一人称視点シューティングゲームにおける深層強化学習

研究背景

・深層強化学習によるAIの学習の進歩
e.g.)ゲームプレイ，ロボット制御，自動運転

＜ビデオゲームのプレイングを学習するAI＞
ゲーム画面を入力とした深層学習の出力に強化学習の価値関数等を近似する.

最適な行動への報酬信号が，頻繁かつ容易に獲得できる環境では人間のプロに匹敵するプレイングを学習することができる.（Atari，3D迷路 etc.）

↓

報酬信号がスパースな環境下では，深層強化学習を適用するだけでは学習がうまく進まない.

↓

報酬が不十分な環境での学習は，Reward Shaping（追加報酬），Curriculum Learningなどで克服できる.

↓

しかしこれらの手法は，複雑な環境になるほど実装側の事前知識や時間消費が多く必要なこと実装が困難であるという欠点がある.

研究目的

報酬がスパースな環境における深層強化学習で，追加報酬やCurriculumの設定を自動化する

実験環境

＜ViZDoom＞
一人称視点シューティングゲーム（FPS）「Doom」をベースにした視覚情報から学習を行う研究用のフレームワーク.

報酬がスパースな環境や，問題設定が複雑な環境など様々なシナリオが実装されていて，カスタマイズも可能

Basic シナリオ

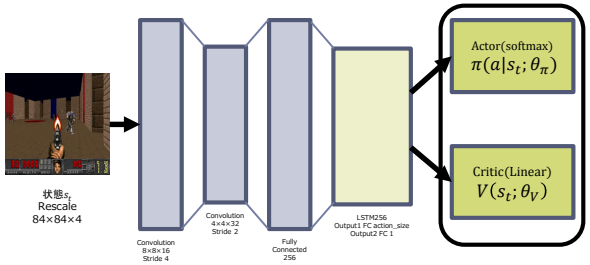
Deathmatch シナリオ

一人称視点シューティングゲームにおける深層強化学習

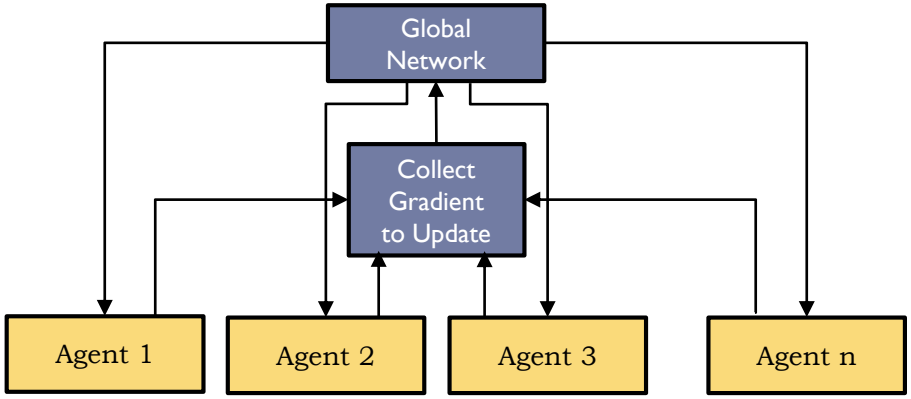
利用手法①

・Advantage Actor-Critic(A2C)

Actor-Criticを利用した深層強化学習手法の一つ
Actorと呼ばれる方策器$\pi(a|s_t;\theta_\pi)$とCriticと呼ばれる状態評価器$V(s_t;\theta_V)$を同時に最適化する



複数スレッドでエージェントを並列に実行し，それぞれの環境で得られた学習結果のシーケンスから勾配を計算し，グローバルネットワークを更新する



利用手法②

・Rarity of Events(RoE)

事前に定義したイベント（アイテムを拾う，敵を攻撃する etc.）を用いて，AIに頻繁に経験するイベントより希少なイベントを経験することに高い報酬を即時的に与えるような報酬関数を設定する．

$$R_t(\mathbf{x}) = \sum_{i=1}^{|\mathbf{x}|} x_i \frac{1}{\max(\mu_t(\epsilon_i), \tau)}$$

・ε_t：エピソード集合
・x_t：イベントの発生回数
・$\mu_t(\epsilon_i)$：エピソードの平均発生回数
・τ：下限値（$\tau = 100$）

イベントの設定という，環境に対する事前知識を最小限に留めることができる．

簡単なイベントから学習し，徐々に困難なイベントを探索するようになるので自動化したCurriculum Learningのような形式となる