

転移学習における知識の評価手法の検討

2019 年 3 月

東京電機大学工学部 情報通信工学科

15EC067

鈴木 溪太

指導教員 教授 鈴木 剛

目 次

第1章 序論	1
1.1 研究背景	1
1.2 研究目的	2
1.3 本論文の構成	2
第2章 強化学習と転移学習	3
2.1 緒言	3
2.2 強化学習	3
2.3 Q 学習	4
2.4 転移学習	5
2.5 転移先環境に合わせた転移学習	6
2.6 結言	6
第3章 ニューラルネットワーク	7
3.1 緒言	7
3.2 パーセプトロン	8
3.3 順伝播型ニューラルネットワーク	9
3.3.1 訓練データによる学習	10
3.3.2 損失関数	10
3.3.3 勾配降下法	11
3.4 結言	11
第4章 提案手法	12
4.1 緒言	12
4.2 先行研究に対する課題と提案手法	12
4.3 最短経路問題における知識の特徴	13
4.4 ニューラルネットワークの構成	15
4.5 カテゴリによる転移学習手法	16
4.6 データ収集	17
4.6.1 転移知識の収集	17
4.6.2 カテゴリライズ環境の収集	18
4.7 結言	18
第5章 実験	19
5.1 緒言	19
5.2 提案手法を用いたニューラルネットワークの構築	19
5.2.1 実験目的	19
5.2.2 実験概要	19
5.2.3 実験結果と考察	21
5.3 カテゴリによる転移学習の検証	22

5.3.1 実験目的	22
5.3.2 実験概要	22
5.3.3 実験結果と考察	24
5.4 結言	26
第6章 結論	27
6.1 本論文の結論	27
6.2 今後の展望	27
謝辞	28
参考文献	29

図 目 次

図 2. 1 想定学習環境	3
図 3. 1 単純パーセプトロン	8
図 3. 2 三層ニューラルネットワーク	9
図 4. 1 最短経路問題のグリッドフィールド例	13
図 4. 2 提案手法で用いるニューラルネットワーク構成	15
図 5. 1 ニューラルネットワークの学習結果	21
図 5. 2 カテゴリを用いた転移学習結果	26

表 目 次

表 4. 1 カテゴリ分類	14
表 5. 1 データセット内訳	20
表 5. 2 カテゴリ毎の平均総ステップ数	25
表 5. 3 カテゴリ毎の減少数	26
表 5. 4 カテゴリ全体の平均総ステップ数	27

第1章 序論

1.1 研究背景

近年, 少子高齢化の影響により, 労働人口の減少が懸念されている. みずほ総合研究所の調査によると, 2016 年の労働人口は 6648 万人, 労働力率は 60%であるが, 男女別, 年齢 5 歳階級別の労働力率を同じとすれば 2065 年の労働人口は 4000 万人弱と約 4 割減少し, 労働力率は 50%程度まで低下する[1]とされている.

労働人口の減少を補うため, ロボットによる省人化が期待されている. このことから, 倉庫・工場・公共交通など様々な環境でロボットを用いることによって人手不足の解消を図ろうという動きが盛んになっている.

しかし, ロボットを用いる場所により環境は大きく変化するため, ロボットの操作者がロボットに環境に応じてプログラムを施す事は困難である. また, そのコストは非常に大きくなると予想される.

このことから, ロボットが環境に応じて自律的に行動を行うことが求められている[2].

1.2 研究目的

ロボットが多様な環境に適応する手法として、強化学習と呼ばれる手法がある。強化学習とは、学習システムとしてのエージェントが、環境から得られる報酬を最大化することを目的として、観測した状態に対して適切な行動を自律的に学習する機械学習手法である。

本研究では強化学習の一手法である Q 学習を用いる。Q 学習は、エージェントが報酬を得られるように行動を選択することで、最適な行動を自律的に学ぶため、報酬の獲得方法を定義すれば確実に学習が可能である。その反面、学習は試行錯誤的に行われるため、試行回数の増加が考えられる。

試行回数を減らす手法として、他の環境で学習した知識である転移知識を転移先の環境のエージェントに転移し、新たな環境に適応させる転移学習がある。転移学習により、試行回数を削減し、エージェントは未知の環境にいち早く適応する事ができる。しかし、転移知識は転移元の環境で学んだ知識であるため、転移先の環境が大きく異なる場合にはそのまま利用することができない。従って、知識をより有効に活用する手法が必要である。その手法の一つとして、複数の知識の中から選択的に知識を転移させる方法がある。

2018 年高桑によって行われた研究により、転移知識の再利用における選択手法が確立され、知識の特徴であるカテゴリごとに分類された知識(以下、カテゴリ化知識と呼称)効率的に選択して知識を転移する事が可能となった。しかし、カテゴリ化知識が転移学習の効率に与える影響は未検討である。

本研究では、転移知識を特徴ごとに分類する際の評価指標を求め、評価指標から未評価の知識を分類する手法を提案し、カテゴリ化知識が転移学習の効率に与える影響について検討を行う。

1.3 本論文の構成

本論文は全 6 章から構成される。本章では序論として研究背景、研究目的について述べた。第 2 章では、強化学習と転移学習について説明する。第 3 章ではニューラルネットワークについて説明する。第 4 章では、本研究にて提案するニューラルネットワークの構成とカテゴリによる転移学習手法について述べる。第 5 章では提案手法を用いてニューラルネットワークを構築し、カテゴリによる転移学習を行う。第 6 章では結論を述べる。

第2章 強化学習と転移学習

2.1 緒言

本章では、機械学習の一種である強化学習及び強化学習によって獲得した知識を転移する転移学習、転移先の環境に合わせて転移学習を行う手法について説明し、それらの課題点について述べる。

2.2 強化学習

強化学習とは、機械学習の一種であり、学習システムとしてのエージェントが、環境から得られる報酬を最大化することを目的として、観測した状態に対して適切な行動を自律的に選択・学習していく学習手法である。

