

SDNN を用いた強化学習による格闘ゲームの動的難易度調整

Dynamic Difficulty Adjustment in Fighting Games by SDNN-based Reinforcement Learning

手塚 大貴[†] 吉田 健人[†] 山根 健[†]
Taiki Tetsuka Taketo Yoshida Ken Yamane

1. はじめに

格闘ゲームにおいて、人と対戦する自律的なキャラクタを動かす AI は重要な要素である[1]. 人を楽しませるためには、人の多様な戦略や熟練度に対して AI の行動を素早く適切に調整する必要がある[2, 3]. これに関して、我々は選択的不感化ニューラルネット (以下、SDNN) [4-6] を価値関数近似器として強化学習を行う AI を提案してきた[7, 8]. 全ての情報をパターンコーディングし、選択的不感化法を用いて情報統合するため、パターンに基づく強力な類推能力により学習効率が低い。しかし、複数の対戦相手への適応能力や難易度調整などの性能については不明であった。

そこで本研究では、複数のサンプル AI を用いて提案 AI を事前学習した上で、適度に弱体化させることで、未知の対戦相手に対して強さを調整可能であるか検証する。

2. SDNN を価値関数近似器として用いた強化学習

提案 AI を図 1 に示す。AI は対戦相手を含む環境と実時間で相互作用する。格闘ゲームのプラットフォームとして FightingICE (<https://www.ice.ci.ritsumei.ac.jp/~ftgaic/>) を、強化学習アルゴリズムとして Q 学習を用いる。最大の特徴は行動価値関数近似器として SDNN[4-6]を用いる点である。

2.1 複数の行動や対戦相手に対応する方法

多数の行動を扱えるように、SDNN の出力層にある n 個の素子から行動毎に予め決められた m 個を選択して行動価値を表現する[7, 8]. この方法は分解能や計算効率の点で利点がある。また、行動間の類似性に基づき行動情報をコード化すれば、行動方向でも SDNN の類推能力を活用できる。

さらに、戦略の異なる複数の対戦相手に対応するため、対戦相手の情報を中間層の素子に選択的不感化法を用いて修飾する。先行研究[8]では、対戦相手の情報を表すコードパターンをランダムに生成したため、SDNN の類推能力を効果的に活用できなかった。そこで本研究では、対戦相手の AI 毎に生成したランダムパターンと相手の行動頻度に基づいて生成したパターンの両方を修飾に用いる。

2.2 強化学習 AI の構築

状態 s_t については、対戦相手との距離、HP、エネルギー残量など 11 変数を用いて構成する。行動 a_t は、移動系、防御系、攻撃系行動から 27 種類とする。行動選択については、 ϵ -Greedy 法を用いて、ランダム行動率 ϵ を勝率に応じて決定する。報酬 r_{t+1} は、行動実行後から次の行動を受け付けるまでの間の (自分の HP の変化量) - (相手の HP の変化量) $+\delta$ とする。 δ については、一定の距離以上であるなど、予め設定した条件で小さな負の報酬を与える。

Q 値の更新に伴い、 r_{t+1} が得られた時刻 $t+1$ に、SDNN の学習を 1 回だけ行う。つまり、ゲームの進行と共に逐次的

[†] 帝京大学 理工学部 Faculty of Science and Engineering, Teikyo University

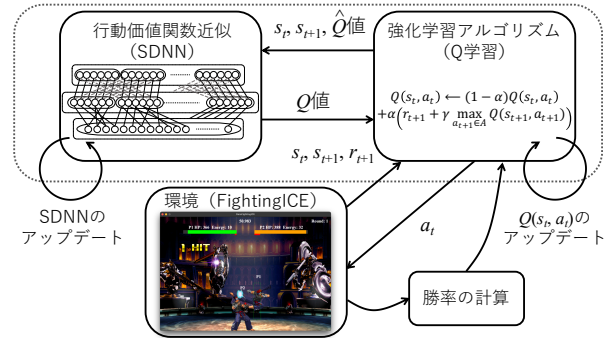


図 1 提案する格闘ゲーム AI の構成

に学習を行い、モデル (シミュレータなど) を使った学習や保持している過去のデータを使った学習 (Experience Replay など) は行わない。なお、1 ステップの処理時間は可変であり、強化学習の学習率 $\alpha = 0.1$ 、割引率 $\gamma = 0.9$ とする。詳細については、文献[8]を参照されたい。

