F-057

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
Reinforcement Learning Agent for Handling Diverse Opponents in Fighting Games

小川 拓実† 阿久津 光範† 岸 愛斗‡ 山根 健†‡
Takumi Ogawa Mitsunori Akutsu Manato Kishi Ken Yamane

1. はじめに

対戦型格闘ゲームにおいて人プレイヤと対戦する自律的なキャラクタ（以下，エージェント）は人プレイヤを楽しませるための重要な要素である[1, 2]．その設計方法として，ルールベースの手法は強さに限界があるだけでなく，柔軟な強さの調整が難しい．これに対して，深層強化学習方法はプロレベルの強いエージェントを作ることができるが，学習効率が悪いなどの問題がある[3]．また，モンテカルロ木探索を利用した行動最適化手法も提案されているが，リアルタイムでゲームが進行する場合に探索時間などが課題となる．取り得る状態，行動，作戦などの組み合わせが爆発的に多い戦略的な格闘ゲームにおいて，対戦の中で相手に応じて素早く行動を最適化する方法が求められる．

これに関して，価値関数近似器として選択的不感化ニューラルネット（以下，SDNN[4]）を用いた強化学習は少ない学習サンプルでも効率的に学習できるなどの利点から大きな可能性がある[5, 6]．そこで，我々は，ゲーム AI への応用を目指して，複数の行動を扱えるように方法を拡張した強化学習エージェントを提案した[7]．提案手法は学習効率が高く， 1 対 1 の対戦において強いエージェントにも素早く対応できることを確認した．しかし，複数の戦略や複数の相手への対応能力には課題が残されていた．

そこで，本研究では，さらに複数の対戦相手に対応できるように拡張して，多様な対戦相手に適応する能力や学習効率について評価を行う．

2. 強化学習を用いたエージェントの構築

提案エージェントを図 1 に示す．エージェントは環境とリアルタイムに相互作用して Q 学習する．行動価値関数近似として SDNN を用いることが特徴である．状態 s_t を構成する各入力変数 x が入力層において分散表現され，中間層で選択的不感化法[4]により情報が統合される．これにより，分散表現に基づく強力な類推能力を利用できる．また，追加的な学習でも干渉が少ないため，過去のデータをランダムに取り出して学習する Experience Replay などは必要ない．

以降では，複数の行動や複数の対戦相手に対応する方法について説明する．関数近似の詳細については文献[4-6]を参照されたい．

2.1 複数の行動に対応する方法

複数の行動に対する価値を一度に計算するために，出力層の n 個の素子から行動毎に m 個の素子を選択して行動価値を表現する（ただし，n>m）[7]．具体的には，それぞれの素子について選択する場合 1，しない場合-1 とし，±1 を並べた n 次元 2 値パターンとして行動情報を表す．つまり，

† 帝京大学理工学研究科　Graduate School of Science and Engineering, Teikyo University
‡ 帝京大学理工学部　Faculty of Science and Engineering, Teikyo University

