 1 を加算，それ以外のペアに対しては減衰率λと割引率のパラメータγをかけることで，その行動を選択した瞬間が現在に近しいほど適格度トレース$e(s,a)$が高い数値となるように更新を行っていく．そのため，長期的な視点での行動選択が可能となっている．Sarsa(λ)アルゴリズムでは Q 値を更新する際に全ての状態sと行動aを更新する．

4. 提案手法

本研究では，ゲーム AI に藤井らが定義した生物学的制約を導入し，さらに従来の Q 学習の代わりに Sarsa(λ)アルゴリズムを導入する．これは人間が選択を行う際の非合理性を表現するためである．人間は選択を迫られた際に必ずしも期待値の高い行動を取るわけではなく，選択後に発生する可能性のあるリスクも考慮して行動を選択している．そのため，選択後のミスによるリスクも考慮する Sarsa(λ)アルゴリズムで学習させることにより，人間らしい振る舞いの獲得を目指す．

4.1. アルゴリズムの構成要素

強化学習でゲーム AI が実行した行動に対し評価を返す報酬rの計算式は，AI の性能を左右する重要な要素である．Sarsa(λ)アルゴリズムでは適格度トレースを導入することで，現在の報酬を過去の行動にも与えられるため，長期的な評価が出来る．しかし，本研究で扱う格闘ゲームでは，自身もしくは相手の HP に変化が生じる瞬間よりも，互いの HP に変化がない時間の方が圧倒的に多い．そのため，キー操作による疲れのみが与えられる時間が多く，序盤で獲得した正の報酬が疲れの累積により，終盤でかき消されてしまう可能性がある．そこで，相手に

ダメージを与えた際の正の報酬を 10 倍に設定し，疲れによる負の報酬との差を大きくすることで，この問題を解決する．

学習に用いる報酬rの計算式は以下の通り定義する．

$$10 \times \mathrm{Damage}_{\mathrm{enemy}} - \frac{\mathrm{HP}_{\mathrm{enemy}}}{\mathrm{HP}_{\mathrm{self}}} \times \mathrm{Damage}_{\mathrm{self}} - \mathrm{fatigue}$$

相手に与えたダメージ量$\mathrm{Damage}_{\mathrm{enemy}}$を正，自身が受けたダメージ量$\mathrm{Damage}_{\mathrm{self}}$を負の基本的な報酬とする．自身が受けたダメージ量$\mathrm{Damage}_{\mathrm{self}}$による報酬は，自身の HP 量$\mathrm{HP}_{\mathrm{self}}$と相手の HP 量$\mathrm{HP}_{\mathrm{enemy}}$の差によって価値を変動させるために$\frac{\mathrm{HP}_{\mathrm{enemy}}}{\mathrm{HP}_{\mathrm{self}}}$倍にしている．生物学的制約における疲れは計算式内の変数fatigueを導入することで表現している．この疲れによる負の報酬はキー操作の変更を行う度に発生するため，互いの HP が変化のない場合でも負の報酬を与える仕組みとなっている．

4.2. FightingICE への適用

提案手法の検証を行う環境としては，格闘ゲームの 1 つである FightingICE を採用する．FightingICE では 1 フレーム毎のゲーム情報を取得出来るため，機械学習の計算式に生物学的制約を表現したパラメータを加えるのに適している．生物学的制約の 1 つである遅れは FightingICE の情報の取得に 15 フレーム以上必要であるという特徴で表現する．

格闘ゲームでは，過去の行動によって引き起こされた状況が，その瞬間の優劣に大きく影響を及ぼすことがある．そのため，本研究では Sarsa(λ)アルゴリズムに導入している適格度トレースによって，AI に長期的な視点を持たせることで，人間らしい振る舞いを獲得出来ると考える．

Glavin らの研究で採用されていた Unreal Tournament 2004 は複数人のプレイヤが 1 つのマップ内でアイテムを拾いながら撃ち合うといったゲーム内容であるのに対し，FightingICE ではキャラクターの行動のみでの 1 対 1 の対戦となっている．そのため，ゲーム内の暗黙的なランダム性は少なく，パラメータの鈍化の課題を解決出来ると考える．

