

# Deep Belief Networkを用いたQ Learningの実現

樋貝隼人 長名優子

東京工科大学 コンピュータサイエンス学部

## 1 はじめに

人工知能の一分野として研究されている機械学習、その中で Deep Learning の手法は、画像認識や音声認識などの有効性が示され、注目されている [1]。2013 年に Volodymyr Mnih らが提案した Deep Q-Network[6] と呼ばれる手法は、強化学習の一種の Q Learning を、Deep Learning の一手法である畳み込みニューラルネットワークで実現したものである。ゲームのプレイ画面を観測として畳み込みニューラルネットワーク [2] に入力し、その状態におけるすべての行動価値を出力するように学習を行う。Deep Q-Network は Atari2600 の複数のゲームに適用されており、従来手法に勝る結果が得られることが示されている。しかし、Deep Q-Network では畳み込みニューラルネットワークを用いるため、非常に多くの学習データが必要になる。

Deep Belief Network では入力データの生成モデルを獲得するように学習を行う。それに対し、畳み込みニューラルネットワークでは、入力データと出力データの間を学習していく。これは生成モデルと識別モデルの違いであり、Deep Belief Network では、畳み込みニューラルネットワークに比べ、学習データが少なくても学習が行える可能性があると言われている。

そこで本研究では、Deep Q-Network で用いられている畳み込みニューラルネットワークの代わりに Deep Belief Network を用いる手法を提案する。Deep Belief Network を用いることで、畳み込みニューラルネットワークを用いた場合よりも学習データが少なくても学習が行える可能性があると考えられる。本研究の成果は、Deep Q-Network と比較し性能を評価する。

## 2 Deep Q-Network

Deep Q-Network[6] は、Q Learning における行動価値の学習を行う畳み込みニューラルネットワークである。Deep Q-Network は畳み込みニューラルネットワークに基づくモデルであり、3 層の畳み込み層と 2

Q Learning using Deep Belief Network  
Hayato Higai and Yuko Osana (Tokyo University of  
Technology, osana@stf.teu.ac.jp)

層の全結合層から構成されている。ゲームのプレイ画面を観測として畳み込みニューラルネットワークに入力し、その状態におけるすべての行動価値を出力するように学習を行う。

## 3 Deep Belief Network (DBN)

Deep Belief Network[3] は複数の制約付きボルツマンマシン (Restricted Boltzmann Machine : RBM)[7] を階層構造に積み重ねたネットワークである。この手法では、事前学習で、学習サンプルの観測分布を表す生成モデルを学習する。従来の研究では、多層構造にするほど精度が低下するという問題があったが、2006 年にこの手法が発表され、これが現在の Deep Learning のブームの元となった。この手法では各層で学習を行ってパラメータを調整し、学習済みのネットワークの出力を次の層への入力にすることで多層にするメリットを活かし、従来の研究より精度を高めている。

### 3.1 制約付きボルツマンマシン

制約付きボルツマンマシン [7] は、可視層と隠れ層のニューロン間にのみ結合が存在するボルツマンマシンで RBM と呼ばれている。図 1 に RBM の構造を示す。RBM は可視層と隠れ層の 2 つの層から構成されており、可視層にデータが入力されると隠れ層にそれに対応した何らかのパターンが出力される。隠れ層に出力されるパターンは入力データに対応する内部表現として捉えることができ、入力データの特徴を抽出したものであると考えられる。Deep Belief Network では、RBM を階層構造に積み重ね、多数の段数を経ることで入力データから特徴を抽出する。

