7P-08

横スクロールアクションゲームの強化学習による攻略

菅野　明日風†　　秋岡　明香‡
明治大学　総合数理学部　ネットワークデザイン学科

1.　はじめに

本稿では、「操作の難しいゲームを、強化学習を利用してスムーズなプレイの再現をする」ことを目指す。「OpenAI」[1]が提供する開発フレームワーク「OpenAI Gym Retro」[2]、「OpenAI Baselines」[3]を活用して『Sonic The Hedgehog』を対象とした強化学習を行っていく。

『Sonic The Hedgehog』はセガ・エンタープライゼスが 1991 年にメガドライブ用に発売したゲームである。ジャンルは横スクロールアクションであり、左から右に向かってキャラクターを操作し特定座標にあるゴールを目指す。ひとつのステージにつき３面構成となっている。操作キャラクターであるソニックを左右移動、ジャンプの基本動作で操作する。移動しながらのジャンプは敵性キャラクターへの攻撃判定を持ったジャンプになる。またソニックには「緩急」の概念があり、高速な状態であれば坂を駆け上がることができるようになるといった変化が現れる。よってステージ攻略のためには高速な状態のほうが有利なことが多いが、その速さ故に敵キャラクターの存在に気づくのが遅れる、ギミックに反応できなくなるということも往々にしてある。更にステージを進めていくと１撃でミスになってしまうギミックが登場するなど、総じて熟練者ではない人間がプレイするには操作が難しいゲームであると言える。

2.　関連研究

ソニックを題材とした研究で、エージェントが将来の環境の変化に適応できるようにする（汎化という）ことを目的としたものがある。OpenAI が汎化の訓練のために『Sonic The Hedgehog』シリーズ数十ステージを対象とした機械学習の研究である「OpenAI Retro Contest」[5]を開催した。本研究では、コンテスト用に公開されたプログラムを基に報酬の調整等を行う。

3.　提案手法

強化学習において、行動する主体を「エージェント」、エージェントがいる世界を「環境」と呼ぶ。本研究では、エージェントはプレイヤーが操作するキャラクターである「ソニック」である。エージェントの行動による環境の変化を通して、変化に応じた報酬をエージェントが獲得する。このサイクル（ステップという）の繰り返しによって学習する。強化学習の訓練１回分のことを「１エピソード」と言い、１エピソードで獲得する報酬を最大化する、つまり収益を最大化することが基本的な目標である。

本研究では Proximal Policy Optimization（PPO）[4]を強化学習アルゴリズムとして採用する。このアルゴリズムは高い報酬が得られる行動を優先し、低い報酬しか得られない行動を避けるようにして次の行動を最適化する学習法である。学習に必要なステップ数が多いが、時間をかければ優れた結果が得られる。PPO は OpenAI の標準アルゴリズムとなっている。本研究ではこれまで、メガドライブ版『Sonic The Hedgehog』第１ステージ「Green Hill Zone」１面と２面の強化

Beating Side-Scrolling action game Using reinforcement learning
†Asukaze Sugano, Meiji University
‡Sayaka Akioka, Meiji University

学習による攻略を行う。学習の流れを以下に示す。

・各ステージ、ゴールが設定されている特定座標にエージェントが到着した際に学習終了とする。
・エピソード完了の条件は、ソニックの残基数に関わらず１ミスした時、またはエージェントがゴールの座標に到着することである。
・各ステージ、そのステージを攻略するために報酬を調節する。

4. 結果・考察

Green Hill Zone 第１面は速度を利用したループが中盤にあるだけのギミックであり、シンプルな横スクロールである。よって単純に右に進む、つまり x 座標の値を大きくすることに報酬を設定した。（報酬）＝（現在の x 座標）-（1 ステップ前の座標）と設定し学習させた。結果としてはゴールの座標に到達できなかった。学習済みモデルを再生すると、助走せず高速な状態でなかったためループを抜けられていなかった。上記の報酬プログラムではエージェントが左に進んだ場合、負の報酬を得てしまうため、引き返して助走するという選択を取らない。

そこで（報酬）＝（現在の x 座標）-（1 ステップ前までの x 座標の最大値）とすれば、現在地より左に進んでも負の報酬を得ない。これにより高速な状態でループに突入できると考えられる。再度学習させた結果、今度はループを抜け、そのままゴールの座標まで到達し、クリアできた。初回の学習とは違い、助走が足りずにループから弾かれてしまった場合も、報酬の影響でかなり後方まで戻ってから再度助走をつけてループ突破に成功しているパターンもあった。この結果から、当ゲームにおいてこの報酬座標が適していると判断し、今後のステージにも活用できると考えた。

第２面は上下ルートに分かれており、上ルートに比べ下ルートにはダメージギミック等が存在するため難度が高くなっている。第１面と同様の報酬を設定し学習させたところソニックが下ルートに落ちてしまい途中でダメージ、タイムアップによるミスでクリアできない状態になってしまった。改善案として、ステージに配置されている「リング」というアイテムに注目した。リングは模範ルートを描くように置かれている事が多く、これをなぞることで上ルートを走行することができるのではないかと考えた。リング獲得を重く見て、獲得できる報酬を多めに設定し学習させたところ、上ルートを通り、初回学習で詰まったところを突破した。その後詰まることはなく、ゴール座標に到達することができた。

5. おわりに

本稿では、強化学習を用いた操作の難しいゲームの模範的攻略を目指した研究を行った。本稿で取り上げた２つのステージは攻略済みだが、今後のステージは右から左に進む要素の出現、即死ギミックの追加など学習が難化する要素が多く出現する。報酬プログラムのさらなる追加などを視野に入れ、多くのステージの攻略ができる学習プログラムの完成を目指す。

参考文献

[1]OpenAI, https://openai.com/（参照：2020-12-30）
[2]prabhatnagarajan, openai, GitHub repository, https://github.com/openai/retro（参照：2020-12-30）
[3] harry uglow, openai, GitHub repository, https://github.com/openai/baselines（参照：2020-12-30）
[4]John Schulman; Oleg Klimov; Filip Wolski, Prafulla Dhariwal; Alec Radford; 2017, Proximal Policy Optimization, https://openai.com/blog/openai-baselines-ppo/（参照：2021-01-06）
[5]Christopher Hesse; John Schulman; Vicki Pfau; Alex Nichol; Oleg Klimov; Larissa Schiavo; 2018, Retro Contest, https://openai.com/blog/retro-contest/（参照：2021-01-06）