自作カードゲーム環境への深層強化学習手法の適用

創発ソフトウェア研究グループ
B3　西村 昭賢

発表の流れ

- はじめに
- 要素技術

発表の流れ

- はじめに
- 要素技術

深層強化学習による問題解決

- 自動運転
- ロボットの制御

⇒様々な実世界の問題に応用されている

ゲームへの応用

- 完全情報ゲーム(囲碁,将棋)への応用
⇒プロを圧倒(AlphaZero)

- 不完全情報ゲーム(ポーカー,麻雀)への応用
⇒現在注目されている

本研究の目的

- 自作カードゲーム(不完全情報ゲーム)を作成

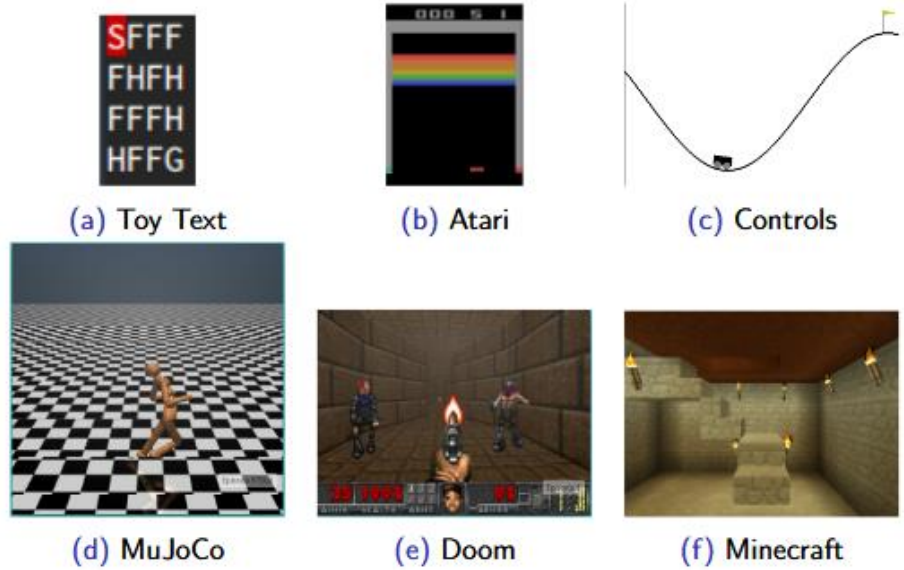
⇒ 自作の環境で強化学習を行えるか確認

発表の流れ

- はじめに
- 要素技術

OpenAI Gym

- OpenAI社が提供する強化学習用シミュレーションライブラリ
- 様々な学習環境が提供されている



- 自作の学習環境をgymのインターフェースを用いて定義が可能

画像出典: https://ai-mrkogao.github.io/reinforcement%20learning/openaigymtutorial/

Q学習

- 代表的な価値ベースの強化学習手法の1つ
- Q値を以下の式に従って1ステップごとに更新していく

TD誤差

$$\mathrm{Q}(s_t, a_t) \leftarrow \mathrm{Q}(s_t, a_t) + \alpha\{r_{t+1} + max_{a_{t+1}}\mathrm{Q}(s_{t+1}, a_{t+1}) - \mathrm{Q}(s_t, a_t)\}$$

α：学習率(Q値の更新の度合い) γ：割引率(将来の価値の割引度合い)

Deep Q Network (DQN)

- Q学習では状態や行動の次元数が増えると現実的に計算ができなくなる

⇒深層学習を用いることで学習可能に

- Experience Replay や Fixed Target Networkにより
  安定した学習が安定

モンテカルロ探索

- 価値ベースの強化学習手法の1つ

- Q値の更新を1エピソード終了後に以下の式に従って行う

エピソード中のステップ$t = 0 \sim T - 1$について

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\{G_t - Q(s_t, a_t)\}$$
$$G_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \ \ldots + \gamma^{T-t-1} r_T$$