本研究で想定している学習環境は図 2. 1 のようなグリッドフィールドである。

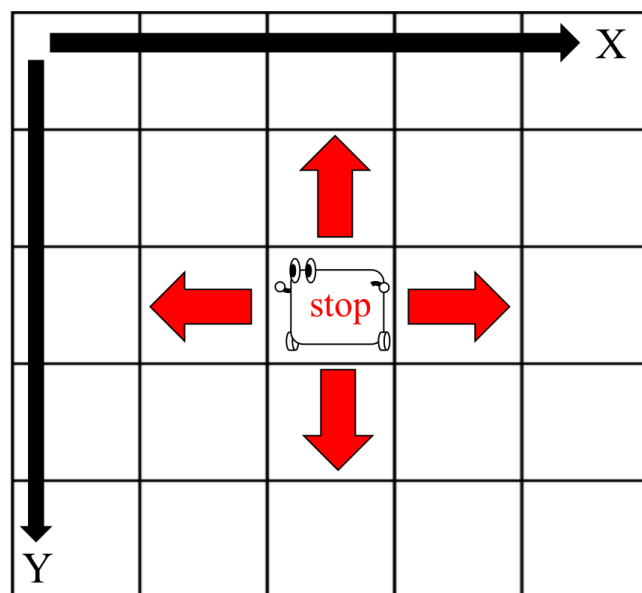


図 2. 1: 想定学習環境

2.3 Q 学習

Q 学習とは、強化学習の一手法であり、エージェントの状態 s における実行可能な行動 a を選択する行動価値 (以下 Q 値と呼称) を学習する学習手法である。よって、Q 値は状態 s , 行動 a の時にとる行動確率の高さを表している。

式(2. 1)の Q 値の更新式を用いることによって、エージェントは Q 値の更新を行い、より最適な行動を自律的に学習していく。

$$Q(s_t, a) \leftarrow Q(s_t, a) + \alpha \{r + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a)\} \quad (2. 1)$$

式(2. 1)において、 t とは、ある時刻を表し、 $t+1$ は次の時刻を表している。また、本研究における状態 s はグリッドフィールドにおけるエージェントの x 座標と y 座標を指し、行動 a はエージェントの上下左右への移動と停止という合計 5 つの行動にそれぞれ番号を割り当てた物である。 α は学習率、 r は報酬、 γ は割引率を表し、これらのパラメータの設定によって学習の速度が変化する。

エージェントが報酬 r を得るように行動を選択することで、最適な行動を自律的に学ぶため、報酬の獲得方法を定義することができれば確実に学習が可能である。

エージェントの行動選択手法として、本研究ではボルツマン選択手法を用いる。

ボルツマン選択手法とは Q 値の大きさに応じて相応の重みを掛け、確率的に行動するソフトマックス行動選択手法にボルツマン分布を用いる手法である。

ボルツマン選択手法を用いることで、Q 値の大きさに応じて確率的に Q 値を選択することが可能になり、学習が一意に定まらずランダム性を持つ事が可能になる。

ボルツマン選択手法は、ある行動をとる確率を $\pi(s, a)$ とし、式(2. 2)で表される。

$$\pi(s, a) = \frac{\exp\left(\frac{Q(s, a)}{T}\right)}{\sum_{b \in A} \exp\left(\frac{Q(s, b)}{T}\right)} \quad (2. 2)$$

s はエージェントの状態、 a はエージェントの行動、 A はエージェントの行動の集合を表す。 T はボルツマン温度と呼ばれる係数であり、この値が大きくなるとエージェントの行動をランダムに選択する確率が増える。本研究ではボルツマン温度に定数を用いる。

2.4 転移学習

式(2. 1)を用いて Q 値の更新を行い試行錯誤的に学習する Q 学習は, 環境の複雑性が増加するに従って試行回数が増加すると考えられる. そこで, 他の環境で学習した知識(以下, 転移知識と呼称)を利用し, 新たな環境に適応させる転移学習という手法を用いる. 転移学習を用いる事で試行回数を削減し, エージェントは新たな環境にいち早く適応する事が可能である. 転移学習では, 転移知識 Q_s と新たに得た知識 Q_t を統合した, 統合知識 Q_c を更新することで式(2. 3)を用いて転移学習を行う.

$$Q_c \leftarrow Q_s + Q_t \quad (2. 3)$$

転移学習の行動選択は Q_c を用いて行われる. Q_c は Q_s を元に作られているため, エージェントが現在学習している環境(以下, 転移先環境と呼称)が Q_s を学習した環境(以下, 転移元環境と呼称)と等しい場合, 試行錯誤的な学習回数が減少し, 学習の収束速度が早くなると考えられる.

しかし, 転移先環境が転移元環境と異なる場合, Q_s は転移先環境にとって悪影響を及ぼす可能性がある. この例を図 2 に示す.

図 2 のように, 転移元環境では有用であった Q_s の行動が, 転移先環境では不要な行動になる可能性がある. この学習環境の差異は試行回数の増加の原因となるだけでなく, 学習が膠着し進まなくなる可能性も存在する. この問題は先に述べたように, 転移先環境と転移元環境が異なるという点に起因する. 従って, 転移先環境に適した転移知識を選択して転移学習を行う事が可能であれば, この問題は緩和すると考えられる.

2.5 転移先環境に合わせた転移学習

転移先環境に合わせた転移学習として、2018 年、高桑によって行われた「活性化拡散モデルに基づく強化学習エージェントの転移学習手法」が挙げられる。この研究によって、強化学習を行うエージェントが獲得・保存した複数知識の再利用における転移知識の選択手法の提案[3]がなされた。高桑(2018)は、転移知識の特徴に基づき、類似した転移知識同士を特徴毎にカテゴリとして分類し、カテゴリ毎に転移知識間を結合するモデルとして活性化拡散モデルを適用した。

活性化拡散モデルとは、人間が獲得した概念同士が脳内でネットワーク構造として保存されていることを前提とし、ある概念が想起されることで、関連する概念も活性化され、概念の利用が促進されるモデルである[4]。活性化拡散モデルには、関連性の強さに応じて関連している概念間の距離を変動させて配置する意味的距離が存在する。概念の活性化は、関連性によって構築されたネットワークを通じて行われ、各概念間に意味的な関連性の表現が存在する。

高桑(2018)の行った転移学習では、転移知識が適切に想起された場合、すなわち転移学習をした結果が効率的な場合には、転移知識間の意味的距離を縮める。逆に、転移知識が適切に想起されなかった場合、すなわち転移学習をした結果が非効率な場合には、転移知識間の意味的距離を伸ばす。これにより、複数知識の関係を転移先環境に応じて選択することが可能となる。この研究によって、複数の知識を用いた転移学習は強化学習や単一方策のみの転移学習と比較して有用である事が検証された。その一方で、類似した転移知識同士を特徴毎にカテゴリに分類する際の分類基準は手法利用者に委ねられているため、学習の効率は転移知識のカテゴリ分類に依存すると考えられる。

2.6 終言

本章では、強化学習と転移学習、転移先環境に合わせた転移学習について説明を行い、その課題点について述べた。次章では、本研究の提案手法に取り入れるニューラルネットワークについて述べる。

第3章 ニューラルネットワーク

3.1 緒言

本章では, 人間の脳機能をモデル化したネットワーク構造であるニューラルネットワークについて説明する.

