平成２１年度

卒 業 研 究 論 文

題　　目　環境認識能力の変化が学習に及ぼす影響について

提 出 者　　室蘭工業大学 情報工学科

氏　名　　中南　義典

学籍番号　1823066

提出年月日　　平成 22 年　2 月 12 日

室蘭工業大学
情報工学科

目次

第 1 章　はじめに

1.1　研究背景

近年，技術の発達によりロボット開発の分野は著しい発達を遂げている．ロボットの活躍は工業等での産業用ロボットが主であるが，最近では一般の家庭や施設にもロボットが普及し始めている[1]-[3]．さらに、これからは人間の日常生活をより良いものにするための活躍が期待され，そのための研究が進められている[4][5]．

ロボット技術が発達した理由のひとつに，機械学習[6]の発展がある．機械学習とは，自身の経験を知識として蓄え，自律的に行動を獲得できるというものである．この機械学習を用いることで，より幅広い環境への適応が可能となっている．

ロボットの学習の従来研究は，学習の手法に注目するものが多かった．効率の良い学習を行うためのアルゴリズムがいくつも開発され，その有効性の検証が行われてきた[7]-[9]．このとき，ロボットの身体構造の違いを考慮して学習手法の有効性を検証している研究は少ない．しかし近年，ロボットの身体構造に注目し，身体構造を改良することで学習の効果を高める研究がいくつか行われている[10]-[13]．つまり，学習手法が同じであっても，身体構造が違えば学習の効果が異なる可能性がある．

ここで，ロボットの身体構造と学習との関係を考える．まず，ロボットは主に以下の 3 つの身体構造により構成されている．

- センサ
- 内部構造
- アクチュエータ

これらの構造が学習にどのように関わるのかを示したのが図 1.1 である．ロボットはまず，センサによって自分の置かれている状況を認識する．次に，認識した情報を元に内部構造によって学習を行い，行動を選択する．最後に，選択された行動をアクチュエータによって実行する．その結果，変化した状況を再びセンサで読み取る，という流れを繰り返して学習を行っている．

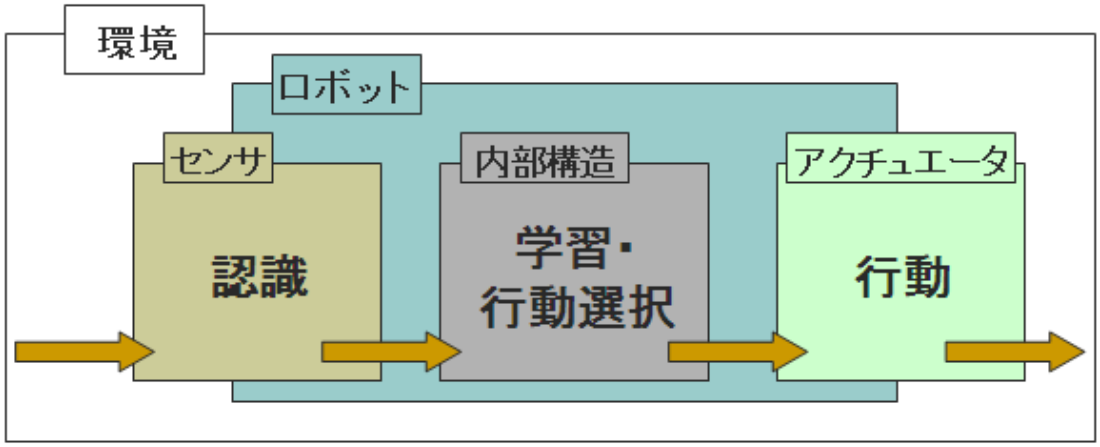

図 1.1 ロボットの学習の様子

このように学習は，センサからの情報とアクチュエータによる行動結果を基に，内部構造で知識を蓄えることで行われている．そのため，学習の効果はセンサやアクチュエータの性能にも依存する．センサから受け取る情報が違えば，学習の元となる情報も違うことになる．また，アクチュエータが行える行動の種類が違えば得られる行動の結果も違ったものとなる．そういった場合には，同じ学習手法を用いていても学習効果に違いが生まれる可能性がある．

従来研究におけるロボットの設計は，研究者が研究に必要なセンサ等を自分で考えて設計する．しかしこのとき，技術面の問題や研究者の考え方の違い等から，研究毎に異なった設計となる．そのため，身体構造の違いを考慮せずに学習手法の有効性を検証すると，正確な検証結果とならない可能性がある．

1.2 研究目的

前節で述べたように，従来研究におけるロボットの設計は研究毎に異なったものとなっている．しかし，従来研究における学習手法の有効性の検証は身体構造の違いについてほとんど触れられないまま行われてきた．そのため，従来研究における学習手法の有効性の比較は正しく行われていない可能性がある．そこで本研究では，身体構造の違いが学習に与える影響を検証することを大きな目的とする．

前節で述べたように，身体構造は大きく三つに分けることができる．その中でも本研究ではセンサのみに注目した．同じ学習手法とアクチュエータを持つロボットについて，センサ能力のみを変化させたとき，学習にどのような影響を及ぼすかの検証を行う．

1.3 本論文の構成

以下に本論文の構成を述べる

第 2 章では，本論文で用いる環境やセンサについての定義を行う．

第 3 章では，本論文で用いる強化学習について説明した後，センサと学習の関係性について述べる．また，それらをふまえて，環境認識能力の違いが学習に及ぼす影響の検証方法を述べる．

第 4 章では，第 3 章で述べた検証方法に基づいて行った実験について述べる．また，それらの結果を示し，その結果から考察を行う．

第 5 章では，本論文のまとめとして，本論文全体について考察を行う．また，本論文では扱えなかった将来の研究課題について述べる．

第 2 章　ロボットのセンサと環境との関係

2.1　環境の定義

本論文では，センサの種類数の変化が学習に与える影響について，一般的な環境に対して検証を行う．そのため，一般的な環境について定義を行う必要がある．そこで，本論文で用いる環境について，以下のように定義する．

(1)　実際の環境は，光や音といったいくつもの要素から構成されている．本論文で用いる環境は，エージェント外部の全ての要素から構成される．

- エージェント外部の要素の数を N 個とすると，その環境は N 個の要素 E_i（i＝1～N）から構成される環境となる．

(2)　環境を構成する要素 E_i はそれぞれ要素の値を持つ．ここで E_i は離散値とする．

(3)　環境は，環境を構成する要素の組み合わせによって，異なる状態を取る．

- N 種類の要素によって構成される環境の場合，環境は N 次元の状態空間を持つ．N=3，E_1＝{0,1,2,3}，E_2＝{0,1}，E_3＝{0,1,2}の場合の状態空間を図 2.1 に示す
- E_i の取り得る状態の個数を $S(E_i)$とすると，環境の取り得る状態の数 S(E)は次の式で表される

$$S(E) = \prod_{i=1}^{N} S(E_i) \tag{2.1}$$

(4)　環境の状態は，時間の経過や外的要因などによって変化する．エージェントの行動も外的要因であり，環境の状態に変化を起こす．

- 時刻 t における環境の状態を s_t とすると，エージェントの行動 a_t により，環境の状態は s_{t+1} へと遷移する．
- 環境の遷移はノードとリンクによって表される．

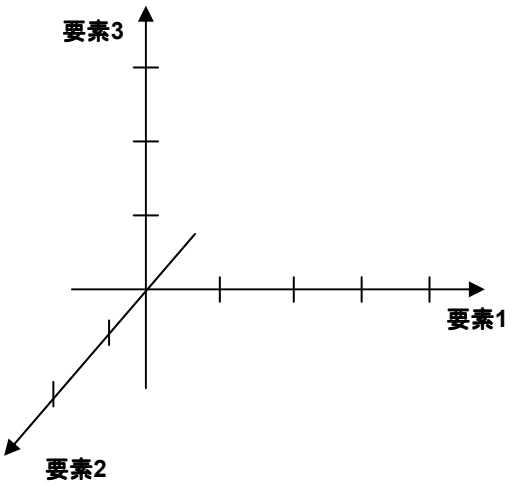

図 2.1 N=3，Vi={0,1}の時の状態空間

本論文で用いる環境の例として，N=2，E_i＝{0,1}，i={1,2}の環境について具体例を用いて考える．この環境は，音(=E_1)と光(=E_2)という 2 種類の要素を持つとする．ここで，音と光は OFF（=0）と ON(=1)の二つの値を取ることとする．また，エージェントは各状態において「音のスイッチを切り替える」「光のスイッチを切り替える」という 2 つの行動のどちらかを選択することができる．この環境の状態空間を，図 2 に示す．