### 3.2 Deep Belief Network の流れ

Deep Belief Network[3] の学習は入力に近い層から順に行う。図 2 に Deep Belief Network の構造を示す。

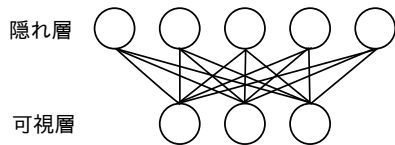


図 1: 制約ボルツマンマシンの構造

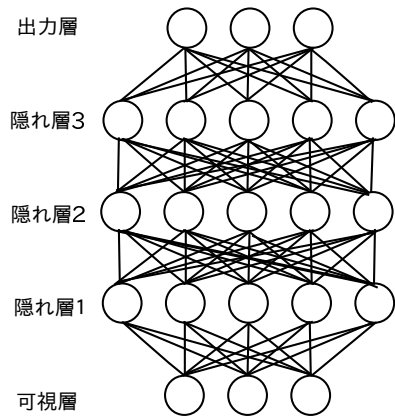


図 2: Deep Belief Network の構造

図 2 に示すような構造の Deep Belief Network では、まず可視層と隠れ層 1 を RBM として学習を行う。次に隠れ層 1 の出力を入力とし、隠れ層 1 を可視層、隠れ層 1 を隠れ層とした RBM と見なし学習を行う。次にして、隠れ層 2 と隠れ層 3 を RBM として学習を行う。最後の出力層は認識を行うためのニューロンが配置され、出力層に対して教師信号を与え、確率的勾配降下法を用いて出力誤差を逆伝播することで学習を行う。確率的勾配降下法では学習データの中からランダムにデータを 1 つ取り出して誤差を計算し、パラメータを勾配降下法に基づいて更新する。この方法では、毎回、別のデータに対する誤差関数に対して勾配を求め、勾配を下ろうとするため、局所解におちいりやすく最適解にたどり着きやすいという特徴がある。

#### 4 Deep Belief Network を用いた Q Learning の実現

この手法は Deep Q-Network の畳み込みニューラルネットワークを Deep Belief Network で置き換えたものである。図 3 に、Deep Belief Network を用いた Q Learning の構造を示す。Deep Belief Network を用いて Q Learning を実現する場合は、観測を入力とし、それに対する行動価値を出力層のニューロンから出力するように学習を行う。ゲームの学習を行う場合には、ゲーム画面を観測として入力することになる。教師信号としては Q Learning の価値を使用する。

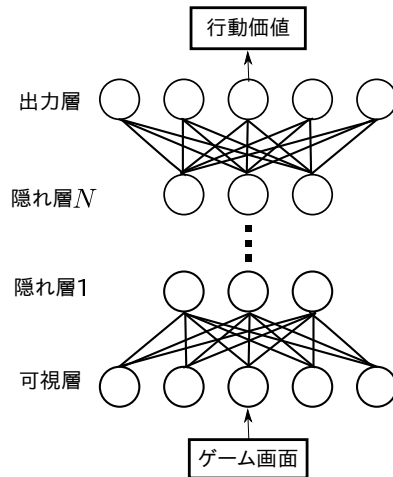


図 3: Deep Belief Network を用いた Q Learning の構造

#### 5 計算機実験

計算機実験を行い、提案手法において学習が行えることを確認した。

#### 参考文献

- [1] A. Krizhevsky, I. Sutsuker and G. E. Hinton : “ImageNet classification with deep convolutional neural networks,” In Advances in NIPS, pp.1097–1105, 2012.
- [2] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner : “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, Vol.86, No.11, pp.2278–2324, 1998.
- [3] G. E. Hinton, S. Osindero and Y. Teh : “A fast learning algorithm for deep belief nets,” Neural Computation, Vol.18, pp.1527–1544, 2006.
- [4] C. J. C. H. Watkins and P. Dayan : “Technical Note: Q Learning,” Machine Learning, Vol.8, pp.55–68, 1992.
- [5] J. J. Grefenstette : “Credit assignment in rule discovery systems based on genetic algorithms,” Machine Learning, Vol.3, pp.225–245, 1988.
- [6] V. Mnih *et al.* : “Human-level control through deep reinforcement learning,” Nature, No.518, pp.529–533, 2015.
- [7] G. E. Hinton : “A practical guide to training restricted boltzmann machines,” Technical Report, 2010.