5. 実装

本研究では，提案した手法の有効性を検証するために，以下の 4 つのゲーム AI を実装した．

① Q 学習のみの AI
② Q 学習に生物学的制約を導入した AI
③ Sarsa(λ)アルゴリズムのみの AI
④ Sarsa(λ)アルゴリズムに生物学的制約を導入した AI

実装した全ての AI で使用する学習率α，割引率γ，ε-greedy 法で用いる基本確率εは，藤井ら [2]が使用したパラメータを基にそれぞれ 0.2，0.9，0.0125 とした．生物学的制約を用いた AI で，疲れを表現する負の報酬 fatigue は-5に設定し，ゆらぎを表現するノイズの発生確率は 20%とした．訓練と挑戦のバランスによって高められたεは藤井らの研究を基に 0.2125 とし，負の報酬を取得してしまった状況が 5 回連続で発生した場合を，失敗を繰り返している局面として定義した．Sarsa(λ)アルゴリズムを用いた AI では，適格度トレースで使用する減衰率λの適切な数値を調べるために予備実験を行った．

学習に用いる状態s，行動aは以下の通り定義した．

- 状態s：相手との距離(3 種類)，相手との HP の差(3 種類)，相手の行動(5 種類)から成る合計 45 個の組み合わせ
- 行動a：地上で可能な行動(ダッシュを除く)25 種類と空中で可能な行動 15 種類の合計 40 種類の組み合わせ

生物学的制約におけるゆらぎは，AI がゲーム情報を観測する際にノイズを付与し，自身の状態を誤認させることで表現している．FightingICE では，キャラクターが空中にいるか，選択した技を実行するための Energy が足りているかによって実行の可否が決まる技が存在する．そのため，本研究では，現在のキャラクターの状態で実行出来ない技を選択肢から除外するようにしている．また，ダッシュコマンドは連続したフレーム間で選択した場合に AI の挙動にもたつきが生じるため，本研究では除外するものとした．

6. 予備実験

6.1. 手順

本実験は，本研究で実装する Sarsa(λ)アルゴリズムを用いたゲーム AI でのみ使用する減衰率λの適切な数値を調べるための実験である．本実験では 5 名の被験者による主観評価の実験を行った．

実験方法は以下の通りである．まず，減衰率λを 0.1～0.9 まで 0.1 区切りに数値を変更した Sarsa(λ)アルゴリズムに生物学的制約を導入した AI と，FightingICE に標準で実装されている AI (MctsAi23i)との対戦動画をそれぞれ用意し，被験者に順番通りに一度だけ視聴してもらう．動画を視聴する順番としては，順番によって偏った評価にならないように，実験計画法のラテン方格法を参考に実験を行った．その後，それぞれの対戦動画で登場した Sarsa(λ)アルゴリズムに生物学的制約を導入した AI に対して，振る舞いに人間らしさを感じたかを 5 段階で評価してもらう．5 段階評価では，1 の数値に近づくほど機械らしい振る舞いをしていることを表し，逆に 5 に近づくほど人間らしい振る舞いをしていることを表している．

6.2. 結果

図 1 は本実験の結果を全被験者の評価の積み上げグラフとして示したものである．減衰率λは，0.1 から 0.5 までの範囲では数値を増加させるほど人間らしい振る舞いに近づいていくが，0.6 以上の範囲では逆に人間らしい振る舞いから遠ざかっていくことがわかった．これにより，0.5 が現在の状況を引き起こす原因となった過去の行動に最適な報酬を分配出来る減衰率λの数値であることがわかった．これ以降の実験では Sarsa(λ)アルゴリズムを導入した AI の減衰率λに 0.5 を採用するものとする．