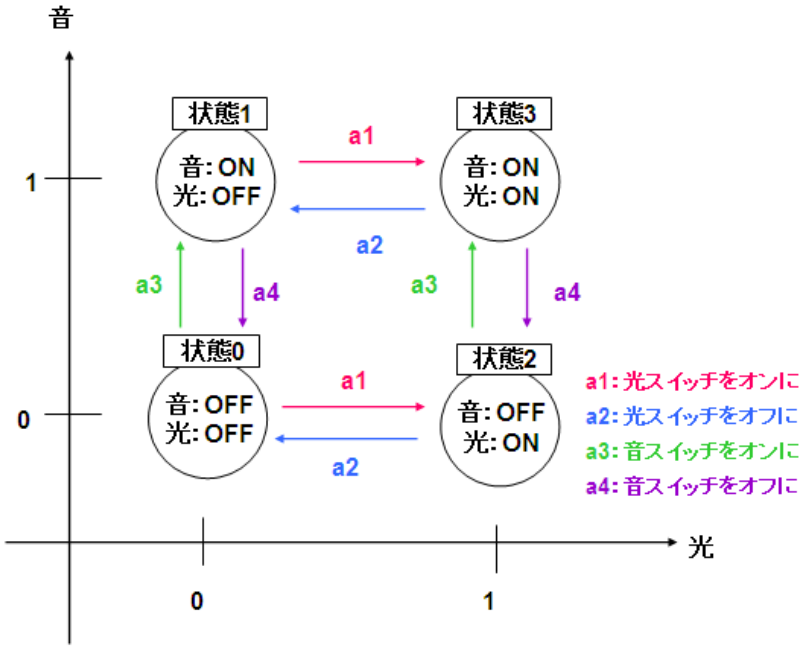

図 2.音と光で構成された環境の状態空間

このとき環境は，状態 0～3 の 4 種類の状態を取ることができる．ここで，エージェントが現在置かれている状態が s_t=状態 0 だとする．エージェントが行動 a1 を選択し，光スイッチをオンにしたとすると，エージェントは s_{t+1}=状態 2 へと推移する．
以上のような環境を，本論文では，エージェントが学習を行う環境として用いる．

2.2 センサの定義

本論文では，センサの種類数の変化が学習に与える影響について検証することを目的としている．そのため，具体的なセンサではなく，抽象的なセンサを対象として検証を行う必要がある．そこで，抽象的なセンサについて，以下のように定義する．

(1) 本論文で用いるセンサは，環境を構成する要素 1 種類に対し，1 つのセンサを用いて認識可能とする．つまり，環境を構成する要素が N 種類存在する場合，環境を正しく認識するためには最低でも N 個の対応センサが必要となる．このとき，要素 1 種類に対して認識可能なセンサが 1 つだけとは限らない.
(2) エージェントは，自身の持つセンサを通して，要素の値を認識する．
(3) エージェントが複数のセンサを持つ場合，各センサが認識できる値の範囲は同じであるとする．

ここで，N=3，E_i ={0,1}(i=1~3)のときの例を図 2.2 に示す.

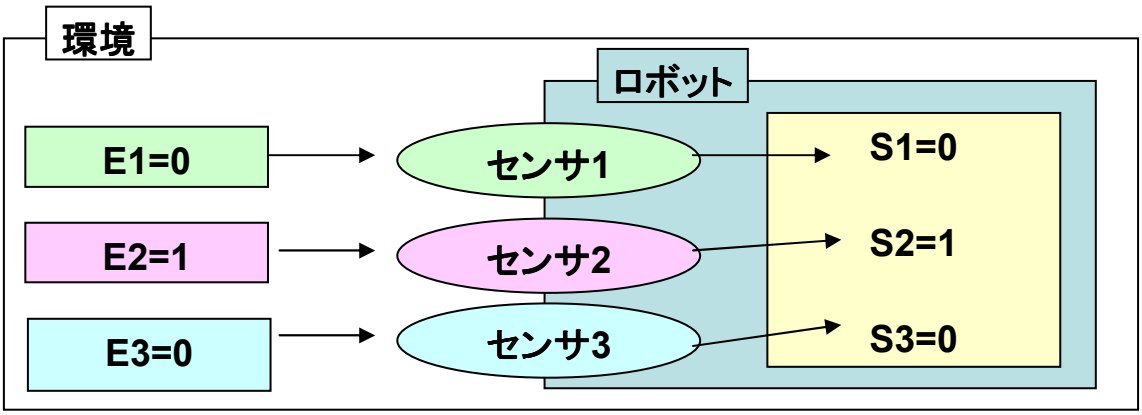

図 2.2 N=3 のときの環境認識の例

図中のセンサ 1～3 はそれぞれ要素 E_1～E_3 を認識するセンサ, S_1～S_3 は各センサが認識した値である．

このときエージェントは，センサ 1 を通じて要素 E_1 の状態を認識することが出来る．同

じようにセンサ 2 が要素 E_2，センサ 3 が要素 E_3 の状態を認識する．そして，それらの状態の組み合わせにより，環境の状態を認識する事が出来る．

2.3　センサと環境の関係

エージェントは，自身のセンサを通して環境の状態を認識する．人間で言えば，視覚や聴覚といった器官を通して環境を認識している，ということである．そのため，どのように環境を認識するかはセンサの種類や性能に大きく依存する．しかし，実際の環境は,エージェントの持つセンサとは無関係に存在している．

ここで，環境とセンサの関係を図 2.3 に示す．まず始めに，様々な要素から構成される環境が存在している．その中で，ある要素のある状態を認識する事ができるのがセンサである．このように，エージェントが認識できるのは環境の一部分でしかない．そのため，エージェントが持つセンサだけでは認識できない要素が実環境に存在している可能性がある．また, 認識できたとしても, センサの性能面から正確な認識が出来ない場合も考えられる．以上のように，エージェントが認識した環境と実環境の間には，差異が生まれるのが一般的である．

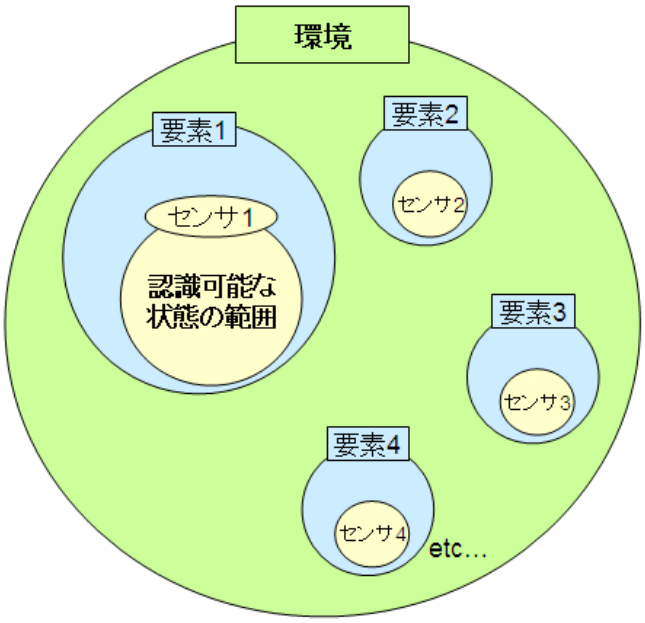

図 2.3　環境とセンサの関係

認識した環境と実環境との間に差異が生まれる例を，N=3，Ei={0,1}(i=1~3)の環境を用いて示す（図 2.4）．ここで，環境の要素 E_1 を認識できるセンサ 0 と，要素 E_2 を認識できるセンサ 2 の，二つのセンサを持ったエージェントを考える．このエージェントは要素 E_1 と

E_2 を認識することができるが，要素 E_3 を認識することはできない．このため，エージェントはこの環境を二つの要素からなる環境であると認識してしまい，実際の環境とは違いが生まれる．