ニューラルネットワークとは, シナプスの結合によりネットワークを形成したユニットが, 学習によってシナプス同士の結合強度を変化させる事で, 任意の問題を解決する事が可能なネットワークを構築する数理的モデルである. 本研究ではニューラルネットワークの一種である順伝播型ニューラルネットワークを用いて学習を行う.

3.2 パーセプトロン

パーセプトロンとは、複数の入力信号を受け取ると、入力それぞれに対して重み付けをした値とバイアスを合計した値を活性化関数に通して一つの信号として出力するニューラルネットワークの一種である。活性化関数とは入力の総和がどのように活性化するか決定する役割を持つ関数である。

入力信号と出力信号はそれぞれユニットで管理され、入力層・中間層(隠れ層)・出力層というような層構造によってそれぞれユニットが配置される。この時、入力層と出力層の二層で構築されるものを単純パーセプトロン、一つ以上の中間層を持つものを多層パーセプトロンと呼ぶ。入力個数を n 、入力を x_1, x_2, \dots, x_n 、重みを w_1, w_2, \dots, w_n 、バイアスを b 、出力を y とすると、単純パーセプトロンは図 3. 1 のように表される。

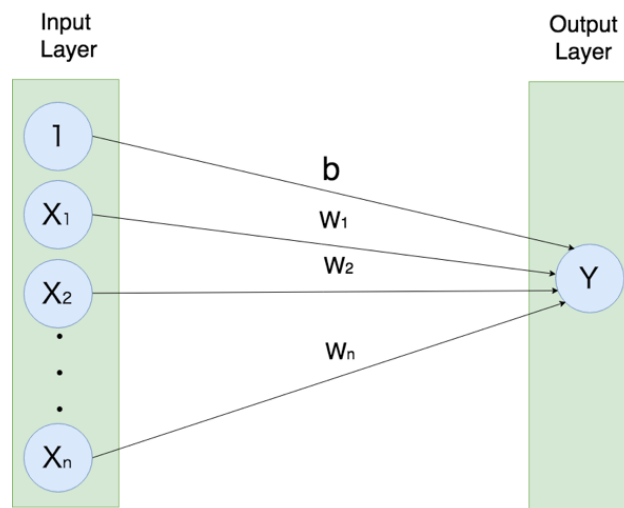


図 3. 1: 単純パーセプトロン

単純パーセプトロンは式(3. 1)として表される。なお、単純パーセプトロンの活性化関数は入力が 0 より大きい場合に 1, 入力が 0 以下の場合は 0 を出力するステップ関数を用いられる。

$$y = \begin{cases} 0 & (w_1x_1 + w_2x_2 + \dots + w_nx_n + b \geq 0) \\ 1 & (w_1x_1 + w_2x_2 + \dots + w_nx_n + b \leq 0) \end{cases} \quad (3. 1)$$

3.3 順伝搬型ニューラルネットワーク

順伝搬型ニューラルネットワーク(以下, ニューラルネットワークと呼称)とは, 単純パーセプトロンを並べた物を一つの層とし, 隣接した層を結合したものである. 例として三層ニューラルネットワークを図 3. 2 に示す.

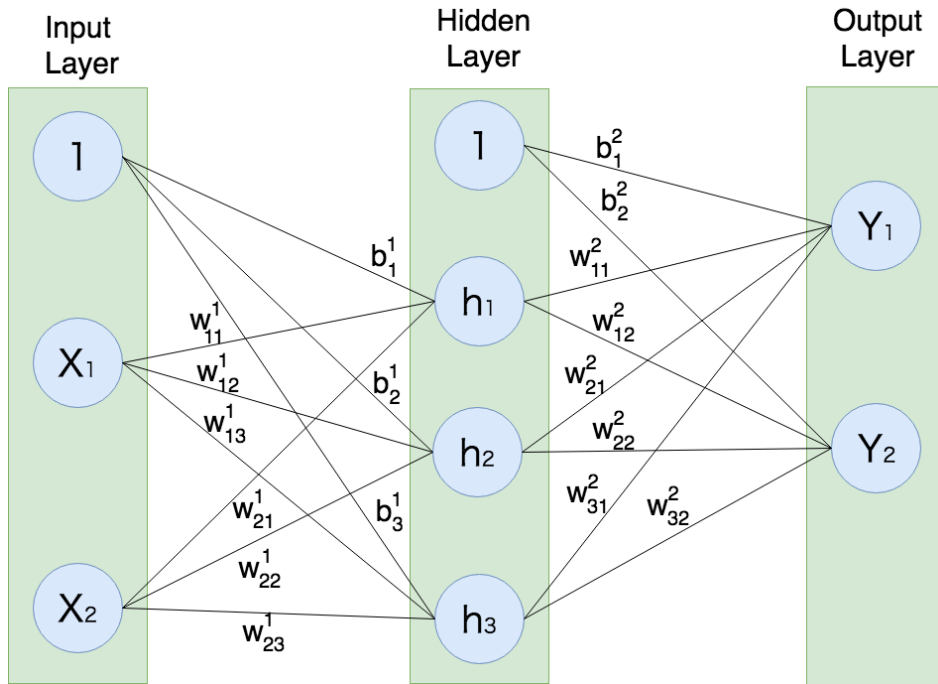


図 3. 2: 三層ニューラルネットワーク

このパーセプトロンの活性化関数を $f(x)$ とすると, 図 Y は式(3. 2)として表される.

$$\begin{aligned}
 h_1 &= f(s_1) , \quad s_1 = w_{11}^1 x_1 + w_{21}^1 x_2 + b_1^1 \\
 h_2 &= f(s_2) , \quad s_2 = w_{12}^1 x_1 + w_{22}^1 x_2 + b_2^1 \\
 h_3 &= f(s_3) , \quad s_3 = w_{13}^1 x_1 + w_{23}^1 x_2 + b_3^1 \\
 y_1 &= w_{11}^2 h_1 + w_{21}^2 h_2 + w_{31}^2 h_3 + b_1^2 \\
 y_2 &= w_{12}^2 h_1 + w_{22}^2 h_2 + w_{32}^2 h_3 + b_2^2
 \end{aligned} \tag{3. 2}$$

三層以上のパーセプトロンの活性化関数にはシグモイド関数, ReLU 関数等, 複数の種類がある. 特徴として, 三層以上のパーセプトロンの活性化関数は非線形である. 従って, 三層以上のパーセプトロンでは任意の関数を近似可能である. この特徴から, ニューラルネットワークは線型分離不可能な問題を解くことができる.