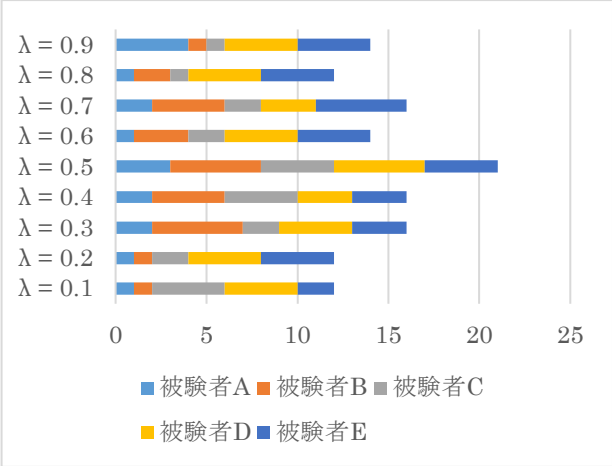

図 1　予備実験の実験結果

6.3. 対戦 AI との勝率

本節では，実装した 4 つのゲーム AI と対戦 AI (MctsAi23i)とのそれぞれの勝率の調査を行った．実装した AI の内，Sarsa(λ)アルゴリズムを導入した AI には，前述で定めた減衰率λを使用した．調査方法としては，実装した 4 つの AI と対戦 AI との対戦をそれぞれ 4 試合(12 ラウンド)行い，各試合の勝敗と取得したラウンドの記録を行った．

調査結果としては，表 1 の通りである．この結果から，全ての AI が対戦 AI に対して，多くの勝利数を収めていることがわかる．そのため，強さという指標では全ての AI が対戦 AI より優れていると考えられる．また，それぞれの AI の勝率としては，Sarsa(λ)アルゴリズムのみを実装した AI と Sarsa(λ)アルゴリズムに生物学的制約を導入した AI が最も高い勝率を示しており，次いで Q 学習のみを実装した AI，その次に Q 学習に生物学的制約を導入した AI という結果となった．この結果を基に考察を行っていく．まず，Sarsa(λ)アルゴリズムを導入した 2 つの AI が最も高い勝率であったことから，Sarsa(λ)アルゴリズムは Q 学習と比べて，安定した結果をもたらせた．これは，Sarsa(λ)アルゴリズムの学習における実際に選択した行動の Q 値を扱うという特徴により，Q 学習と比べてリスクを避けた学習を行えた結果だと考える．次に，Q 学習のみを実装した AI と Q 学習に生物学的制約を導入した AI の勝率に着目すると，生物学的制約を導入することで，勝率が低下してしまっていることがわかる．これは，生物学的制約が AI に人間らしい振る舞いを持たせるために，行動に制限をかけるといった手法であることから，行動の制限が強さという面で悪い結果をもたらしてしまった原因であると考える．

表 1　FightingICE に付随していた AI (MctsAi23i)との対戦結果

	1 試合目	2 試合目	3 試合目	4 試合目
Q 学習のみ	○ 2 - 1	○ 2 - 1	○ 2 - 1	○ 2 - 1
Q 学習+生物学的制約	○ 2 - 1	× 1 – 2	○ 2 - 1	○ 2 - 1
Sarsa(λ)アルゴリズムのみ	○ 3 – 0	○ 2 - 1	○ 2 - 1	○ 2 - 1
Sarsa(λ)アルゴリズム+生物学的制約	○ 3 - 0	○ 2 - 1	○ 2 - 1	○ 2 - 1

7. 実験

7.1. 手順

本実験では，実装した 4 つのゲーム AI の人間らしさを被験者にそれぞれ評価してもらい，提案した手法の有効性を検証する．本実験では 11 名の被験者による主観評価を行った．

実験方法は予備実験と同様，4 つの AI と FightingICE に標準で実装されている AI (MctsAi23i)との対戦動画をそれぞれ用意し，被験者に順番通りに一度だけ視聴してもらうものである．本実験も視聴する順番による影響を防ぐために，ラテン方格法を参考に実験を行った．本実験では，視聴してもらった対戦動画に登場する AI の振る舞いに関して 5 段階評価をしてもらうだけでなく，その理由も記述してもらった．5 段階評価の指標としては予備実験と同様のものを使用した．