こうした差異は，センサの種類が少ない等，センシング能力が低いほど大きくなるものと考えられる．

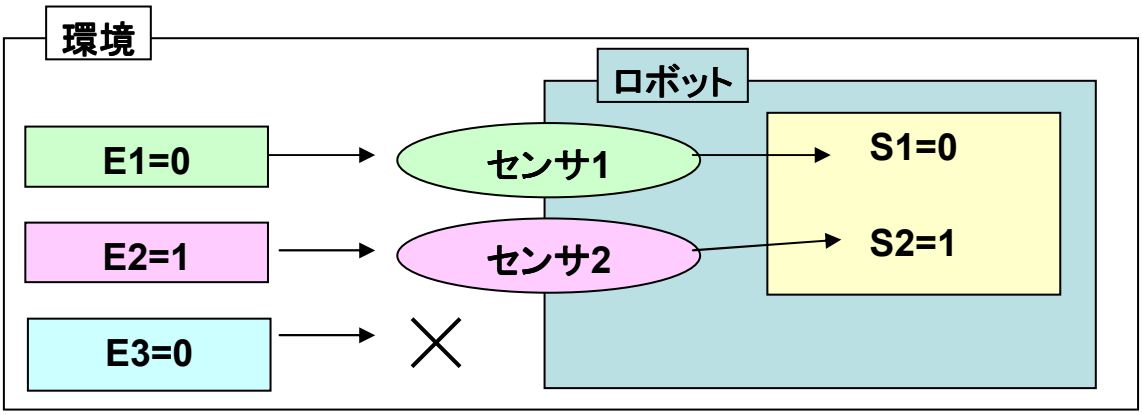

図 2.4　認識した環境と実環境との差異

2.4　環境認識と学習の関係

機械学習は，センサを通して認識した環境の情報を元に行われる．例えば強化学習であれば，図 2.5 に示すように，認識した状態に対して各行動の価値を知識として蓄えていく．このため，センサによって認識される情報，すなわち知覚環境が違えば，学習の元となる情報も違うものとなる．そのため，実環境と知覚環境に差異があれば，学習にも差異が生まれると考えられる．実環境と知覚環境との差異はセンシング能力の差異によって生まれることは前節で述べた．そのため，センサ能力に差異があれば，学習にも差異が生まれると考えられる．

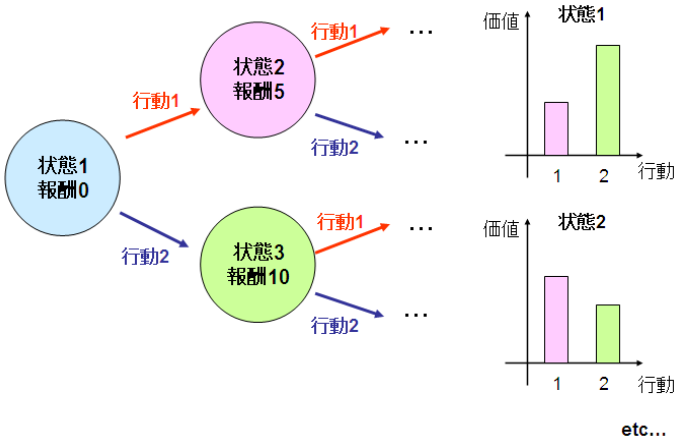

図 2.5　強化学習における状態と行動の利用の仕方

第 3 章　環境認識能力と学習との関係の検証方法

本章では，学習と環境認識能力との関係の検証方法について述べる．また，その検証のために本実験で用いる学習手法について述べる.

3.1　強化学習

本節では，本実験で用いる学習手法である強化学習[14]について述べる.

3.1.1　強化学習の概要

強化学習は，学習者であるエージェントが環境内で行動を起こすことで報酬を獲得する．報酬とは，設計者が目標の代わりに設定するものである．エージェントは報酬の獲得量を最大にすることを目的とし，試行錯誤を繰り返して学習を行う.

強化学習の特徴として，以下の二つが挙げられる.

- 学習に際して，正解が与えられない（教師無し学習）

  強化学習では，「何をすべきか」という目的を与えなくても，各行動に対して報酬を与えれば学習を行うことができる．また，どのように行動していけばいいか，その過程をあらかじめ明確にする必要が無い．つまり強化学習では，報酬さえ与えれば，学習の目的も過程もエージェントが獲得していくことができる.

- 遅延報酬に対応できる

  前述したように，強化学習では報酬を用いて学習を行う．この報酬には，即時報酬と遅延報酬の二種類が存在する．即時報酬とは，ある行動に対してその場で与えられる報酬である．一方遅延報酬とは，ある行動からさらに 2 回，3 回と行動を重ねたときに得られる報酬である．この遅延報酬に対応するためには，ある行動に対して即時報酬だけでなく長期的に得られる報酬を考慮に入れて学習を行う必要がある．そのようなタスクにも対応できるのが強化学習の特徴のひとつである.

強化学習の流れを図 3.1 に示す.

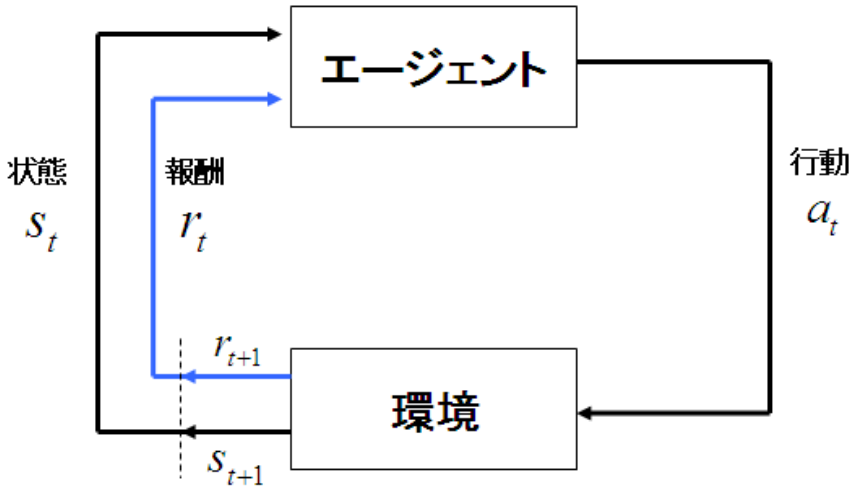

図 3.1　強化学習の流れ

ここで t は現在の時刻，s_t は時刻 t においてエージェントの置かれている状態，r_t は時刻 t においてエージェントが得た報酬，a_t は時刻 t においてエージェントが選択した行動を示す．

エージェントはまず，センサを通して状態 s_t を知覚する．次に，エージェントが選択可能な行動の中から行動 a_t を選択して実行する．するとエージェントは，実行した結果の善し悪しを報酬 r_t という数値で受け取る．この報酬を元にしてエージェントは学習を行う．また，これと同時に状態 s_t は状態 s_{t+1} へと遷移する．

3.1.2　強化学習の構成要素

強化学習は次の二つの要素によって構成される．

- 行動選択部

  エージェントが行う行動を決定するのが行動選択部である．エージェントは行動学習部で得た知識を元に行動を選択する．このとき，知識をどのように利用して行動選択を行うかは行動選択手法によって決まる．

- 行動学習部

  エージェントが実際に学習を行い，知識を蓄えるのが行動学習部である．エージェントがある状態である行動を行ったとき，その状態と行動の組み合わせの価値を知識として蓄える．この価値をどのように推定するかを決めるのが行動学習手法である．強化学

習におけるタスクには，遅延報酬が存在することは前節で述べた．状態と行動の組み合わせの価値は，この遅延報酬を考慮して推定する必要がある．そのため強化学習では，どのように価値を推定するかという行動学習手法が非常に重要となる．

3.1.3 行動選択手法

本研究に使用した行動選択手法について述べる

・ softmax

softmax 法は，推定される行動価値に基づいた確率で行動を選択するという行動選択手法である．一般的に Gibbs 分布，もしくは Bltzmann 分布に基づいて行動を選択する．具体的には，t 回目の試行における行動 a について行動価値 $Q(s_t,a_t)$が与えられたとき，以下の式（3.1）によって行動 a を選択する確率 π （s_t,a_t）を決定する．

$$\pi(s_t,a_t)=\frac{e^{Q(s_t,a_t)/\tau}}{\sum_{b=1}^{n} e^{Q(s_t,b)/\tau}} \tag{3.1}$$

ここで τ は，温度と呼ばれる正定数である．温度が低いほど，推定した価値が高い行動を選びやすくなり，温度が高いと，推定した価値が低くても選択される確率が高くなる．

3.1.4 行動学習手法

本研究で用いた行動学習手法について述べる．

・ Q 学習

Q 学習は，現在選択した行動の価値と，行動によって遷移した先の状態の状態価値の二つを用いて，現在の行動価値を更新していく学習法である．また，ある方策に基づいて行動しながら，最適方策を学習するという特徴を持つ．例えば softmax 法を用いた場合，softmax 法に基づいた行動選択を行いながら，実際には最適方策を学習する．