3.3.1 訓練データによる学習

ニューラルネットワークにおける学習とは、出力が最適な値となるように、重みとバイアスを自動的に修正することである。

ニューラルネットワークの学習の順序は以下のようになる。

1. 入力データとそれに対する解答データをセットとし、訓練データ群を作成する。
2. 訓練データの入力データをニューラルネットワークに入力し、得られた出力と対応する解答データの誤差を損失関数によって求める。
3. 勾配降下法を用いて、求めた誤差が 0 に近づくように、重みとバイアスの値を修正する。

3.3.2 損失関数

損失関数とは、ニューラルネットワークの入力に対する出力 y と、入力に対する解答である教師データ t から、出力 y と教師データ t の近さの指標を求める関数である。本研究では損失関数として交差エントロピー誤差を用いる。 k 組の訓練データを用いた場合の交差エントロピー誤差 E は式(3. 3)で求められる。

$$E = \sum_{n=1}^k t_n \log(y_n) \quad (3. 3)$$

なお、損失関数の値が小さくなればなるほどニューラルネットワークの学習が進んでいると言えるが、これは訓練データに対する学習であるため、訓練データ以外のデータに対して学習がうまく進んでいるとは限らない。この状態を過学習という。従って、訓練データを用いて学習を行ったのち、学習に用いていない検証データを用いて学習の汎化性能を確かめる。また、ニューラルネットワークの学習が終了した時点で、学習にも検証にも用いていないテストデータを用いてニューラルネットワークの最終的な予測結果を評価する。

3.3.3 勾配降下法

勾配降下法とは, 損失関数で得た誤差 E を極小化するように重み w を変化させ, ニューラルネットワークの学習を進める方法である.

損失関数の勾配を $\frac{\partial E}{\partial w}$, 現在の重みを w^t , 動かした後の重みを w^{t+1} とすると, 更新後の w^{t+1} は式(3. 4)で求められる.

$$w^{t+1} = w^t - \eta \frac{\partial E}{\partial w} \quad (3. 4)$$

式(3. 3)の η は学習率と呼ばれ, 重み w の更新の大きさを決めるパラメータである. なお, 学習率 η には, 確実に学習が収束可能で最も大きい値を設定する方法が存在しないため, 非常に小さな値を用いることが推奨されている. このように適所解が存在しないパラメータをハイパーパラメータと呼ぶ.

3.4 結言

本章では, ニューラルネットワークの概要と理論について説明を行った. 次章では, 本研究の提案手法について述べる.

第4章 提案手法

4.1 緒言

本章では, 提案手法についての説明とその検証方法について述べる.

4.2 先行研究に対する課題と提案手法

第2章に述べた先行研究における課題として, 複数の知識を用いた転移学習は強化学習や単一方策のみの転移学習と比較して有用である一方, 学習の効率は転移知識のカテゴリ分類に依存する事がわかった. 従って, 効率の良い学習を行うためには転移知識をカテゴリに分類する際の指標を求める事が必要だと考えられる.

そこで, 本研究では転移知識をカテゴリに分類する際の評価指標を導出し, その値を元にカテゴリ分類を行い, 転移知識とカテゴリを訓練データとしてニューラルネットワークを構築することで, カテゴリ分類されていない転移知識に対しても一定のカテゴリ分類が可能なニューラルネットワークを作成できると考え, その手法を提案する. また, カテゴリ分類が転移学習の効率に与える影響について検討を行う.

4.3 最短経路問題における知識の特徴

カテゴリ分類の評価指標は、転移知識の特徴から求めることができると考えられる。本研究ではタスクとして、グリッドフィールド上における最短経路問題を取り扱う。グリッドフィールドの例を図 4. 1 に示す。

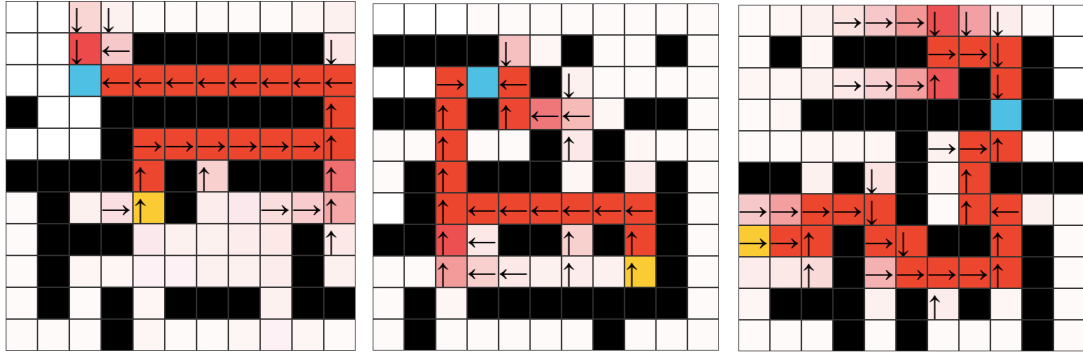


図 4. 1:最短経路問題のグリッドフィールド例

グリッドフィールドは縦 11 マス・横 11 マスで構成されており、黒マスはエージェントの通行できない障害物、白マスはエージェントの通行可能な道、黄色マス及び青マスはエージェントのスタート地点とゴール地点をそれぞれ表している。また、強化学習後、エージェントの行動価値 Q が高くなるほど道は赤くなる。

これらの最短経路問題の強化学習結果を比較し、転移知識の特徴から分類基準を求める。本研究では「スタートからゴールまでの経路長」「スタートから見たゴール方向」の 2 つを特徴として用いる事とした。選定理由として、これらの特徴が経路長の長短、ゴール方向の四方位というようにカテゴリとして表現しやすい事、これら 2 つの特徴に直接の相関性がない事が挙げられる。

この 2 つの特徴を基準としてランダムな環境で強化学習を行った知識群を分類し、表 4. 1 に示す 8 つのカテゴリに分類する。

表 4. 1:カテゴリ分類

カテゴリ	経路長	ゴール方角
1	平均以上	北
2	平均以上	東
3	平均以上	南
4	平均以上	西
5	平均以下	北
6	平均以下	東
7	平均以下	南
8	平均以下	西

この指標を用いることで転移知識をカテゴリに分類することが可能となる. その一方, 表 4. 1 のカテゴリ分類を行うためには転移知識として, Q 値の他にスタート地点とゴール地点の情報, エージェントが通行可能な道と通行できない障害物を表したマップ情報というような環境情報が必要となる. 一般に転移知識には状態 s における行動 a の場合の行動価値 Q 以外の情報は記録されていないため, この指標を転移知識に適用してカテゴリ分類することは出来ない. よって, Q 値のみが記録された転移知識をカテゴリ分類するため, 次節に示すニューラルネットワークを用いる.