7.2. 結果

図 2 と表 2 はそれぞれ本実験の評価の積み上げグラフと自由記述による評価理由を示したものである．図 2 より，Q 学習のみを実装した AI が最も人間らしいと評価され，次いで Sarsa(λ)アルゴリズムに生物学的制約を導入した AI が人間らしいと評価された．表 2 の自由記述による理由に関しては，まず，Q 学習のみを実装した AI が初心者らしく，最も攻めの姿勢を取っているように感じられたといった意見があった．Q 学習に生物学的制約を導入した AI と Sarsa(λ)アルゴリズムのみの AI の評価は似ており，どちらも相手のジャンプ攻撃に対する反応が速すぎるという意見が多くあった．しかし，Sarsa(λ)アルゴリズムのみの AI に関しては，後退することで相手の攻撃を回避している場面が見られ，そこが人間らしいという意見もあった．Sarsa(λ)アルゴリズムに生物学的制約を導入した AI では後退だけではなく，ジャンプ攻撃や遠距離攻撃といった攻撃手段も多く使用されており，4 つの AI の中で最も攻撃のパターンが多いという意見があった．

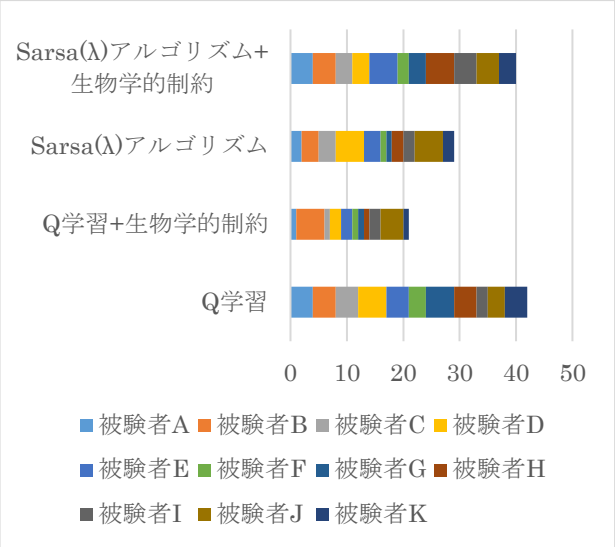

図 2　最終実験の評価結果

表 2　最終実験の評価理由

	人間らしい振る舞い	機械的な振る舞い
Q 学習のみ	攻めの姿勢が見られた 前ジャンプ攻撃が多く 初心者のように見えた	
Q 学習+生物学的制約		相手の跳びに対する反応が速すぎる ずっと同じ動き
Sarsa(λ)アルゴリズムのみ	距離を取るところが人間らしい	相手の跳びに対する反応が速すぎる
Sarsa(λ)アルゴリズム+生物学的制約	ジャンプ，遠距離攻撃，回避といった動きのパターンが増えた	

8. 議論

図 2 より，本研究で提案した Sarsa(λ)アルゴリズムに生物学的制約を導入したゲーム AI は，Q 学習に生物学的制約を導入した AI と Sarsa(λ)アルゴリズムのみの AI に比べ，優れた性能を発揮したが，Q 学習のみを実装した AI には劣っている結果となった．この結果と自由記述による評価理由を基に提案した手法の有効性を考察する．

まず，Q 学習のみを実装した AI に対する評価理由から，この AI が最も評価されたのは初心者らしく，最も攻めの姿勢を取っていることが原因であることがわかった．これは生物学的制約がないことや，Q 学習が期待値のみを考慮して学習していることにより，とにかくボタンを連打しながら前に出るという姿勢となり，それが生物学的制約の想定に反して高く評価されていることが原因だと考えられる．

次に，Q 学習に生物学的制約を導入した AI に対する評価理由から，相手の攻撃に対する反応の速さが評価に悪影響をもたらしていることがわかった．この原因としては，対戦 AI の行動が大きく関わっていると考えた．これは本研究で使用した対戦 AI が特定の攻撃パターンで前に出てくる傾向があったため，AI が連続で同じ迎撃行動を行うことを最適解にしたと考える．

