DEEP Q NETWORK

金沢人工知能勉強会・交流会
(金沢工業大学工学部情報工学科４年)
上野友裕

Kanazawa AI Meetup

deep Q-networkとは？

- V. Mnih *et al.*, "Playing atari with deep reinforcement learning,“http://arxiv.org/pdf/1312.5602.pdf
- Google DeepMindがAtari 2600のゲームをdeep Q-networkにプレイさせたところ、人間のスコアを上回った。

事前知識

Q値とは？

- 報酬の予測値のこと
- ある行動をとる際に、現在から無限に未来までの報酬の和を表すもの。

∇(ナブラ)とは？

- 勾配のこと
- ∇Q_Θと書かれていたら、Q_Θで表せられるネットワークを学習する（勾配を降下する）という意味

$X \sim N(0,3)$の意味

- Xは平均0,分散3の確率分布に従う
- 「~」の記号の意味は、「○○の確率分布は…」　という意味

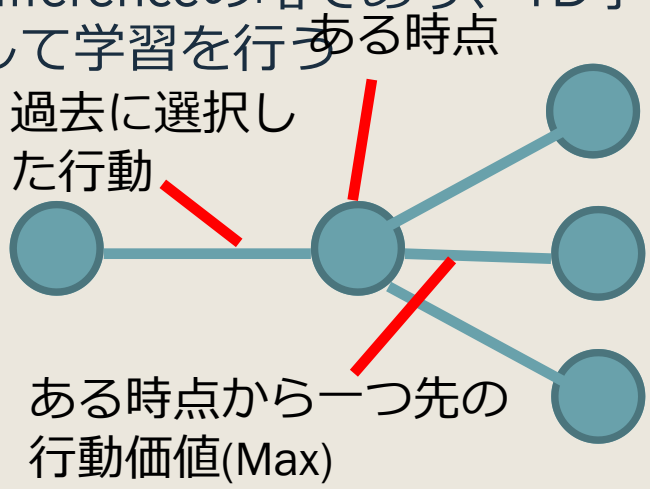
DQNのアルゴリズム

TD学習とは?

- ある時点から一つ先の行動価値の最大値と、ある時点から一つ前に観測された行動価値の差分に重みをつけて、ある時点から一つ前時点のニューラルネットからの推定価値(行動価値)に足しこみ、学習させる手法

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left(r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right)$$

- TDはTemporal Differenceの略であり、TD学習は時間的な価値の差分を考慮して学習を行う



追加するスライド

- イプシロングリーディー
- Deepの部分の説明
- フーバー損失（ロバストにできる）
- QT 型の損失

Experience Replayとは?

- エージェントが経験した全ての「状態、行動、報酬、行動を取った後の次の状態」を保存する
- 保存したデータの中からランダムにデータを取り出し、学習させる
- 以上の方法により学習が安定する
- マシンの制約などにより、メモリ上にデータが乗り切らない場合は保存する数を決めて、溢れた場合はランダムに古いデータを削除する

参考になるプログラム

- https://github.com/rlcode/reinforcement-learning（オススメ!）
- https://github.com/yukiB/keras-dqn-test
- https://github.com/icoxfog417/chainer_pong

参考文献

- **『DQNの生い立ち　＋　Deep Q-NetworkをChainerで書いた』**,https://qiita.com/UgoNama/items/08c6a5f6a571335972d5
- **『ゼロからDeepまで学ぶ強化学習』**,https://qiita.com/icoxfog417/items/242439ecd1a477ece312

KAIM