4.4 ニューラルネットワークの構成

4.3 で述べた評価指標によるカテゴリ分類の問題は、 Q 値のみでカテゴリ分類を行えないことにあった。このことから、ニューラルネットワークを用いて Q 値からカテゴリを予想することにより、カテゴリ分類を行う必要がある。

事前に集めた最短経路問題の転移知識群と表 4.1 に示したカテゴリを訓練データとして、ニューラルネットワークを構築する。このニューラルネットワークにより、カテゴリ分類されていない転移知識をカテゴリ分類された転移知識（以下、カテゴリ化知識と呼称）として評価することが可能となる。また、カテゴリ分類にニューラルネットワークを用いることで、表 4.1 に示した特徴以外の隠れた特徴を抽出し、カテゴリ化の指標として利用することが期待できる。

提案手法で用いるニューラルネットワーク構成を図 4.2 に示す。入力層には行動価値 Q を用いる。本研究のタスクは 11×11 のグリッドフィールドであるため、入力層のユニット数は 121 個となる。中間層のユニット数は 500 個とした。また、出力層には 8 種類のカテゴリが出力されるため、出力層のユニット数は 8 個となる。

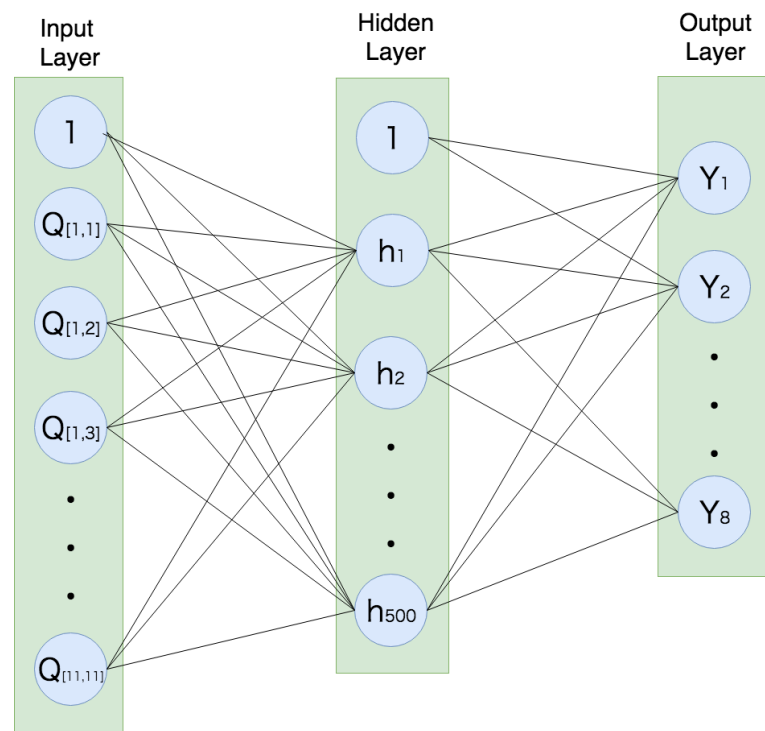


図 4.2: 提案手法で用いるニューラルネットワーク構成

活性化関数には ReLU 関数を用いる. ReLU 関数は入力を x , 出力を $\varphi(x)$ とする
と式(4. 1)として表される. Yann LeCun(2015)によると, 2015 年 5 月時点では,
ReLU 関数が最も高速で動作する活性化関数であると述べている. [5]

$$\varphi(x) = \begin{cases} 0 & (x \leq 0) \\ 1 & (x > 0) \end{cases} \quad (4. 1)$$

提案手法として用いるニューラルネットワークは以上の構成となる.

4.5 カテゴリを用いた転移学習手法

カテゴリを用いた転移学習手法を以下に示す.

1. 4. 4 で構築したニューラルネットワークにカテゴリ分類されていない転移知識を入力し, 出力としてカテゴリ化知識を得る
2. カテゴリに分類された学習環境(以下, カテゴリ化環境と呼称)を用意
3. カテゴリ化環境と同じカテゴリのカテゴリ化知識群の中から無作為に知識を選び転移学習を行う
4. カテゴリ化環境と関係なく, 転移知識群の中から無作為に知識を選び転移学習を行う

手順 4 と手順 5 の転移学習が収束するまでに掛かった総ステップ数を比較することで, カテゴリ分類が転移学習の効率に与える影響を検証可能であると考えられる.

4.6 データ収集

4.4 に示したニューラルネットワークの学習及び 4.5 に示したカテゴリを用いた転移学習で用いるデータの収集方法を行う。

4.6.1 転移知識の収集

以下の条件で強化学習による転移知識の収集を行う。

1. 学習タスクは最短経路問題とする
2. 学習環境は 11×11 のグリッドフィールドとする。学習環境を生成するアルゴリズムとして壁伸ばし法を用いる。壁伸ばし法とは全面が通路の状態から障害物を生成し伸ばすことで、迷路を作り上げるアルゴリズム[6]である
3. エージェントは上下左右と停止の合計 5 種類の行動が可能
4. Q 値の更新を行う際の学習率 α は 0.1, 割引率 γ は 0.99 とする
5. 報酬 r は 1.0 とする。報酬はエージェントがゴールに達した場合に与える
6. 行動選択手法にはボルツマン選択手法を用いる
7. 最大エピソード数は 500 エピソードとする

なお、本稿では、エージェントの一回の行動を 1 ステップ(step), スタートからゴールまでを 1 エピソードと表現する。

以上の条件で転移知識を収集した後、表 4.1 に示したカテゴリ分類によって転移知識のカテゴリを求める。この転移知識とカテゴリのセットをニューラルネットワークの訓練データとして用いる。

4.6.2 カテゴリ環境の収集

カテゴリ環境の収集手法を以下に示す.