また，Sarsa(λ)アルゴリズムのみの AI も相手の攻撃に対する反応の速さが原因で低い評価となった．しかし，この AI には後退によって相手の攻撃を回避する場面も見られ，その点で Q 学習に生物学的制約を導入した AI より高い評価を獲得出来た．これは Sarsa(λ)アルゴリズムの適格度トレースによって，報酬には直接影響しない後退という行為に報酬が与えられているからと考える．

最後に，Sarsa(λ)アルゴリズムに生物学的制約を導入した AI に対する評価理由から，この AI が他の 3 つの AI と比べ，最も攻撃パターンの多い AI として高い評価を得られたことがわかった．これは Sarsa(λ)アルゴリズムのみの AI と同様に，一手では報酬に直結しない行動が評価されたことと，生物学的制約による訓練と挑戦のバランスにより，多彩な攻撃方法を学べたからと考える．

これらのことから，本研究で提案した手法は既存の手法を用いて作成した AI と比べて，多彩な行動を学習出来，その点で高い評価を得られたと考える．しかし，人間は保守的な動きに比べ，攻撃的な動きをしている AI に人間らしい振る舞いを感じる傾向があると考えられるため，本研究で提案した手法により攻撃性を持たせることで，プレイヤのゲーム体験を向上させる人間らしいゲーム AI を獲得出来ると考える．そのためには，状態区分や報酬式における正の報酬と負の報酬のバランス，学習に用いる対戦相手を再検討することが必要であると考える．

9. おわりに

本研究では，Sarsa(λ)アルゴリズムに藤井らの生物学的制約を導入することで，従来の Q 学習ベースのゲーム AI と比べてより人間らしい振る舞いを獲得する手法の提案を目的として実験を行った．実験結果としては，Q 学習による AI が最も人間らしいと評価された．これは，この AI が期待値のみを考慮して行動を評価する Q 学習を採用していることや，生物学的制約がないことで攻撃的な姿勢となっているのに対し，他の 3 つの AI は対戦相手が攻撃的すぎたことで Sarsa(λ)アルゴリズムの学習による現実的な行動が保守的になりすぎてしまったことや，生物学的制約による行動の制限があることにより，自分から攻撃を仕掛けるという行動が少なかったことが原因であると考える．しかし，被験者からは攻撃の多彩性という点で，Sarsa(λ)アルゴリズムに生物学的制約を導入した AI が最も高い評価を得られた．これは Sarsa(λ)アルゴリズムの適格度トレースによって，一手では報酬に直結しない行動が評価されたことや，生物学的制約の訓練と挑戦のバランスにより，多様なパターンを試行錯誤出来たことがこの評価につながったと考える．そのため，本研究で提案した手法は，AI に多彩な行動を学ばせるという点で有効であると考える．

今後の課題として，Sarsa(λ)アルゴリズムに生物学的制約を導入した AI に，より人間らしい振る舞いを獲得させる．そのために，AI が攻撃的になるような状態区分や報酬式，対戦相手を再検討する必要があると考える．

文　献

[1] F. G. Glavin and M. G. Madden, "DRE-Bot: A Hierarchical First Person Shooter Bot Using Multiple Sarsa(λ) Reinforcement Learners," *Proc. 17th International Conference on Computer Games,* pp. 148-152, 2012.

[2] 藤井叙人, 佐藤祐一, 若間弘典, 風井浩志, 片寄晴弘, "生物学的制約の導入によるビデオゲームエージェントの「人間らしい」振る舞いの自動獲得," 情報処理学会論文誌, vol. 55, no. 7, pp. 1655-1664, 2014.

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

[4] INTELLIGENT COMPUTER ENTERTAINMENT LAB., RITSUMEIKAN UNIVERSITY, “Welcome to Fighting Game AI Competition, ” [Online]. Available: https://www.ice.ci.ritsumei.ac.jp/~ftgaic/. [Accessed 25 1 2026].