Q 学習における行動価値の推定は，1 ステップ毎に式（3.2）を用いて行われる．

$$Q(s_t,a_t)\leftarrow Q(s_t,a_t)+a\left[r_{t+1}+\gamma\max_a Q(s_{t+1},a)-Q(s_t,a_t)\right] \tag{3.2}$$

ここで，s_t は現在の状態，a_t は採用した行動，r_{t+1} は行動によって得られた報酬を示

す．また，s_{t+1} は行動後の新しい状態，a は新しい状態において最も価値の高い行動を示す．$Q(s_t, a_t)$は状態 s_t における行動 a_t の行動価値推定を示し，α $(0<\alpha<1)$は学習率，γ $(0<\gamma<1)$は割引率を表す．

3.2 センサと学習の関係

強化学習は，認識した環境の情報を元に行われる．そのため，センサによって認識される情報，すなわち知覚環境が違えば，学習に大きな影響があると考えられる．実環境と知覚環境との差異はセンシング能力の差異によって生まれることは前章で述べた．そのため，センシング能力の違いは学習に大きな影響を与えると考えられる．

エージェントのセンシング能力を決める要素はセンサの種類数やセンサの分解能など，いくつかの要素によって決定する．そのため，センシング能力に関わる各要素について，それぞれの性能の違いが環境認識や学習にどのような影響を与えるかを考える必要がある．そこで，3.3 ではセンシング能力を決定する要素を 3 つにわけ，それぞれが環境認識や学習に与える影響について考察する．また，3.4 では，エージェントが冗長性のあるセンサを持つときに環境認識や学習に与える影響について考察する．

3.3 センサ能力の不足が学習に与える影響

実環境と知覚環境との間に差異が生まれる原因のひとつとして，センシング能力の不足が考えられる．センシング能力を決定する要素としては，大きく分けて以下の三つが挙げられる．

- センサの種類数
- 各センサの分解能
- 各センサのサンプリング周波数

これら三つの要素によって，実環境と知覚環境との間にどのような差異が生まれるかを考える．また，その差異が学習効率に与える影響について考察する．

3.3.1 センサの種類数

センサの種類数とは, 環境の要素を認識できるセンサの数である. 環境の要素の数を N, センサの種類数を M とすると，環境を正確に認識するためには，最低でも N 種類のセンサが必要である．センサの種類数が N より少ない場合，(N－M)個の要素を認識することができなくなる．

センサの種類数が少ない場合，認識した環境と実環境との間にどのような差異が生まれるか考える．一般的には，認識できる状態の数が違う，確定的な遷移を確率的なものと認識する，等の違いが生まれるといったことが考えられる．その具体例のひとつとして，本来確定的な遷移先を，センサの不足によって確率的な遷移先であると認識する場合について示す．まず，N=2，Ei={0,1,2}(i=1,2)の環境を考える．この環境はエージェントの行動により，図 3.2 で示すように遷移するものとする．ここで，2 体のエージェント A・B を考える．エージェント A は，各要素を認識する 2 種類のセンサを持つとする．一方エージェント B は，センサをひとつしか持たず，要素 E_1 のみを認識できるものとする．この 2 体のエージェントが，それぞれどのように環境を認識するかを考える．

2 体のエージェントが，E1=0 の状態から次の状態へ遷移する場合を考える．要素 E1=0 のとき，取り得る状態は E2=0 と E2=2 の二つが考えられる．このときエージェント A は図 3.3 に示すように，二つの状態が別々のものであると認識する．また，E1=0，E2=0 のときは E1=1，E2=0 へ，E1=0，E2=2 のときは E1=2，E2=1 へ確定的に遷移するものであると認識する．これに対しエージェント B は図 3.4 に示すように，現在の状態が E2=0 か E2=2 か判断することができない．そのため，E1=1 であれば同じ状態であると判断してしまう．また，この状態で次の遷移先を判断するため，E1=0 の遷移先は E1=1 か E1=2 のどちらかであり，確率的に遷移するものであると認識する．

センサの種類数が不足している場合，認識した環境と実環境との間に，以上のような差異が生まれる．

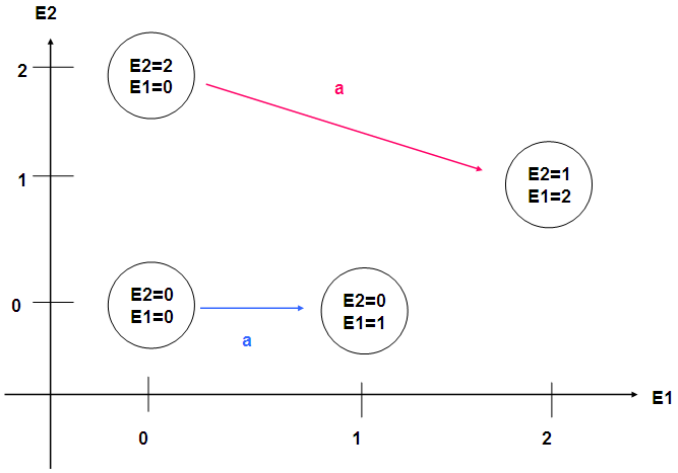

図 3.2　実環境の状態遷移

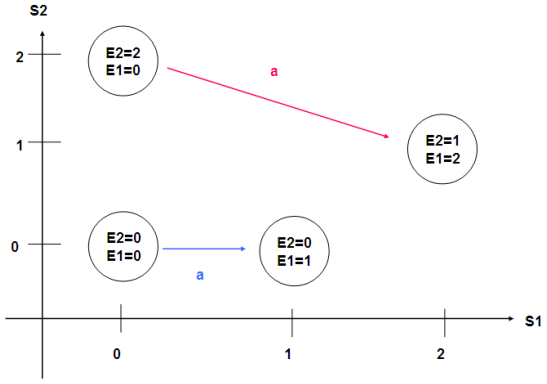

図 3.3　エージェント A が認識した環境

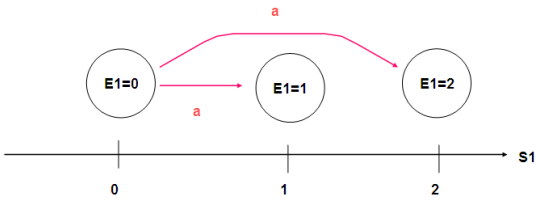

図 3.4 エージェント B が認識した環境

3.3.2　センサの分解能

センサの分解能とは，環境の要素の値をどれだけ分割して読み取れるか，という能力である．より多くの値に分割できるほど，センサの分解能の能力は高くなる．環境の要素 E_i は V 個の値に分割できるとし，エージェントのセンサの分解能を R とすると，環境の値を正確に認識するには，R が最低でも V でなくてはならない．R が V より小さい場合，環境の要素の値に追従しきれず，値が違う状態でも同じ状態であると認識してしまう可能がある．

例として，N=1，V1={0,1,2,3,4}の環境を考える．この環境は図 3.5 に示すように状態遷移するものとする．ここで分解能 R＝5，つまり S1＝{0,1,2,3,4}を読み取れるエージェント A と，R＝3，S1＝{0,2,4}のエージェント B を考える．エージェント A は，図 3.6 で示すように値の変化全てを正確に認識することができる．そのため，E1=0 からは E1=2 へ，E1＝1 からは E1=4 へ確定的に遷移するものであると認識できる．これに対しエージェント B は，

2 刻みでしか値を認識できない．そのため，図 3.7 に示すように，E1=0 と E1=1，E1=2 と E1=3 の違いを認識することができず，同じ状態であると認識してしまう．そのため，E1＝0 の状態からは，E1＝2 と E1=4 のどちらかへ，確率的に遷移するものであると認識してしまう．