1. グリッドフィールド環境を生成する. 条件は 4.6.1 と同様とする
2. 生成環境で強化学習を行う. 条件は 4.6.1 と同様とする
3. 強化学習の結果から表 4.1 に示したカテゴリ分類によってカテゴリを求める
4. 3. で求めたカテゴリを環境のカテゴリとする

以上の手法で, カテゴリを用いた転移学習に用いるカテゴリ環境を収集する.

4.7 結言

本章では, 知識の特徴からカテゴリ分類の指標を求める手法の提案と, カテゴリ分類が転移学習の効率に与える影響の検証方法をおこなった. 次章では, 提案手法を用いてニューラルネットワークを構築, カテゴリを用いた転移学習を行い, 結果の考察を行う.

第5章 実験

5.1 緒言

本章では, 提案手法を用いたニューラルネットワークの構築と, カテゴリによる転移学習の検証を行う.

5.2 提案手法を用いたニューラルネットワークの構築

4. 4 で示したニューラルネットワークの構築と構築の結果の検証を行う.

5. 2. 1 実験目的

Q 値のみが記録されたカテゴリ分類されていない転移知識をカテゴリズ知識として評価するため.

5. 2. 2 実験概要

ニューラルネットワークの学習条件は以下の通りである.

1. ニューラルネットワークの学習用ライブラリとして Chainer を用いる
2. データセットは 6000 個とする. うち 3840 個を訓練データ, 960 個を検証データ, 1200 個をテストデータとして用いる
3. 学習回数である Epoch 数は 1000 とする

また, カテゴリ毎のデータセットの内訳を表 5. 1 に示す.

表 5. 1:データセット内訳

カテゴリ	経路長	ゴール 方角	データセット数	データ割合[%] ([データセット合計] ÷ 8 × 100)
1	平均以上	北	638	10. 6
2	平均以上	東	581	9. 7
3	平均以上	南	574	9. 6
4	平均以上	西	587	9. 8
5	平均以下	北	928	15. 5
6	平均以下	東	927	15. 5
7	平均以下	南	881	14. 7
8	平均以下	西	884	14. 7
合計			6000	100

5. 2. 3 実験結果と考察

ニューラルネットワークの学習結果を次の図 5. 1 に示す. Accuracy は訓練データに対する学習精度, Validity は検証データに対する学習精度を表している. また, Epoch は学習回数を表している.

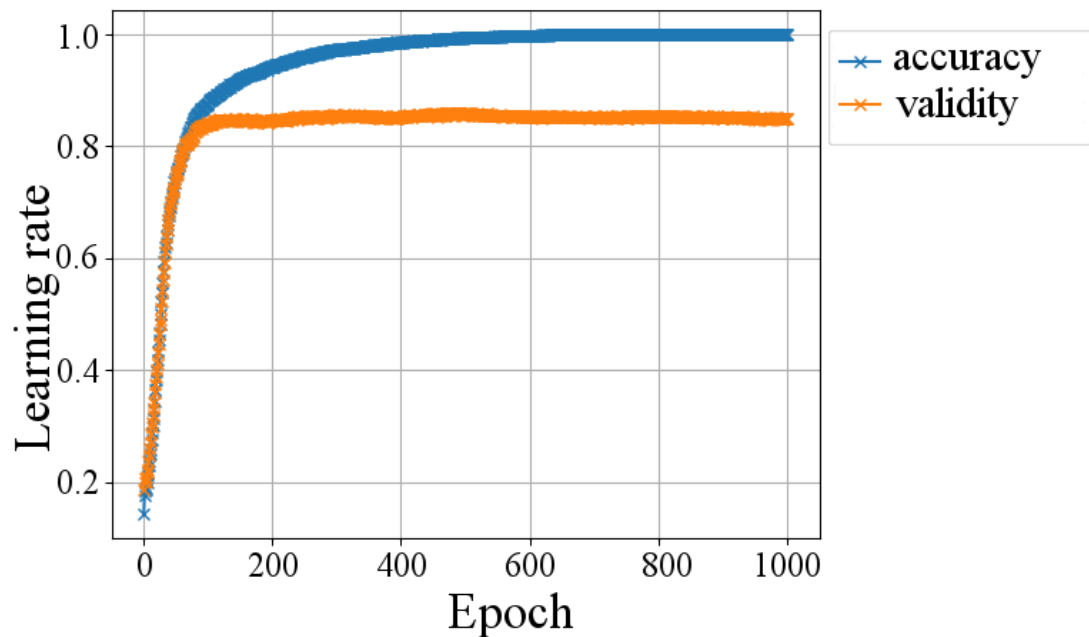


図 5. 1:ニューラルネットワークの学習結果

テストデータによる最終的な予想結果は86. 4%であった. これは86. 4%の確率で意図したカテゴリに分類可能である事を示している. グラフから, 500 Epoch の段階で学習が収束しており, その後のグラフに大きな乱高下がない事が確認できる. このことから, カテゴリ分類の精度は安定していると考えられる.

また, テストデータを用いて学習の結果を予測したところ, 結果は 86.4%となった. これは 86.4%の確率で意図したカテゴリ分類ができる事を示している. このことから, Q 値のみが記録された転移知識をカテゴリライズ知識に分類することが可能であると考えられる.

5.3 カテゴリによる転移学習の検証

4. 5 で示したカテゴリによる転移学習の検証を行う.

5.3.1 実験目的

カテゴリ化知識が転移学習に与える影響を検証するため.

5.3.2 実験概要

転移学習の条件は以下の通りである.