2.3 動的難易度調整方法

未知 (未学習) の対戦相手に対しても強化学習と弱体化により強さを調整する。幾つか弱体化方法が考えられるが、本研究では単純な方法として、対戦相手との勝率が 0.5 以上かつ HP 差が 0 以上である場合に行動価値が中央値の行動を選択し、それ以外は行動価値が最大の行動を選択する。

勝率の計算において、対戦相手の HP と同じ HP でラウンドが終了した場合、提案 AI の敗北とする。また、通常は過去 100 ラウンドにおける勝率を計算するが、100 ラウンド未満の場合はそれまでの全ての結果から勝率を計算する。

3. 実験

10 体のサンプル AI を用いて事前学習した上で、未学習の対戦相手に対して強さを調整する能力について調べた。なお、本実験ではキャラクタとして ZEN のみを用いる。

3.1 事前学習

事前学習の対戦相手として MogakuMono, SimpleAI, MctsAi, Thunder, ReiwaThunder, Toothless, HaibuAI, JayBot_GM, BCP, TOVOR の 10 体の AI を用いて、この順番で対戦した。なお、ラウンド終了毎に対戦相手の AI との勝率を計算し、勝率が 0.5 未満の場合、再び最初の AI である MogakuMono からやり直して対戦した。この方法は、順番に対戦を繰り返すよりも効率的に学習が進む。

事前学習で MogakuMono と対戦したラウンド数は 5 千回以上、SimpleAI とは 3 千回以上であり、順番が早い AI との回数は増加した。それ以外の AI とは 2 千回未満であった。

学習過程における勝率、完全勝率、完全敗北率を図 2 の上パネルに示す。なお、“完全”とは、全てのラウンドに対して、どちらかの HP が 0 になり勝負が決したラウンドの割合を示す。また、各 AI との勝率を図 2 の中央パネルに

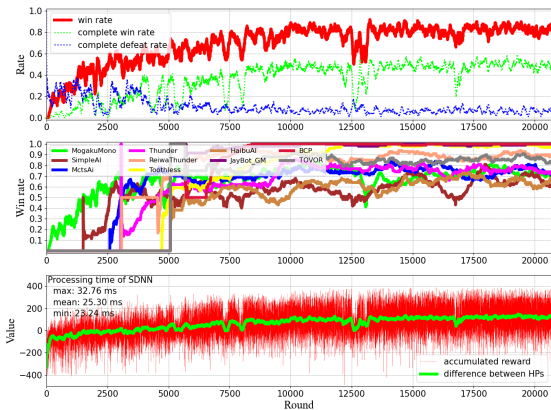


図 2 事前学習における学習過程(100 ラウンド平均)

示す。さらに、ラウンド終了時の累積報酬、HP 差の移動平均、SDNN 部分の 1 ステップの処理時間を、図 2 の下パネルに示す(なお、ゲームの 1 フレームは 17 ms 程度)。合計 2 万ラウンド以上の十分な学習により、全 AI に対する勝率が 0.5 以上に到達して、その後も維持された。

3.2 未学習 AI に対する動的難易度調整

未学習 AI である Dora, FalzAI, KotlinTestAgent, DiceAI, MctsAi23i, KeyBoard, LGIST_Bot の 7 体と個別に対戦して、強さを調整する能力を調べた。なお、提案 AI は、事前学習と同様に常に強化学習していることに注意されたい。

実験の結果、6 体の AI に対しては、素早く強さが調整され、勝率は 0.5 程度を維持した。図 3 に Dora, 図 4 に FalzAI との対戦結果を示す。いずれも直ぐに強さが拮抗した状態になり、ラウンド終了時の HP 差が 0 に近づいた。結果として、数ラウンドで勝率が 0.5 程度に調整されている。一方で、LGIST_Bot に対しては、2 千ラウンド以上対戦しても勝率が 0.1~0.2 程度にしか上がらなかった。

3.3 人プレイヤーとの対戦

人プレイヤー(学部生 1 名, 教員 1 名)と初期状態から提案 AI を対戦させた結果、200 ラウンド程度で AI の勝率が 0.2 程度に達した。一方で、学習済みの提案 AI を用いた場合、勝率は高くても 0.3 程度にしか上がらなかった。

初心者が勝てないサンプル AI (MctsAi, ReiwaThunder など)も多いことから、人、提案 AI, サンプル AI で三棘みのような結果が得られていることが興味深い。その理由として、提案 AI の事前学習において多様性が不足している可能性や、未だ明らかではない人の特性がある可能性などが考えられる。例えば、人はうまく行かないと感じると直ぐに戦略を変更しているようであり、ラウンド内での急激な戦略の変更提案 AI が対応できていない可能性がある。

