

時系列データを対象とする Deep Recurrent Q-Network への 学習時間軽減を目的とした Quasi-Recurrent Neural Network の導入

仁科 京介[†]藤田 茂[‡]千葉工業大学大学院情報科学研究科[†]千葉工業大学情報科学部[‡]

1 はじめに

強化学習に深層学習を用いた Deep Q-Network(DQN) は Atari 2600 のゲームにおいて 49 のゲームの内 29 のゲームで人間よりも高い成績を示している [1]. また, DQN に Long-Short Term Memory(LSTM) を導入することにより, 時系列データを扱うことができるため, 人間よりも最大 4024.9% 高い性能を示している [2][3].

しかし, LSTM は 1 つ前のタイムステップの情報をもとに計算するため, 学習に時間がかかる課題がある. この問題を解決した手法の一つに Quasi-Recurrent Neural Network(QRNN) が存在する [4]. QRNN とは層間を共通重みの局所結合で繋いだニューラルネットワークである Convolution Neural Network(CNN) を LSTM のように長い時系列データを処理できるようにしたニューラルネットワークである. これにより学習時間を最大 17 倍高速化し, LSTM と同等の性能を示した.

本研究では, DRQN の画像処理層を変更して QRNN を導入することにより, 学習時間を最大 11.78% 軽減した.

2 Deep Recurrent Q-Network(DRQN)

DRQN とは, DQN に LSTM を組み合わせることで過去の情報を扱えるようにした深層強化学習手法である [2]. 文献 [2] では, Atari2600 のゲームの 1 つである Pong の画面を 50% の確率で隠すことによって部分観測マルコフ決定過程にし, DRQN と DQN の性能を比較した. その結果, DRQN が DQN より 66.4% 高い性能を実現した.

3 Quasi-Recurrent Neural Network

QRNN とは, LSTM が持つ入力ゲート, 忘却ゲート, 出力ゲートの 3 つの計算を畳み込み層で並列処理を行う. その後, プーリング層で過去の履歴と現在の情報を集約することによって LSTM のように長い時系列パターンを学習することを可能にした CNN である. これにより, 学習時間が LSTM よりも最大 17 倍高速化し, 感情分類, 言語モデル, 機械翻訳のタスクにおいて LSTM と同等のスコアを出した.

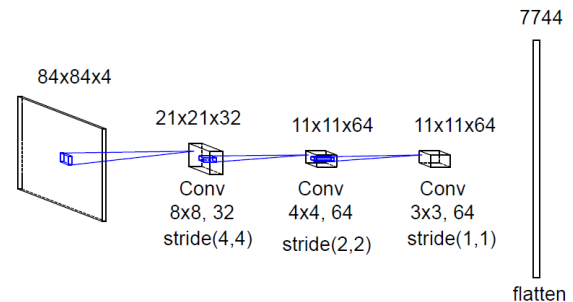


図1 DRQNの画像処理層

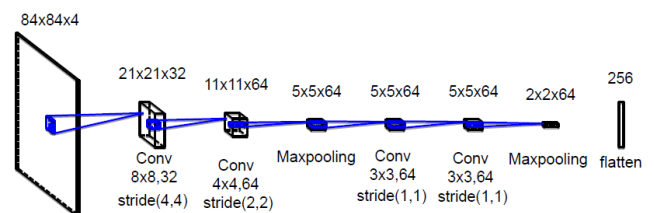


図2 本研究の画像処理層

4 提案手法

本研究では, DRQN で利用されていた LSTM を QRNN に変更することで過去の情報を扱いつつ, 学習時間の軽減を目的とした Q-Network を提案する.

QRNN に変更するにあたって 1 つ課題がある. DRQN では, 画像の処理に図 1 のような畳み込み層を 3 つ重ねたものを使用している. この画像処理層の出力次元数は 7744 である. QRNN のパラメータ数は窓幅を W , 入力次元数を D ユニット数を U とすると $W * D * U * 3 + U * 3$ に対し, LSTM は $D * U * 4 + U$ (記憶セル) $* U * 4 + U * 4$ となるため, より過去の情報を考慮するほど, パラメータの数が LSTM より大きくなる問題がある. そこで本研究では図 2 のように畳み込み層 1 つとプーリング層を 2 つ追加した画像処理層を適用し, QRNN への入力次元数を削減することでパラメータの数を減らす.

5 実験方法

本研究では 3 種類の実験を行った. 全ての実験のハイパーパラメータは全結合層, LSTM 層, QRNN 層のユニット数は 256, 最適化アルゴリズムは Adam を使用し, 学習率は 0.0001 とした. QRNN の窓幅は 4 に固定する.

