ISSN 2186-5647

深層強化学習における状態空間構成の最適化

日大生産工　○小松　泰士　　日大生産工　山内　ゆかり

1. まえがき

近年，強化学習 1) を搭載した自律的にタスクを遂行するエージェントやロボットの実現が強く望まれている．その中で，環境を通じて適応的な学習が可能な強化学習が注目されている．実用性といった観点から，状態空間を予め適切に構成することが問題点の一つである．

強化学習には状態と行動に対する評価値としてQ値を扱うQ-Learning(QL)がある．近年，状態数が膨大な連続空間でも学習が可能な深層強化学習(DQN) 2)が注目されている．この学習方法では，連続空間の膨大な状態数を畳み込みニューラルネットワーク(CNN)を用いて状態数を圧縮し，ニューラルネットワークを通して学習を行っている．

本報告では，DQNのアルゴリズム内で行われている状態数圧縮，つまり状態空間構成を自己組織化マップ(SOM) 3)・自己増殖型ニューラルネットワーク(SOINN) 4)に置き換えて実験を行う．また，既存のSOM・SOINNで状態空間構成を行ったQLと学習結果の比較検討を行う．

2. 研究背景及び従来手法

2.1 Q-Learning (QL)

Q-Learningとはある状態の時にとったある行動の価値を，Q-Tableと呼ばれる表で管理し，行動する毎にテーブルの値(Q値)を更新していく手法である．Q値の更新では式(1)を用いる．

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \gamma \max Q(s_{t+1},a_t) - Q(s_t,a_t)] \quad (1)$$

2.2 Deep Q-Network (DQN)

Deep Q-NetworkとはQ-Tableでは計算できない状態数の複雑になった環境をNeural Networkの関数近似によって計算する手法である．DQNの全体構造をFig.1に示す．

入力ノード数は環境のすべての状態を表現できるだけの膨大な数が必要なため，全結合ニューラルネットワーク層へ渡す前に状態空間の圧縮を行う．既存の論文では圧縮に畳み込みニューラルネットワーク(CNN)を用いることが多いが，本論文では，SOMやSOINNを適用し，DQNの学習効率を測定する．

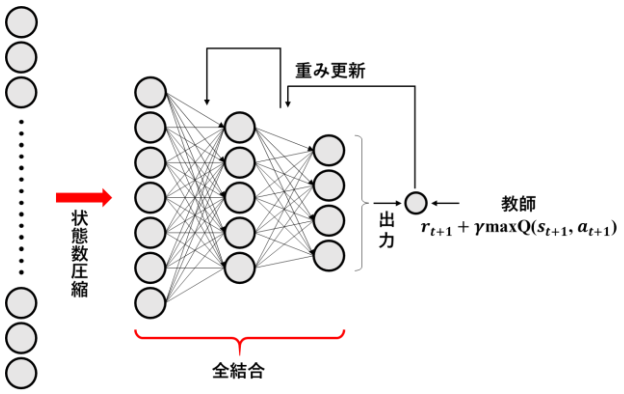

Fig.1 DQNの全体構造

2.3 Self-Organizing Map (SOM)

自己組織化マップとはニューラルネットワークの一種で与えられた入力情報の類似度をマップ上での距離で表現するモデルである．SOMは高次元データの中に存在する傾向や相関関係の発見などに応用することができる．

SOMの特徴とは，様々な高次元データを予備知識なし（教師なし）にクラスタリングできる点にある．

2.4 Self-Organizing Incremental Neural Network (SOINN)

自己増殖型ニューラルネットワークとは，追加学習可能なオンライン教師なし学習手法であり，事前にネットワークの構造を決定する必要がないほか，高いノイズ耐性を有し，計算が軽いなどの特長がある．また，ノード数が常に一定のSOMとは異なりノードの追加・削除を随時行うため，追記学習が可能である．

3. 提案手法

本研究では，DQNアルゴリズム内の状態空間の構成にSOM・SOINNを適用させる手法を提案する．つまりFig.1内の矢印である状態数圧縮にSOM・SOINNを用いる．

Q値の計算には，まず環境状態からSOMの近傍関数を用いて，N × Nのノード内で勝者を見つける．ニューラルネットワークの入力ノードはSOMのノード数と合わせ，勝者ノードには1，その他のノードには0を入力として前向き演算を行う．中間層の誤差関数にはシグモイド関数，出力層の誤差関数にはReLU関数を用いる．

An Optimization of State Space Composition in Deep Reinforcement Learning

Taishi KOMATSU and Yukari YAMAUCHI

重み更新のための教師の設定には式(2)を用いる．

$$Tk = Ok(s_t, a_t) + \alpha[r_{t+1} + \gamma \max Ok(s_{t+1}, a_t) - Ok(s_t, a_t)] \quad (2)$$