センサの分解能の性能が不足している場合，認識した環境と実環境との間に，以上のような差異が生まれる．

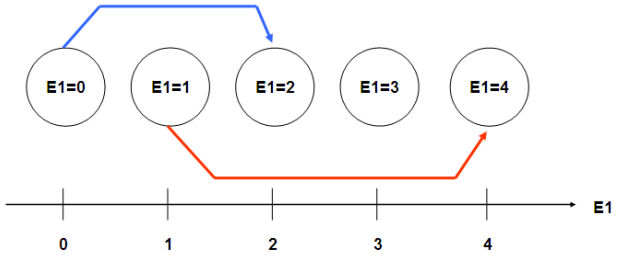

図 3.5　実環境の状態遷移

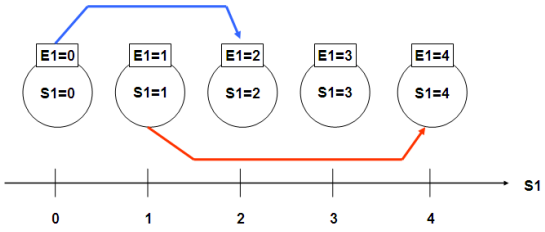

図 3.6　エージェント A が認識する環境

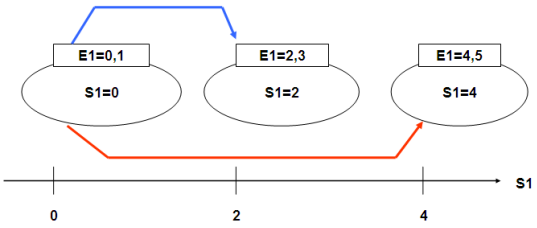

図 3.7　エージェント B が認識する環境

3.3.3　センサのサンプリング周波数

センサのサンプリング周波数とは，センサが環境の状態を認識する周期である．環境の状態の変化は本来連続的であるが，ここでは周期的に変化するものとして考える．環境の周波数を Fc，センサのサンプリング周波数を Fs とすると，環境を正確に読み取るためには Fs＝Fc である必要がある．ただし，現実の環境は連続的なので，サンプリング周波数が環境の状態の変化の周波数を上回る場合については考えないものとする．つまり，環境を正しく認識できない場合は，Fc が Fs より小さい場合である．

例として，N=1，V1={0,1,2}，周波数が Fc=3.334，つまり 0.3[sec]の周期で状態遷移する環境を考える．この環境は図 3.8 に示すように状態遷移するものとする．ここで，周波数が同じ Fs=3.333･･･，0.3[sec]の周期で認識を行うセンサをもつエージェント A と，周波数が Fs=1.0，1.0[sec]の周期で認識を行うセンサをもつエージェント B を考える．エージェント

A は，図 3.9 で示すように，状態の変化の様子をすべて認識できる．そのため，各状態の遷移先について確定的であると認識できる．これに対しエージェント B は，図 3.10 で示すように，状態が変化していく途中の様子を見落としてしまう．その結果，E1＝2 の遷移先が E1=0 と E1＝2 の二通りとなり，確率的な遷移であると認識してしまう．

センサのサンプリング周波数が低い場合，認識した環境と実環境との間に，以上のような差異が生まれる．

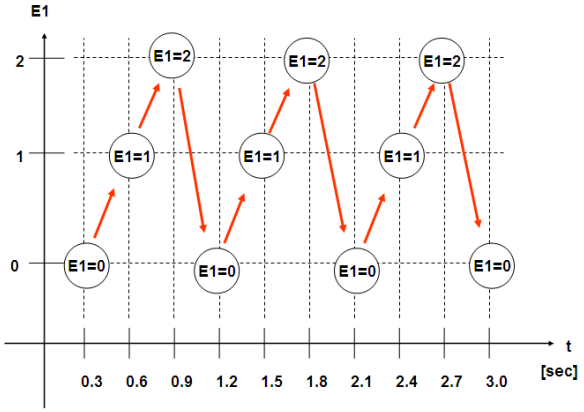

図 3.8　実環境の状態遷移

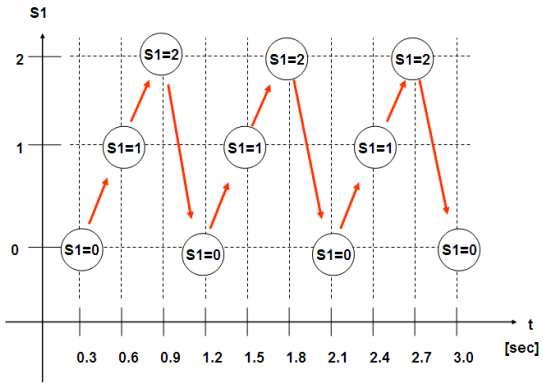

図 3.9　エージェント A が認識する環境

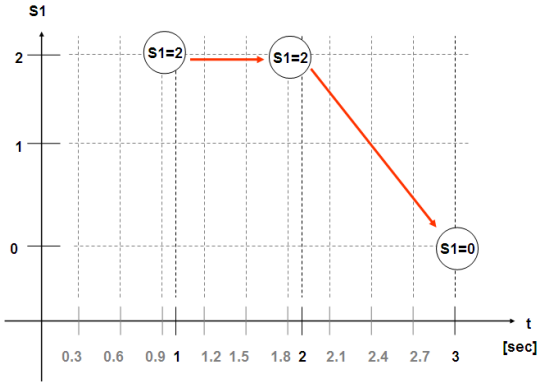

図 3.10　エージェント B が認識する環境

以上，センサの能力について紹介した．本論文では，この中でも特にセンサの種類数に注目し，センサの種類数の違いが学習に与える影響について検証を行う．

3.4　冗長なセンサが学習に与える影響

エージェントが環境を認識するとき，たくさんセンサを取り付ければいいというもので

はない．センサが環境に上手く対応できていない場合や，対応している環境の要素がタスクの遂行と無関係な場合など，エージェントによっては必要のないセンサが存在する可能性がある．このように，ロボットが学習を行うために本来必要のない無駄なセンサを冗長センサと定義する．また，これまでに述べた冗長でないセンサを対応センサと定義し，冗長センサと区別する．本節では冗長センサが環境認識や学習に及ぼす影響について考える．

冗長センサの種類は，以下の三つに大きく分けるとする．

- 重複センサ
- ノイズセンサ
- 不用センサ

これら三種類の冗長センサが学習効率に与える影響について考察する．

3.4.1 重複センサ

本論文では，環境の要素 1 つを認識するためには 1 つのセンサが必要であるとした．このとき，1 つのセンサで複数の環境の要素を認識することは出来ないが，複数のセンサが 1 つの環境の要素を認識するという状況が考えられる（図 3.11）．しかし，1 つの環境の要素を正しく認識するためには 1 つのセンサがあれば十分であり，同じ要素を認識するセンサを 2 つ以上設置する必要はない．そこで，同じ要素を認識するセンサが 2 つ以上設置されているとき，2 つ目以降のセンサを重複センサと呼ぶ．

重複センサが実際に存在する場合の例を図 3.12 に示す．本実験で扱うセンサは環境の要素を一つだけ認識可能としたが，実際には複数の環境の要素を認識可能なセンサが存在する．例えば超音波センサを持つエージェントは，物体との距離や物体の形状などが認識可能である．また，光学センサを持つエージェントは，物体との距離や物体の材質などを認識可能である．このように複数の環境の要素を認識可能なセンサがいくつか設置されている場合，互いの能力の一部が重複する可能性がある．