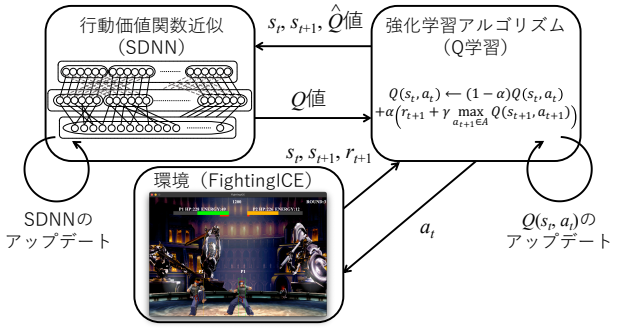

図 1　提案する強化学習エージェントの構成

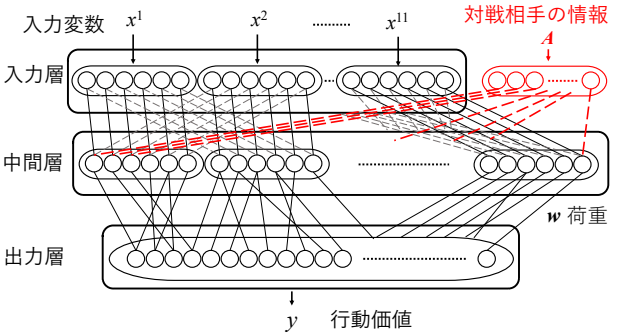

図 2　複数の行動および対戦相手に対応する方法

m 個の素子を選択するパターンが行動情報を表す．本方法により，行動毎に異なる価値を表現することができる．また，Q 値の分解能や計算量の点からも効率が良い．

さらに，選択される素子の一部は行動間で重なるため，コード化方法を工夫することで行動方向でも分散表現に基づく類推を利用できる．そこで，次の方法を用いた．まず，扱う行動について，行動タイプ，コマンド長，与ダメージなど 10 項目を考えて，それぞれの項目において 4 つのクラスを設定する．そして，±1 を成分にもつ μ 次元 2 値パターンをそれぞれのクラスに割り当てる．ただし，n=10μ かつ成分 1 の数を n/40 個とする．これら 10 項目の組み合わせを考えて，すべてのパターンの結合により得られる n 次元 2 値パターンとして行動情報を表現する．なお，実験で扱う行動は数十種類であり，計算上は 4^{10} 乗の組み合わせの行動を扱える．また，このコード化方法では n=4m となる．

2.2 複数の対戦相手に対応する方法

複数の対戦相手に対応するため，図 2 に示すように対戦相手の情報 A を中間層の素子に選択的不感化法[4]を用いて修飾する．この時，既に中間層の表現は，入力変数の組み合わせの数だけ選択的不感化された表現が並んでいるため，重ねて選択的不感化される．なお，対戦相手を表すパターンとしてランダムに生成した高次元 2 値パターンを用いる．

ラウンド内では同一の相手と対戦するため，常に同じパターンで修飾される．従って，対戦相手が増えても，必要な計算時間やメモリは増加しない．

表 1　SDNN のハイパーパラメータ

入力層の素子数	11×100
中間層の素子数	11×10×100
出力層の素子数（n）	4000
行動を表す素子数（m）	1000

2.3 強化学習エージェントの構築

状態 s_t については，両者の距離や HP，エネルギー残量など 11 変数とする．これらの変数は人によって恣意的に選択されるが，効果的だと思われる変数を次々と追加しても必要な計算時間やメモリが爆発的には増加しない．

行動 a_t に関しては，移動系 8 種類，防御系 2 種類，攻撃系 17 種類，合計 27 種類の行動を扱う．なお，ジャンプが伴う空中での行動については明示的には扱わない．行動選択方法として，ε-Greedy 法を採用する．ランダム行動率を ε として，過去 100 ラウンドの勝率に応じて設定し，勝率が低い場合には 0.1，高い場合には段々と 0.0 にする．

報酬 r_{t+1} については，行動選択したフレームから，行動が実行あるいはキャンセルされて次の行動を受け付け可能になるフレームまでの HP の変化量に注目して，**報酬＝（自分の HP の変化量）－（相手の HP の変化量）＋δ** として，次の行動を選択する直前に与える．なお，δ については，予め設定した状況において小さな報酬を入れる．また，勝敗が決した時にはその時点での HP の差を δ に足し込む．

Q 学習では，r_{t+1} が得られた時刻 t+1 に，SDNN において s_t を入力変数として，a_t および A を用いて，新しい Q 値を教師信号として 1 回だけ誤り訂正学習を行う．なお，時間ステップは可変であり，学習率 α=0.1，割引率 γ=0.9 とする．

3. 実験

提案方法の性能評価のため，格闘ゲーム FightingICE においてサンプルエージェント 6 体と 1 ラウンド（最大 60 秒）ずつ順番に対戦を繰り返した．用いた SDNN のハイパーパラメータを表 1 に示す．また，学習係数 c=0.01 とする．