SOINNの場合，入力ノード数がN × NのSOMとはことなり，入力ノード数が増減する．Fig.2にSOINN+Q-Networkの全体構造を示す．

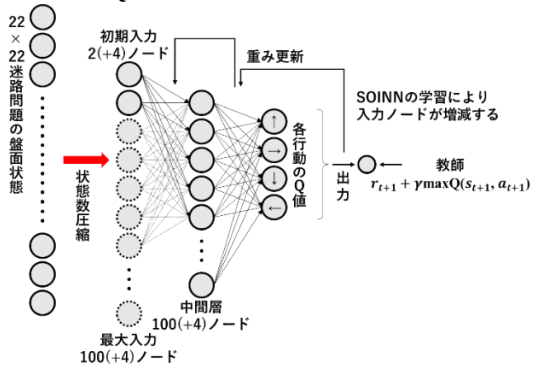

Fig.2 SOINN+Q-Networkの全体構造

4. 実験方法および測定方法

提案手法の有効性を検証するために，エージェントのナビゲーションタスク(迷路問題)における計算機シミュレーションを行った．

エージェントが置かれる環境は2次元の(x, y)平面環境である．位置を表す各座標は外壁も含め，0~21 の整数値をとるものとする．エージェントは図中の黒塗りの部分(外壁及び障害物)には侵入できず，侵入するような行動を行うと強制的に行動をとる前の位置に戻されるものとする．

このタスクでは，エージェントはゴール地点に到達したときに報酬を与えられ，スタート地点からゴール地点までの最短経路を学習する．

行動は『上，下，左，右』への四方向への移動から 1 つ選択されるものとし，各行動でそれぞれ 1 マス進めるものとする．行動選択の方策にはボルツマンを用いる．実験では，

・SOM を用いて状態空間構成を行った場合
・SOINN を用いて状態空間構成を行った場合

の 2 つの手法を適用して実験を行う．

5. 実験結果および検討

今回の実験で検討すべきことは，学習の早さと探索した経路の深さの2つである．SOM・SOINNを用いてQL・QNを学習させた結果をそれぞれFig.3・Fig.4示す．

SOM・SOINNどちらとも，Q-Networkの方が学習の進みが早く，ゴールまでのTime数が短い経路が探索できていることが分かる．

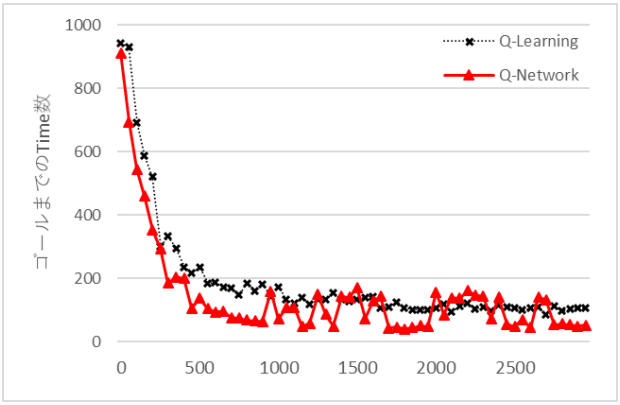

Fig.3 SOMによる状態空間構成の学習結果

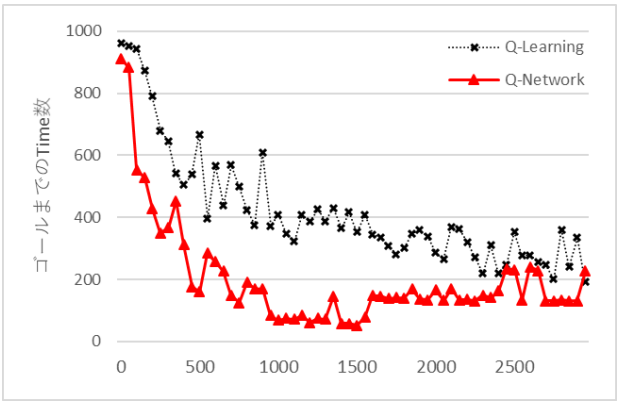

Fig.4 SOINNによる状態空間構成の学習結果

6. まとめ

DQNアルゴリズムへのSOM・SOINNの適用を提案し，QLとの学習結果を比較した．結果，学習の早さ・深さともにQ-Networkの方が上回った．これは，状態ノードが増加し，固まって配置された際にノードを1個ずつ表計算で学習させるよりも，大まかに特徴を抽出して学習させるニューラルネットワークの方が適していたためだと考えられる．

【参考文献】

1) Sutton, R. & Barto A., “Reinforcement Learning: An Introduction”, MIT Press, 1998
2) Volodymyr Mnih et al., “Human-level control through deep reinforcement learning”, Nature 14236, 2015
3) T. Kohonen, "Self-organized formation of topologically correct feature maps", Biol. Cybern., vol. 43, pp. 59-69, 1982
4) 山崎和博 他, “自己増殖型ニューラルネットワーク SOINN とその実践”, 日本神経回路学会誌, Vol.17, No.4, pp.187-196, 2010