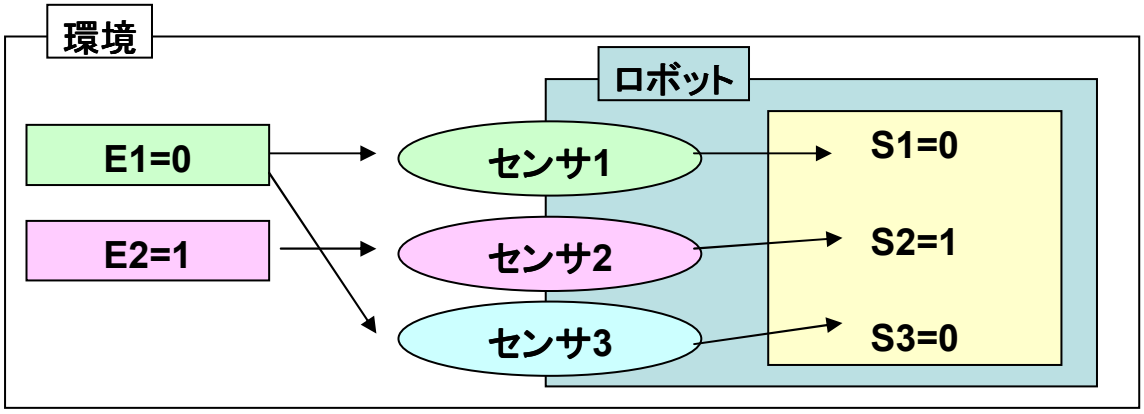

図 3.11　重複センサの例

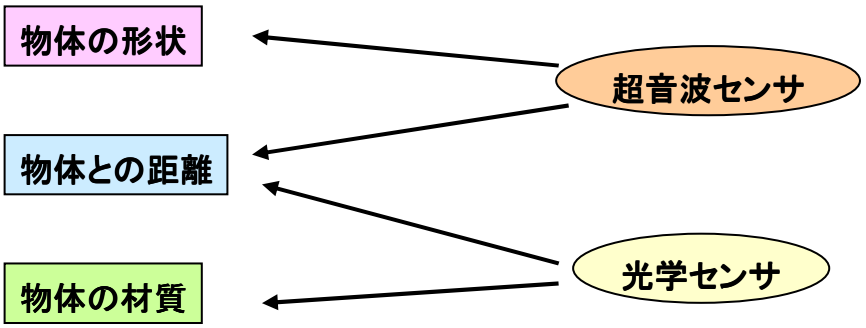

図 3.12　重複センサの実例

重複センサが環境認識に影響を与える例として，N=2，E_i={0,1}(i=1,2)の環境について考える．この環境を正しく認識するためには，要素 0 を読み取るセンサ 0 と，要素 1 を読み取るセンサ 1 の二つがあれば良い．その場合の状態空間は図 3.13 のような二次元空間となる．ここでさらに要素 0 を読み取るセンサ 2 を設置したとする．このとき，エージェントが認識する可能性のある状態空間は図 3.14 のような三次元空間となり，状態空間が増えることになる．さらにセンサが設置されれば，センサの数だけ状態空間の次元数が増えることになる．状態空間の次元数が増えるということは，それだけ学習が必要な状態数が増えることになる．そのため，学習に必要な試行回数が増える可能性がある．

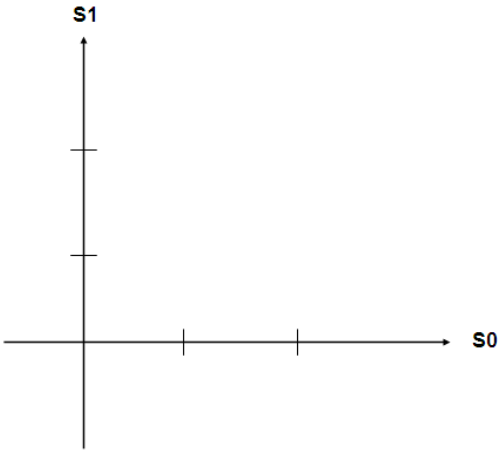

図 3.13　冗長センサのない状態空間

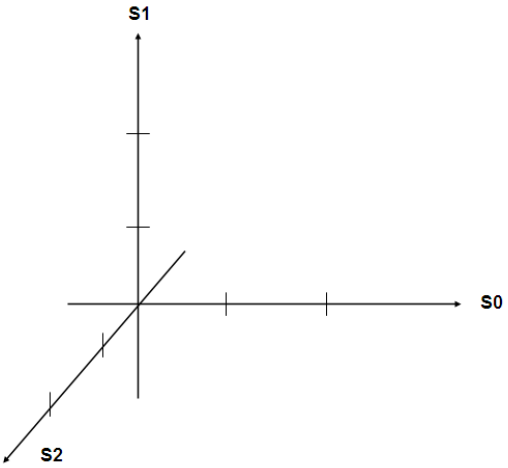

図 3.14　冗長センサを持つ状態空間

3.4.2　ノイズセンサ

ノイズセンサとは，環境の要素と関係のない数値を出力するセンサである．対応センサは状況に合わせて数値を取得するのに対し，ノイズセンサは状況に関係のない数値を出力してしまう．そのため，誤った状況を認識してしまう恐れがある（図 3.15）．

ノイズセンサが実際に存在する場合の例として，故障したセンサが挙げられる．故障によって，時々要素と関係のない数値を出力したり，常に同じ数値を出力するようになってしまうといった状況が考えられる．

このように，ノイズセンサが出力する数値のパターンはいくつかある．そこで本論文で扱うノイズセンサは，環境の要素とは関係なく，常にランダムの数値を出力するものとする．また，出力する数値は 0 と 1 の二通りとする．

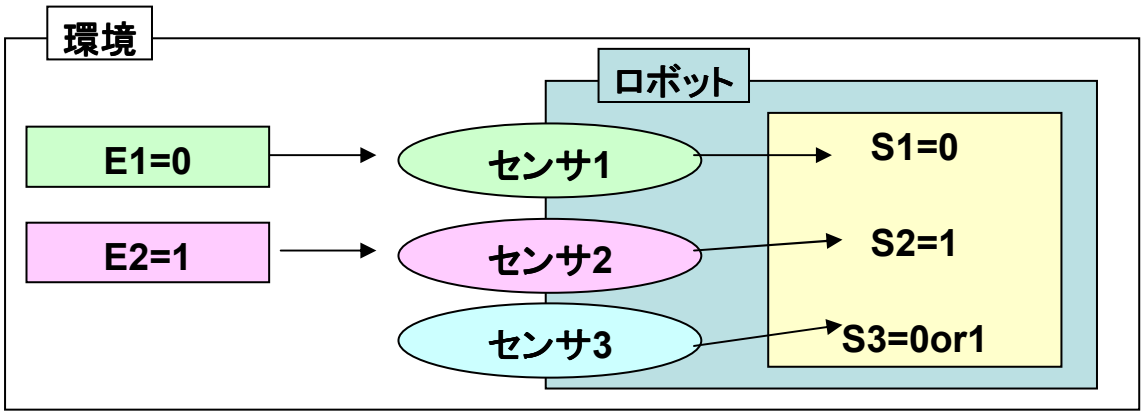

図 3.15　ノイズセンサの例

ノイズセンサが環境認識や学習に影響を与える例として，N=1，E_1={0,1}の環境上の，二体のエージェント A･B について考える．エージェント A は，要素 0 を読み取ることのできる正常なセンサ 0 のみをもつ．一方エージェント B は，センサ 0 に加え，ノイズセンサ 1 をもつ．このとき各エージェントがどのように環境を認識するか考える．エージェント A の場合環境を正しく認識するので，認識しうる状態は図 3.16 のようになる．しかしエージェント B の場合センサ 1 の情報が加わるため，認識しうる状態は図 3.17 のようになり，エージェント A よりも状態数が増えてしまう．その上センサ 1 の情報は環境と全く無関係なので，センサ 1 によって増えた状態は実際には同じ状態である（図 3.18）．そのため，状態数が必要以上に増え，学習に影響を及ぼす可能性がある．

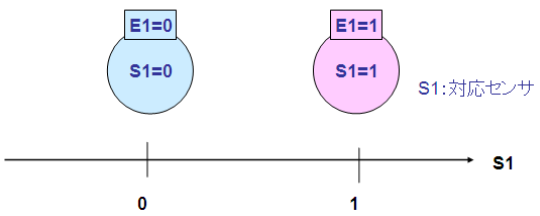

図 3.16　ノイズセンサのない状態空間

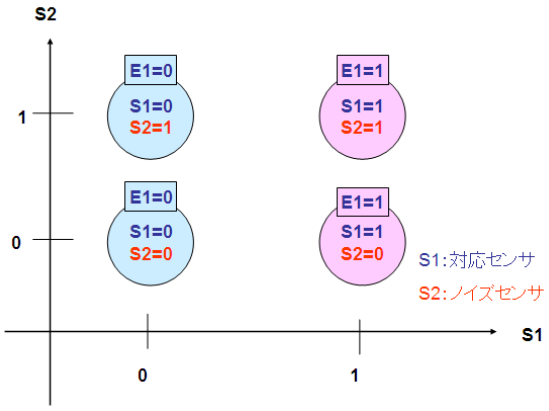

図 3.17　ノイズセンサを持つ状態空間

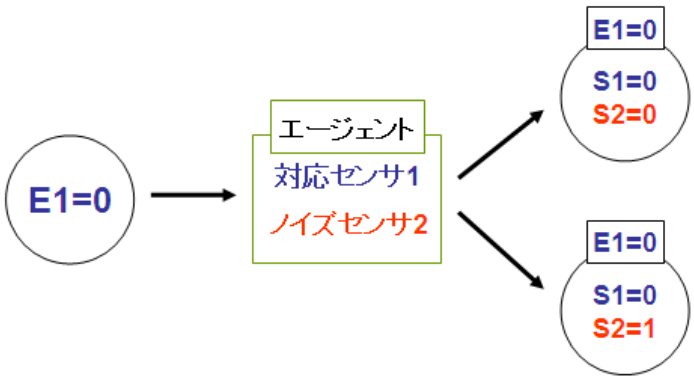

図 3.18　ノイズセンサの誤認識の様子

3.4.3　不用センサ

エージェントがタスクを達成する上で，環境のありとあらゆる要素が必要であるわけではない．例えば，障害物を避けて走るだけの簡単なロボットであれば，障害物を認識する

距離センサひとつあれば十分である．そのようなロボットが音センサや熱センサなどを持っていても，障害物を避けて走る，というタスクの達成には全く関係が無い．このように，たとえ環境の要素と対応していても，それがエージェントのタスク達成と無関係なセンサであれば設置する必要は無い．このような冗長センサを不用センサと呼ぶ．