1. 学習タスクは最短経路問題とする
2. 学習環境は 4. 6. 2 に示した手法で収集したカテゴリ化環境を用いる
3. 転移知識は 5. 2 で示した学習済みニューラルネットワークにカテゴリ分類されていない転移知識を入力して得られたカテゴリ化知識を用いる
4. 学習環境に転移知識を反映させる転移率 β は 0. 1 とする
5. エージェントは上下左右と停止の合計 5 種類の行動が可能
6. Q 値の更新を行う際の学習率 α は 0. 1, 割引率 γ は 0. 99 とする
7. 報酬 r は 1. 0 とする. 正の報酬はエージェントがゴールに達した場合に与え, 負の報酬はエージェントが障害物に衝突するような行動を選択した場合に与える
8. 行動選択手法にはボルツマン選択手法を用いる
9. 最大エピソード数は 500 エピソードとする

学習回数は表 4. 1 に示した 8 種類のカテゴリにおいて、提案手法であるカテゴリ環境と同じカテゴリのカテゴリ知識群の中から無作為に知識を選んだ場合の転移学習を 50 回、カテゴリ環境と関係なく転移知識群の中から無作為に知識を選んだ場合の転移学習を 50 回行う事とする。

また、比較の基準として平均総ステップ数 $\overline{S_{(c, X)}}$ を用いる。平均総ステップ数

$\overline{S_{(c, X)}}$ は式(5. 1)で算出される。

$$\overline{S_{(c, X)}} = \frac{1}{50} \sum_{i=1}^{50} S_i \quad (5. 1)$$

c はカテゴリ、 X は知識選択手法、 i は転移学習の回数、 S_i は i 回目の転移学習の総ステップ数とする。

知識選択手法 X として、提案手法であるカテゴリ環境と同じカテゴリのカテゴリ知識群の中から無作為に知識を選んだ場合の転移学習を行った結果を A 、カテゴリ環境と関係なく転移知識群の中から無作為に知識を選んだ場合の転移学習を行った結果を B とする。これを用いて、カテゴリ c において知識選択手法 A を用いた場合の平均総ステップ数を $\overline{S_{(c, A)}}$ 、知識選択手法 B を用いた場合の平均総

ステップ数を $\overline{S_{(c, B)}}$ と表す。また、カテゴリ全体で知識選択手法 A を用いた場合の平

均総ステップ数を $\overline{S_A}$ と表す。同様に、カテゴリ全体で知識選択手法 B を用いた場合の平均総ステップ数を $\overline{S_B}$ と表す。

5.3.3 結果と考察

カテゴリによる転移学習結果を表 5. 2, 図 5. 2 に示す.

表 5. 2, 5.3 及び図 5. 2 から, カテゴリ1~4では $\overline{S_{(c, A)}}$ は $\overline{S_{(c, B)}}$ よりも多く, カテゴリ5~8では $\overline{S_{(c, A)}}$ は $\overline{S_{(c, B)}}$ よりも少なくなる事が分かる. これは, 経路長が平均より長い場合, カテゴリによる転移学習は無作為に知識を転移した転移学習と比べて学習効率が悪く, 経路長が平均より短い場合, カテゴリによる転移学習は無作為に知識を転移した転移学習と比べて学習効率が優れている事を示している. この結果から, 最短経路問題におけるカテゴリを用いた転移学習では, 経路長の長さによって学習効率が異なると考えられる.

一方, ゴール方角による学習効率の変化に大きな差は見られない. 例として, ゴール方角が同じ東の方角である $\overline{S_{(2, x)}}$ と $\overline{S_{(6, x)}}$ を比較すると, $\overline{S_{(6, x)}}$ は全カテゴリの中で最も平均総ステップ数が減少しているが, $\overline{S_{(2, x)}}$ は全カテゴリの中で最も平均総ステップ数が増加している.

これらのことから, カテゴリによる転移学習はカテゴリ分類の条件によって転移学習結果が大きく変化することが分かる.

また, カテゴリ全体の平均総ステップ数 $\overline{S_A}$ 及び $\overline{S_B}$ を表 5. 3 に示す. 表 5. 3 から, $\overline{S_A}$ は $\overline{S_B}$ と比べて多い事が分かる. これは, カテゴリによる転移学習は無作為に知識を転移した転移学習と比べて全体的な学習効率が優れている事を示している. また, 提案手法を用いたことによる学習総ステップの減少量は2855である. これは $\overline{S_B}$ の4%にあたる. 4%という値は減少量として小さいが, ニューラルネットワークでの転移知識の分類が 86.4%, つまり 13.6%の誤認識が生じている事を加味した上でも全体として学習時間が短くなるという結果が得られた.

表 5. 2: カテゴリ毎の平均総ステップ数

カテゴリ	知識選択手法	表記	平均総ステップ数 $\overline{S_{(c, x)}}$
1	A	$\overline{S_{(1, A)}}$	75461
1	B	$\overline{S_{(1, B)}}$	71578
2	A	$\overline{S_{(2, A)}}$	224242
2	B	$\overline{S_{(2, B)}}$	115889
3	A	$\overline{S_{(3, A)}}$	95370
3	B	$\overline{S_{(3, B)}}$	75985
4	A	$\overline{S_{(4, A)}}$	126203
4	B	$\overline{S_{(4, B)}}$	66810
5	A	$\overline{S_{(5, A)}}$	13076
5	B	$\overline{S_{(5, B)}}$	60589
6	A	$\overline{S_{(6, A)}}$	12145
6	B	$\overline{S_{(6, B)}}$	77964
7	A	$\overline{S_{(7, A)}}$	14089
7	B	$\overline{S_{(7, B)}}$	52944
8	A	$\overline{S_{(8, A)}}$	23228
8	B	$\overline{S_{(8, B)}}$	50132

表 5. 3: カテゴリ毎の減少数

カテゴリ c	$\overline{S}_{(c, A)}$	$\overline{S}_{(c, B)}$	減少数[step] $\overline{S}_{(c, B)} - \overline{S}_{(c, A)}$	減少率[%] $\frac{(\overline{S}_{(c, B)} - \overline{S}_{(c, A)})}{100}$
$\overline{S}_{(1, x)}$	75461	71578	-3883	-5.42
$\overline{S}_{(2, x)}$	224242	115889	-108353	-93.5
$\overline{S}_{(3, x)}$	95370	75985	-19385	-25.5
$\overline{S}_{(4, x)}$	126203	66810	-59393	-88.9
$\overline{S}_{(5, x)}$	13076	60589	47513	78.4
$\overline{S}_{(6, x)}$	12145	77964	65819	84.4
$\overline{S}_{(7, x)}$	14089	52944	38855	73.4
$\overline{S}_{(8, x)}$	23228	50132	26904	53.7

