7R-04

Deep Q-Network に基づくクラス分類

Classification in DeepQ-Network

山村 光平
Kohei YAMAMURA

1 はじめに

近年，機械学習によるパターン認識技術が目覚ましく発展し，手法には主に教師あり学習が用いられている．そこで，本稿ではパターン認識におけるクラス分類課題を取り上げ，強化学習の手法によるアプローチを行い，その有用性を検証する．実験では人工データを使用し深層強化学習の一手法である Deep Q-Network によりクラス分類を行う．

2 Deep Q-Network

強化学習ではエージェントが環境との相互作用を通して学習を行う [1]．学習の流れを図 1 に示す．まずエージェントが環境から状態を入力として受け取り，行動を出力する．次に環境が受け取った行動を基にエージェントに新たな状態と受け取った行動の評価にあたる報酬とを出力する．以上を繰り返すことで学習を行う．

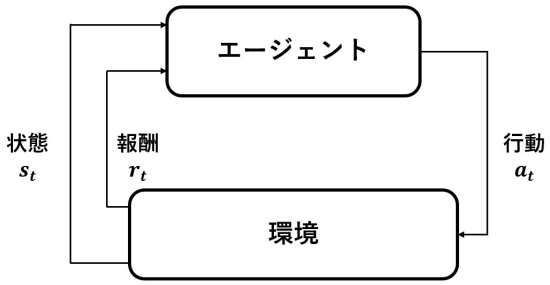

図 1　学習の流れ

最終的に得ることができる報酬の合計，すなわち収益は式 (1) で表され，γ は割引率を表す．強化学習における学習の目標は収益が最大となる方策をエージェントが獲得することである．

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{1}$$

本稿では状態を連続値として扱うため強化学習にニューラルネットワークを応用した DQN（Deep Q-Network）を用いる．DQN では行動価値関数を深層ニューラルネットワークによって近似を行う．行動価値関数とは状態 s において取り得る行動 a の価値を出力する関数であり更新式は式 (2) で表される．

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha\delta_{t+1} \tag{2}$$

このとき α は学習率にあたり，δ_{t+1} は TD 誤差（temporal difference error）と呼ばれる差分であり，次式で表される．

$$\delta_{t+1} = r_{t+1} + \gamma \max_a Q_t(s_{t+1},a) - Q(s_t,a_t) \tag{3}$$

DQN では TD 誤差を基にしたニューラルネットワークにおける損失が計算される．本稿では TD 誤差の平均二乗誤差 (MSE) を損失 L とし，損失 L を最小化するように学習を行う [2]．またエージェントにおける Q ネットワークの構造は次のようになる．

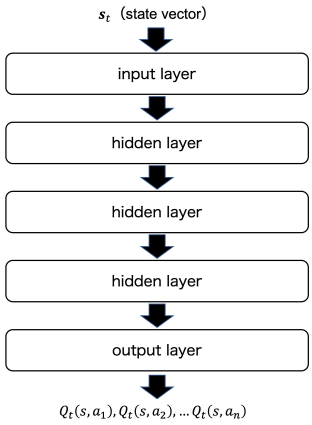

図 2　Q ネットワークの構造

3 実験

3.1 概要

本稿におけるクラス分類課題はシャッフルされた推測問題とする [3]．エージェントは各タイムステップにおいてサンプルの特徴（状態）を受け取り，サンプルが属するクラスの推測を行う．環境はエージェントが行った推測（行動）を基に予めシャッフルされたデータセットから次のサンプルと報酬をエージェントに返す．このとき，報酬はエージェントの予想が正解であれば報酬を 1，不正解であれば報酬を-1 とする．

分類を行うデータセットは, 人為的に作成した「人工データ」である GMM300 に対して実験評価を行う. データセットの性質を表 1 に示す. 人工データセットの GMM300 は混合正規分布を母集団として生成した後, 分布毎にクラスラベルを付与している [4]. データセットは学習用標本と試験用標本に分割し，学習には学習用標本のみを用いる．

表 1　データセット

データセット	学習用標本数	試験用標本数	次元数	クラス数
GMM300	300	11700	2	2

学習実験では，学習用標本の全クラスの推測を 1 エピソードとし，20000 エピソード学習を行う．Q ネットワークの学習率は 0.001 とする．評価実験は学習用標本と試験用標本とでそれぞれ行う．