強化学習を用いた例として，N=3，E_i={0,1}(i=1~3)の環境について考える．ただし，ここでは簡単のため遅延報酬は考えない物とする．強化学習は，報酬を最大にすることを目的とするため，報酬設定と関わりのない要素が存在していれば，その要素はタスクの達成とは無関係となる．そこで，要素 E1 と E2 のみが報酬に関係し，E3 は報酬と関わりのない環境の状態空間を図 3.19 に示す．

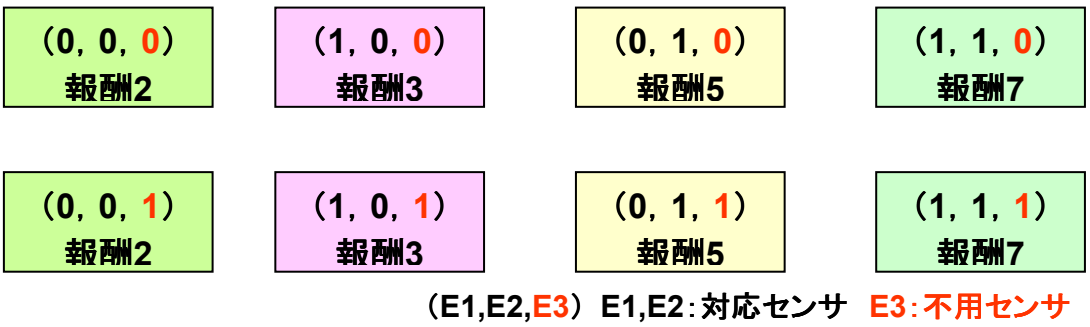

図 3.19　タスク達成と無関係な要素が存在する状態空間

このとき，E3 の値に関わりなく，E1 と E2 の値のみによって報酬が決まる．このため，E3 を認識しなくても，報酬を最大にするというタスクの達成には影響が無いと言える．逆に，E3 を認識する場合，その分状態空間が増えてしまい，学習に悪影響を及ぼす可能性がある．

以上，三つの冗長センサについて述べた．本論文では，この中でも特に重複センサとノイズセンサに注目し，これらが学習に与える影響について調査する．

3.5　検証方法

本節では，エージェントの環境認識能力の差が学習に与える影響を調査するための検証方法を述べる．

まず，本論文における実験では，第 2 章で定義したセンサと環境を用いる．第 2 章で定義したセンサと環境は，一般的なセンサや環境を抽象化するために定義したものである．

しかし，実験の際に実機を用いると，センサの能力はロボット固有の物になってしまう．そのため本論文ではシミュレーションを用いて実験を行う．

本論文では，センサの種類数の違いが学習に与える影響を検証することを目的としている．そのため，環境認識能力の異なるエージェントを複数台用意する必要がある．このとき各エージェントの違いはセンサの種類数のみとし，その他のセンサ能力や身体構造についてはすべて同じものを用いる．具体的には，各エージェントの用いる学習手法・センサの分解能・サンプリング周波数・選択できる行動の数はすべて同一のものとする．これにより，センサの種類数以外の要素が学習に影響を及ぼすのを防ぐ．

以上のようにセンサの種類数の異なるエージェントを複数台用意し，一つの環境内で同時に学習を行わせる（図 3.20）．その学習結果を比較することにより，各センサ数のエージェントの学習効率の違いを検証する．今回は学習手法に強化学習を用いるため，学習結果の比較は各エージェントが獲得した報酬を比較することで行う．

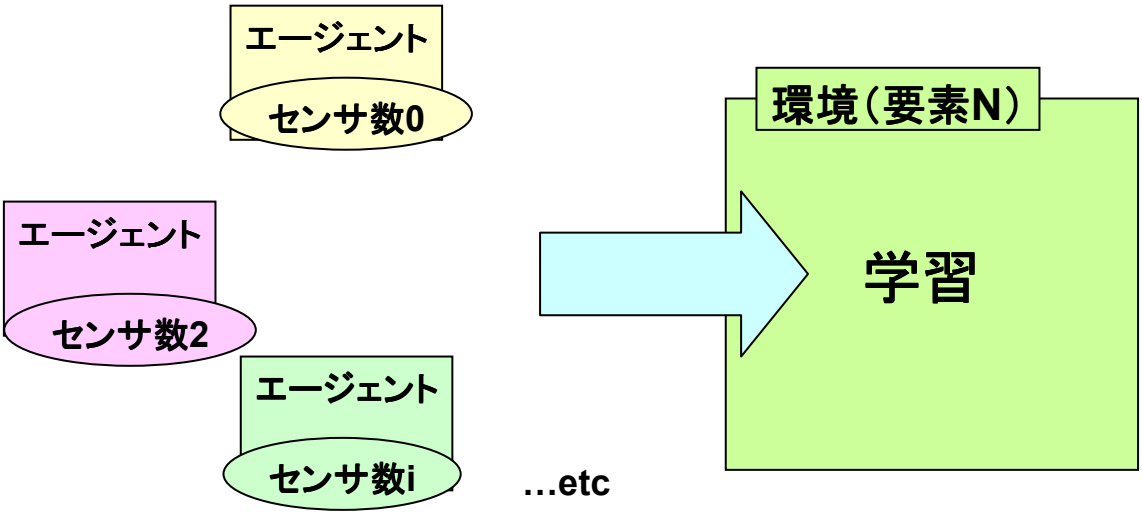

図 3.20　検証方法の概念図

ここで，センサの種類数の設定の仕方について考える．対応センサは環境の要素と 1 対 1 で対応している．そのため，環境の要素数を N とすると，対応センサの最大数も N となる．そのため本論文では，対応センサ数が 0～N のエージェントを用いて実験を行う．

一方，本論文で用いる冗長センサである重複センサとノイズセンサは，環境の要素数とは関係なく存在する．そのため，重複センサとノイズセンサの数には上限が無い．そこで本論文では，比較を行う上で十分な数の冗長センサを実験に合わせて決めることとする．ここで，N=3，冗長センサの最大数が 4 の場合の各センサの組み合わせ例を図 3.21 に示す．

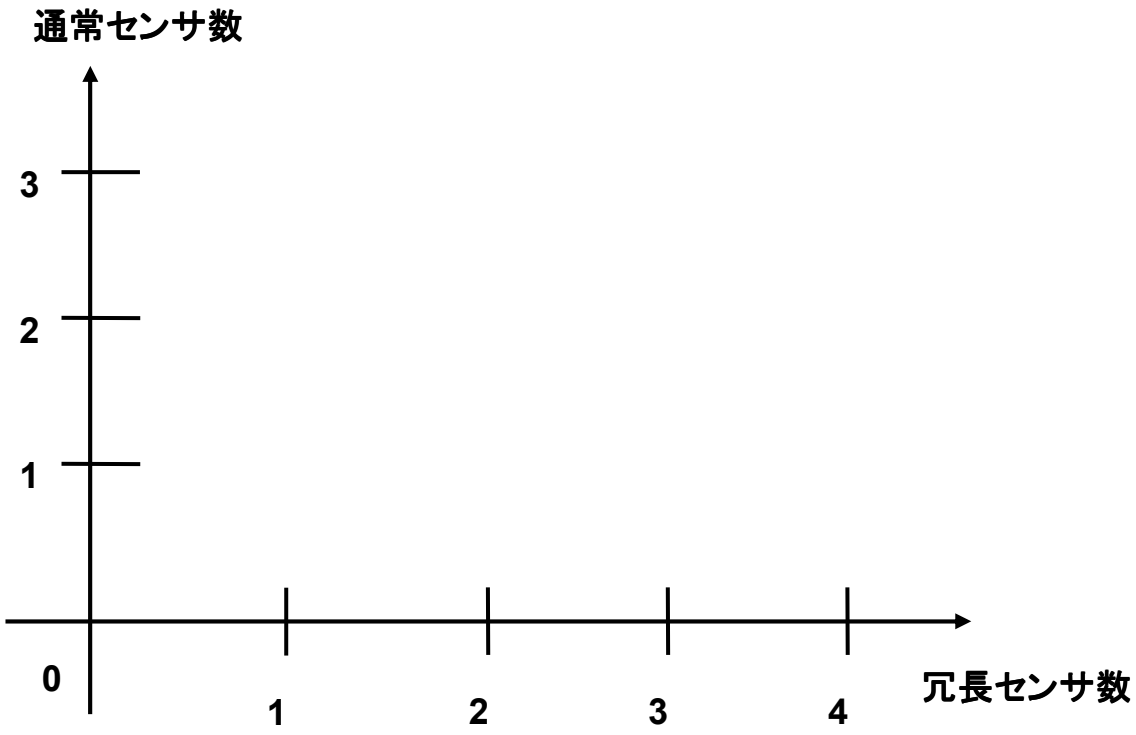

図 3.21　冗長センサの組み合わせ例

次に，検証に用いる環境について述べる．環境の取り得る状態の数は，環境の要素の数と各要素が取り得る値の個数が決まれば，自動的に決定される．環境の要素 E_i の取り得る値の個数を $S(E_i)$とすると，環境の取り得る状態の数 S(E)は次の式（3.3）で表される．