Using Quasi-Recurrent Neural Network to Reduce Training Time on Deep Recurrent Q-Network for Time Series Data

[†] Kyousuke Nishina, Graduate School of Information and Computer Science, Chiba Institute of Technology

[‡] Shigeru Fujita, Faculty of Information and Computer Science, Chiba Institute of Technology

実験 1 は DQN と図 2 の画像処理層を適用した DQN との学習の推移を比較する．これにより，画像処理層の変更による影響を調べる．環境は OpenAI gym が提供している Pong に適用した．Target network のアップデートは 10,000 ステップ毎に行い，1,000,000 ステップで ϵ が 1 から 0.01 に減少する ϵ -greedy 法を用いた．ステップ数は 50,000 から 2,500,000 とし，10000 ステップ毎に 10 回プレイした平均報酬値を比較する．

実験 2 は図 2 を適用した DRQN と QRNN を適用した Q-Network を，ミニバッチサイズを 16 と 32，入力フレームを 4，8，16，32 の時の学習の時間を比較する．実験環境は OpenAI gym が提供している環境の 1 つである Pendulum を 25% の確率で真っ暗な画面に遷移するものに適用した．学習ステップ数は 2,000 から 50,000 とした．

実験 3 は図 2 の画像処理層を適用した DQN，DRQN，QRNN を適用した Q-Network の 3 つの学習の推移を比較する．これにより，QRNN が時系列的特徴を学習しているかを調べる．ミニバッチサイズは 32，入力フレームは 4 とし，実験 2 と同じ環境に適用し，1000 ステップ毎に 10 回プレイした平均報酬値を比較する．

6 結果・考察

図 3 は本研究で提案した画像処理層を適用した DQN と従来の画像処理層の DQN の学習の推移を比較したものである．図 3 から，次元数を削減しても従来の画像処理層と同じように学習が進んでいることがわかる．

実験 2 の結果，ミニバッチサイズが 32，入力フレーム数 4 の時の学習時間の軽減率が 0.66%，ミニバッチサイズが 16，入力フレーム数 32 の時の学習時間の軽減率が 11.78% となった．このことからミニバッチサイズが小さい，または入力フレームの数が多くなるほど提案手法の学習時間が軽減されていくことがわかる．

図 4 は実験 3 の方法で図 2 を適用した DQN，DRQN，提案手法を比較したものである．図 4 から提案手法は DRQN よりも劣る結果となった．しかし，DQN と提案手法を比較すると，提案手法の方が高いため，過去の情報から行動を学習できていると考えられる．

7 おわりに

本研究では，DRQN の画像処理層を変更して QRNN を導入することにより，学習時間を軽減しつつ，時系列データを扱える Q-Network を提案した．実験には，OpenAI gym が提供している一部の環境に適用することで画像処理層の変更による学習の影響や提案手法と従来手法の学習時間及び性能を比較した．実験結果から画像処理層の変更による影響はないと言える．また，性能は DRQN よりも提案手法が最大 30% 劣る結果となった．しかし，DQN より提案手法が 30% 優れていたため，過去の情報から行動を学習できていると考えられる．また，ミニバッチサイズが 16，入力フレームが 32 の時，学習時間は DRQN に比べ提案手法の方が 11.78% 軽減された．

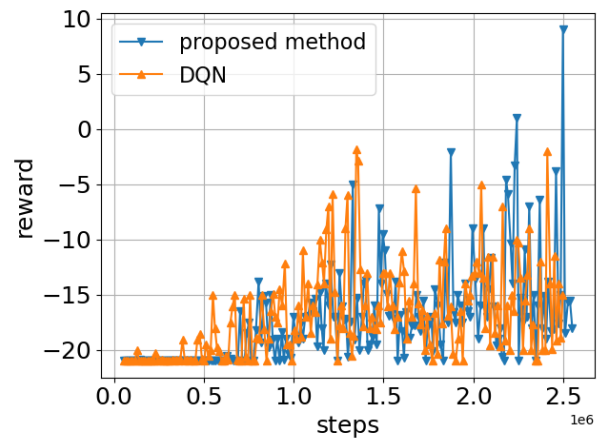


図 3 DQN と次元削減 DQN の学習推移

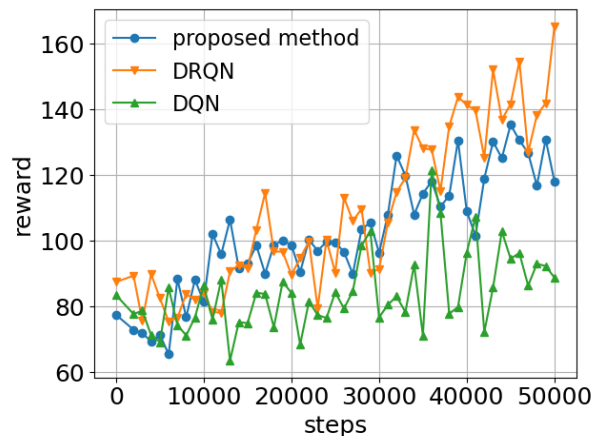


図 4 3つの手法の学習推移

このことから，提案手法は学習時間を最大 11.78% 軽減し，過去の情報から行動を学習できていることがいえる．

参考文献

- [1] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, and M.G. Bellemare et.al. Human-level control through deep reinforcement learning. In *Nature* 518, pp. 529–533. 2015.
- [2] Petter Stone Matthew Hausknecht. Deep Recurrent Q-Learning for Partially Observable MDPs. In *CoRR, abs/1507.06527*. 2015.
- [3] Steven Kapturowski, Georg Ostrovski, Will Dabney, John Quan, and Remi Munos. Recurrent Experience Replay in Distributed Reinforcement Learning. In *International Conference on Learning Representations 2019*.
- [4] J. Bradbury, S. Merity, C. Xiong, and R. Socher. Quasi-recurrent neural networks. In *International Conference on Learning Representations (ICLR 2017)*. 2017.