3.2 結果と考察

学習実験の結果を以下に示す．図 3 では横軸がエピソード，縦軸が報酬の和にあたる総報酬を表す．図 4 では横軸がエピソード，縦軸が Q ネットワークの損失を表す．

GMM300 のデータセットにおける学習用標本数が 300 であり，図 3 より学習が進むにつれ総報酬の値が最大の 300 に近づき，振動していることが伺える．また，図 4 では Q ネットワークの損失が学習が進むごとに小さくなっているため正しく学習が行われていることが示唆された．

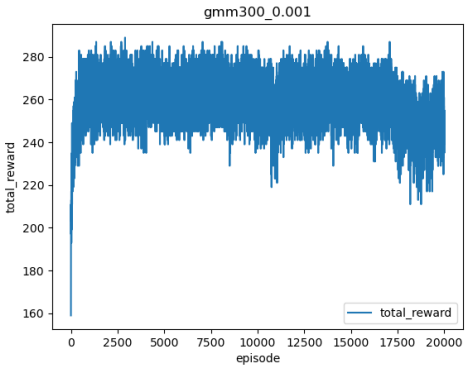

図 3　総報酬の学習過程

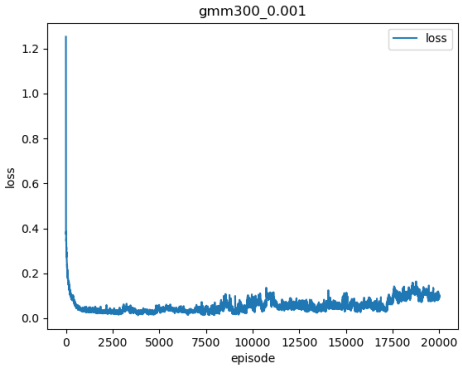

図 4　損失の学習過程

次に，評価実験の結果を以下に示す．図 5 では横軸がエピソード，縦軸が学習用データにおける誤り率を表す．図 6 では横軸がエピソード，縦軸が試験用データにおける誤り率を表す．

図 5 の学習用標本では学習が進むにつれ誤り率が低下している．一方で図 6 の試験用標本では学習が進むにつれ誤り率が徐々に上昇している．以上の結果よりエージェントは学習用標本における最適な方策を獲得することはできたが，学習用標本に対して過適合してしまったために試験用標本においては最適な分類ができなかったことが示唆された．

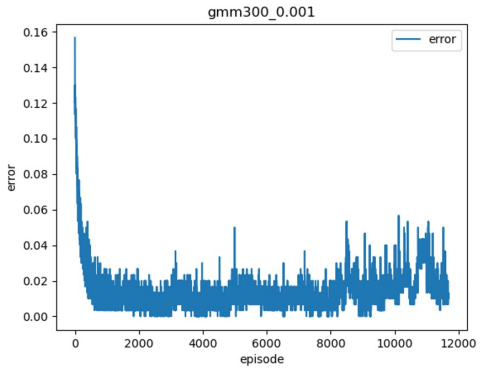

図 5　学習用標本における誤り率

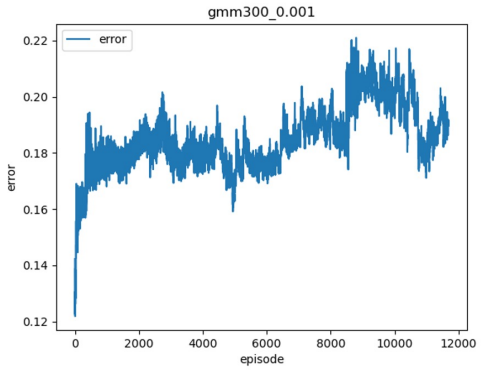

図 6　試験用標本における誤り率

4　まとめ

本稿ではクラス分類課題を取り上げ，強化学習によるアプローチを行い，手法としては深層強化学習の一つである DQN を用いた．獲得した分類器モデルは学習用標本に対して高い分類精度を発揮したが，一方で試験用標本における分類精度はエピソードが進むにつれ過学習になっていることが示唆された．以上の結果より今後の課題としては学習用標本に対する過学習を回避できるモデルを開発する必要がある．

参考文献

1) Richard S. Sutton and Andrew G. Barto. Reinforcement learning, 1998.
2) Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning, Feb 2015.
3) Xiaoming Qi Enlu Lin, Qiong Chen. Deep reinforcement learning for imbalanced classification, Jan 2019.
4) Masahiro Senda. Study of maximum bayes boundary-ness training method, 2020 Jan.