4. おわりに

本研究では、SDNN を用いることで多数の行動や多様な戦略をもつ対戦相手に対して素早く行動価値関数を近似できる利点を活かし、提案 AI を適度に価値の低い行動を選択させて弱体化させることで、動的に強さを調整可能か調べた。その結果、多くの AI に対して調整可能であり、提案方法に可能性があることが分かった。

詳しい分析として、行動頻度に基づいてサンプル AI の行動パターンを定義し、各 AI の行動パターン間の方向余弦(類似度)を計算した。表 1 に、未学習 AI に類似する学習

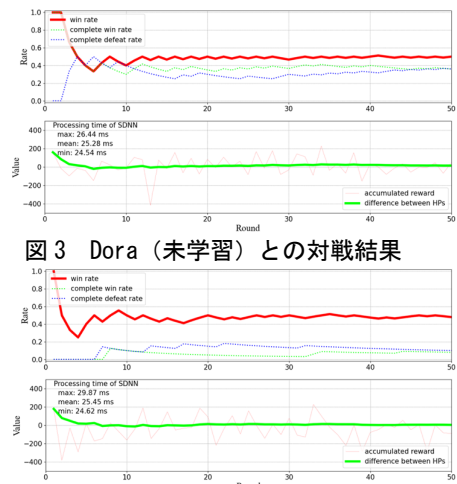


図 3 Dora (未学習) との対戦結果

図 4 FalzAI (未学習) との対戦結果

表 1 未学習 AI と学習済み AI の行動パターンの類似性

未学習 AI 名	類似 AI (方向余弦 0.6 以上)
Dora	Thunder, ReiwaThunder, Toothless
FalzAI	MctsAi
KotlinTestAgent	Thunder, ReiwaThunder
DiceAI	SimpleAI, HaibuAI, TOVOR
MctsAi23i	TOVOR
KeyBoard	ReiwaThunder, MogakuMono
LGIST_Bot	Thunder

済み AI を示す(閾値 0.6)。類似した行動パターンを取る AI が事前学習に含まれたため、SDNN の強力な類推能力を利用して未知の対戦相手にも適応できたと考えられる。

一方で、強い AI の一つである LGIST_Bot は行動選択が速く、行動が洗練されているため、多くのラウンドを重ねなければ勝つことが難しい。最終的に人を楽しませる AI の開発を考えるならば、人でも全く勝てない超人的 AI を対象に含めるか否かは議論の余地がある。

今後の課題として、情報のコード化方法を改善すると共に、人に対して動的難易度調整する方法について検討する。

参考文献

- [1] 石原誠ら, “対戦格闘ゲームにおけるゲーム AI や操作方法の違いがプレイヤーの感じる面白さに与える影響の分析”, 情報処理学会論文誌, Vol.57, No.11, pp.2414-2425 (2016).
- [2] 邓士達ら, “動的な難易度調整により対戦して楽しい格闘ゲーム AI”, 第 25 回ゲームプログラミングワークショップ 2020 論文集, pp.58-61 (2020).
- [3] 澤野圭太ら, “遺伝的アルゴリズムとモンテカルロ木探索を用いた動的難易度調整を行う格闘ゲーム AI”, 第 28 回ゲームプログラミングワークショップ 2023 論文集, pp.38-41 (2023).
- [4] 森田昌彦ら, “選択的不感化法を適用した層状ニューラルネットの情報統合能力”, 電子情報通信学会 D-II, Vol.J87-D-II, No.12, pp.2242-2252 (2004).
- [5] 新保智之ら, “選択的不感化ニューラルネットを用いた強化学習の価値関数近似”, 電子情報通信学会 D, Vol.J93-D, No.6, pp.837-847 (2010).
- [6] 小林高彰ら, “選択的不感化ニューラルネットを用いた連続状態行動空間における Q 学習”, 電子情報通信学会誌 D, Vol.J98-D, No.2, pp.287-299 (2015).
- [7] 小川拓実ら, “強化学習の価値関数近似器として SDNN を用いた格闘ゲーム AI”, 情報処理学会第 85 回全国大会講演論文集, 第 2 分冊, pp.149-150 (2023).
- [8] 小川拓実ら, “対戦型格闘ゲームにおける多様な対戦相手に対応する強化学習エージェント”, 第 22 回情報科学技術フォーラム講演論文集 (FIT2023), 第 2 分冊, pp.533-534 (2023).