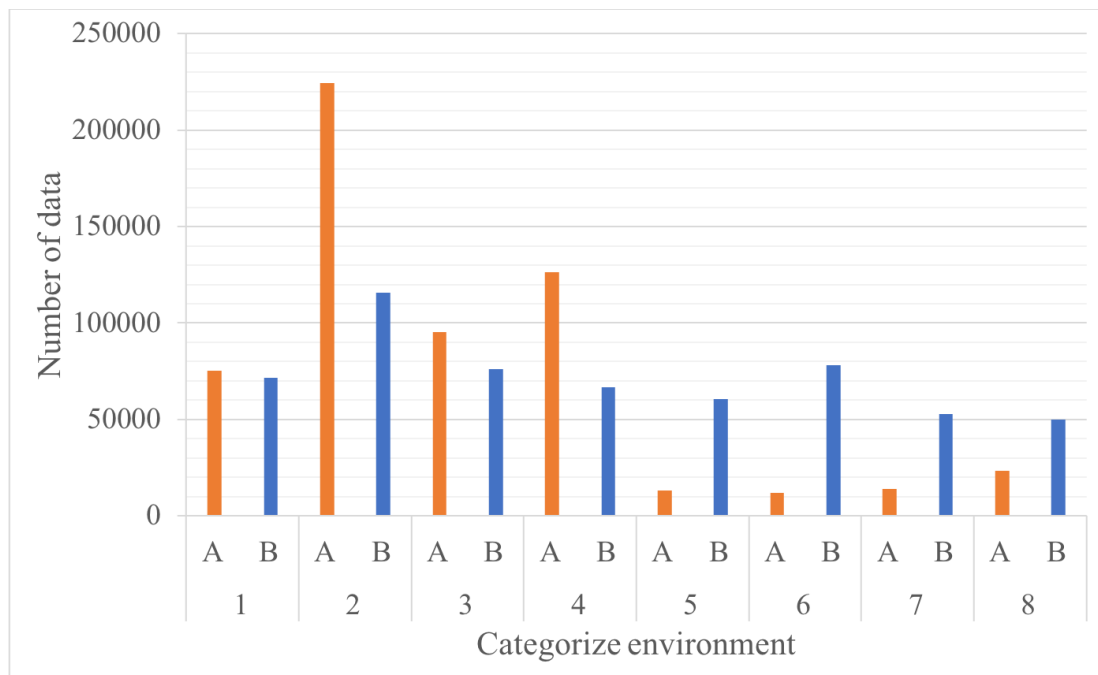


図 5. 2: カテゴリを用いた転移学習結果

表 5. 4: カテゴリ全体の平均総ステップ数

知識選択手法	表記	平均総ステップ数 \overline{S}_x
A	\overline{S}_A	68631
B	\overline{S}_B	71578

5.4 結言

本章では提案手法を用いたニューラルネットワークの構築結果を述べ, カテゴリによる転移学習と無作為に知識を転移した転移学習の結果の比較を行った.

ニューラルネットワークを用いることで, Q 値のみが記録された転移知識からカテゴリライズ知識を得ることが可能である. よって, 図 4. 1 に示したような強化学習後の環境と Q 値の知識からカテゴリ分類の指標を求める事が可能であれば, 環境情報を削り, Q 値のみのデータとなった知識からでもカテゴリ分類が可能であると考えられる.

カテゴリによる転移学習と無作為に知識を転移した転移学習の結果の比較から, カテゴリによる転移学習はカテゴリの条件によって結果が大きく異なる事が分かる. この事から, カテゴリによる転移学習はカテゴリ分類の質によって転移学習結果が大きく変化すると考えられる.

第6章 結論

6.1 本論文の結言

本論文では、まず、ロボットの代表的な学習方法である強化学習と転移学習の説明を行い、その課題点を述べた。強化学習の課題点は、試行錯誤による行動回数の増加であることから、転移知識を用いた転移学習により解決可能であるものの、転移先の環境が転移知識の環境と大きく異なる場合、転移知識が却って学習の妨害をしてしまうということが分かった。この問題を解決するため、先行研究では転移知識をカテゴリに分類し、認知心理学の知見である活性化拡散モデルを適用することで複数の転移知識を利用して転移先の環境に合わせた転移学習手法が提案された。この手法により、複数の知識を用いた転移学習は強化学習や単一方策のみの転移学習と比較して有用である一方、学習の効率は転移知識のカテゴリ分類に依存する事がわかった。従って、効率の良い学習を行うためには転移知識をカテゴリに分類する際の指標を求める事が必要であることから、カテゴリ分類の評価指標を導出し、カテゴリを用いた転移学習を行うことで、転移学習の効率化を図る提案を行った。そして、カテゴリを用いた転移学習とカテゴリを用いない転移学習を比較することで、カテゴリによる転移学習はカテゴリの条件によって結果が大きく異なる事が分かった。

6.2 今後の展望

提案手法のカテゴリの評価指標はタスクとしてグリッドフィールド上における最短経路問題を扱うことが前提であるため、今後はより一般的な評価指標を導出する方法の検討を行う必要がある。

謝辞

本研究において、ご指導頂いた東京電機大学 工学部 情報通信工学科 鈴木剛教授に深く感謝申し上げます。1年間、研究に関する助言、方針の提示等のご指導をしてくださいました。

鈴木研修士課程の皆様に感謝申し上げます。予稿の添削や発表のアドバイスをいただきました。

特に修士課程2年の高桑優作氏に感謝申し上げます。高桑氏の研究が私の研究テーマの発端となっているのはもちろんのこと、研究のアプローチについて相談していただきました。

鈴木研学部4年の皆様に感謝します。研究室での生活を楽しく過ごすことができました。

最後に4年間お世話になった東京電機大学と、東京の私立大学に通わせてくれた家族に心より感謝申し上げます。

参考文献

- [1] 堀江奈保子, “少子高齢化で労働力人口は 4 割減”, みずほインサイト, p1(2017)
- [2] 日本経済再生本部, “ロボット新戦略”, ロボット革命イニシアティブ協議会, p6-p7(2015)
- [3] 高桑優作, “活性化拡散モデルに基づく強化学習エージェントの方策選択手法”, ロボティクス・メカトロニクス講演会, 2017
- [4] A.M.Collins, E.F.Loftus, “A Spreading-Activation Theory of Semantic Processing”, Psychological Review, Vol. 82, No. 6, pp. 407-428, 1975
- [5] Yann LeCun, Leon Bottou, Genevieve B.Orr, Klaus-Robert Muller, “Efficient BackProp”, Neural Networks: Tricks of the Trade, pp. 9-48, 1988.
- [6] 平石 真也, 郭 清蓮, “文法と迷路を融合したデジタルコンテンツに関する研究”, 情報処理学会研究報告, vol. 2013-MBL-68, No. 20, pp. 3, 2013