我々が調べた限りでは，エージェント間で強い順に ReiwaThunder＞Dora, Toothless＞JayBot_GM＞BCP＞TOVOR である．また，初心者のプレイヤは BCP に勝ち越し，TOVOR，Toothless と同程度の強さ，その他に対して歯が立たなかった．なお，我々が試作した単純で弱いルールベースのエージェントは Fighting Game AI Competition（2019）において優勝した ReiwaThunder に必勝した．また，中程度の強さのエージェント同士にはジャンケンのような強さの構造があった．つまり，絶対的に強いエージェントはおらず，戦略の相性がある．つまり，本タスクは戦略的な要素が大きいと言える．

結果を図 3 に示す．図の横軸にラウンド数，縦軸に各ラウンドにおける累積報酬と過去 100 ラウンドの平均累積報酬を示す．ラウンドを重ねるとほぼすべての相手に対して累積報酬が正方向へ急激に増加した．

また，過去 100 ラウンドの平均勝率を図 4 に示す．すべての相手に対して勝率が素早く上がっている．さらに，複数の相手と順番に学習することで性能が大きく悪化することはなかった．むしろ，一部の相手に対しては 1 対 1 で学習する場合と比較して効率よく学習できていた．

この結果から，多様な相手と順番に対戦すると初心者程度の弱い相手に対して素早く対応できるだけでなく，強い相手にも対応できることがわかった．また，ある相手で学習した知識が別の相手でも利用できる可能性が示唆された．

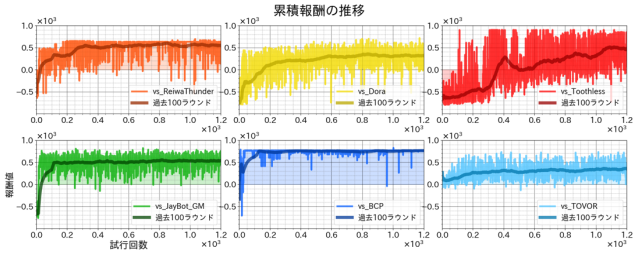

図 3 ラウンド終了時の累積報酬

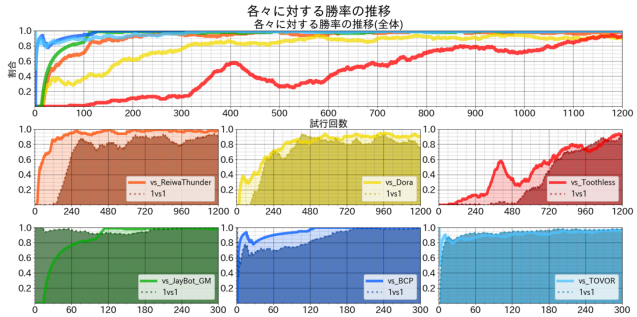

図 4 学習過程における勝率の推移（100 ラウンド平均）

4. おわりに

本研究では，多様な対戦相手に対応する強化学習エージェントを構築して，素早く学習できることを示した．

格闘ゲームにおいて必勝の固定戦略が存在しないと仮定すると，相手の戦略に素早く適応することが重要になる．提案方法は，複数相手に事前学習した上で用いることで，人プレイヤと対戦する中で素早く適応できる可能性があり，動的に強さを自動調節する仕組みに繋がると期待される．

今後の課題として，新しい対戦相手への適応能力や人プレイヤとの対戦の中で性能を調べることなどが挙げられる．

謝辞

本研究は，科学技術融合振興財団補助金助成（調査研究期間 2023～2024 年度）の助成を受けた．

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] 邓士达ら，“動的な難易度調整により対戦して楽しい格闘ゲーム AI，” 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] I. Oh et al., “Creating Pro-Level AI for a Real-Time Fighting Game Using Deep Reinforcement Learning,” IEEE Transactions on Games, vol.14, no.2, pp.212-220, 2022.

[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，” 信学誌 D, 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，” 信学誌 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，” 信学誌 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.