$$S(E) = \prod_{i=1}^{N} S(E_i) \qquad (3.3)$$

最後に, 結果の比較方法について述べる. 各エージェントが行った学習の効果の比較は, 二種類の方法を用いて行う．ひとつは，各エージェントが R ステップの学習で得た報酬の総和を比較するものである．しかし，最終的に得た報酬量が少なくても，学習初期段階では獲得量が多いなど，学習途中の獲得報酬の推移を確認する必要がある．そこで，各エージェントが各ステップに得た報酬の推移についても比較を行う．

まず，各エージェントが獲得した報酬の総和の比較方法について述べる．学習が行われない（ランダムに行動する）エージェントが得た報酬の総和を 0，対応センサ数が最大かつ冗長センサ数が 0 のエージェントが得た報酬の総和を 1 として正規化を行う．ここで対応

センサ数が最大かつ冗長センサ数が 0 のエージェントが得た報酬の総和を 1 としたのは，基本的には対応センサが多く冗長センサが少ないほどエージェントの学習効率が高いと予想されるからである．正規化の式を式（3.4）に示す.

$$RATIO_{i,j} = \frac{\sum_{k=1}^{R} r_{i,j,k} - \sum_{k=1}^{R} r_{random,k}}{\sum_{k=1}^{R} r_{N,0,k} - \sum_{k=1}^{R} r_{random,k}} \tag{3.4}$$

ここで $r_{i,j,k}$ は，i 種類のセンサと j 種類の冗長センサを持つエージェントが k 試行回数目に得た報酬の値である.

次に，各エージェントが各ステップに得た報酬の推移の比較方法について述べる．本実験で用いる環境は，その時点での報酬が低くても，そこから二回，三回と遷移を行うとさらに高い報酬が得られる可能性がある．そのため，一度の遷移で得た報酬ではなく，定常的に得られる報酬の量で比較する必要がある. そこで, L ステップで得た報酬の平均（SMA）を算出し，比較に用いる.

$$SMA = \frac{P_t + P_{t-1} + \cdots + P_{t-L-1}}{L} \tag{3.5}$$

以上のような設定において実験を行い，センサの種類数の違いが学習に与える影響を検証する.

第 4 章　実験

4.1　実験概要

本節では，本論文で行う実験の概要について説明する．本実験は 3 章で述べた検証方法を用いる．また，実験は，実機ではなくシミュレーションによって行う．

実験は次のように分けられる．

・ 冗長なセンサを持たないエージェントによる実験

実験 1：定常環境での実験

実験 2：非定常環境での実験

・ 冗長なセンサを持つエージェントによる実験

実験 3：定常環境における重複センサの実験

実験 4：定常環境におけるノイズセンサの実験

実験 5：非定常環境におけるノイズセンサの実験

まず, 実験 1～2 で冗長なセンサを持たないエージェントについて, 学習効率を比較する．実験 1 では変動の無い定常環境に対して，実験 2 では確率的に変動する非定常環境に対して実験を行う．

次に，実験 1～2 の結果をふまえた上で，冗長なセンサを持つエージェントを用いて実験を行う．実験 3 では，重複センサを用いた実験を，実験 4～5 では，ノイズセンサを用いた実験を行う．

以上の実験により，センサの種類数の変化が学習に及ぼす影響を明らかにする．

4.2　冗長センサを持たないエージェントを用いた実験 1

4.2.1　実験の目的

本節では，エージェントが冗長なセンサを持たない場合について，センサの種類数の違いが学習に及ぼす影響の検証を行う．また，本節では定常環境を用いて検証することを目的とする．

4.2.2　実験方法

実験方法は，3 章で述べた検証方法を用いる．N 種類の要素からなる環境上で，センサの種類数が 0～N 種類のエージェントに同時に学習を行わせる. また, 学習効果の比較のため,

学習を行わずランダムに行動するエージェントも行動させることとする．

実験の流れの概要図を図 4.1 に示す．1 個の環境に対し，各エージェントに R ステップの学習を行わせる．これを S 種類の環境に対して行い，その結果獲得した報酬の平均を用いて学習効率の比較を行う．今回用いる環境はランダムに生成される．そのため，環境の作り方によっては特定のエージェントに対して有利な状況が生まれる可能性がある，そこで S 種類の環境での平均を用いることで，環境によって特定のエージェントに対して有利な状況が生まれる，ということを防ぐ．

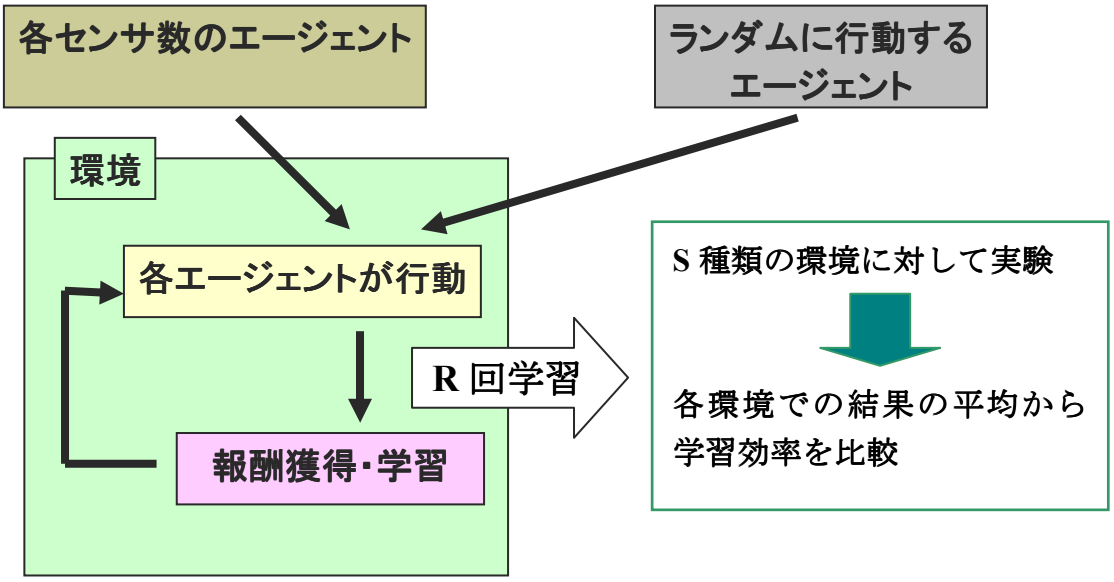

図 4.1　実験の流れの概要図

4.2.3　実験に用いるタスク

エージェントの行うタスクは，最大の報酬が得られるルートの探索である．エージェントは環境の各状態において行動を選択し，次の状態へと遷移することができる．その結果，遷移先の状態に応じてエージェントは報酬を得る．このような環境内で R 回行動を行い，その中で得られる報酬の総和を最大にすることがエージェントの目的となる．

今回用いるタスクでは，遅延報酬が生じる．その例を図 4.2 に示す．図中の状態 6 に遷移した場合，得られる報酬は 1 しかない．しかし，そこからさらに状態 5，状態 1 の順に遷移することで，10 という高い報酬を得られる．このように，その状態で得られる報酬が少なくても，将来的に高い報酬が得られる場合が存在する．そのため，先を見据えた学習が必要となる．このようなタスクでエージェントは学習を行う．

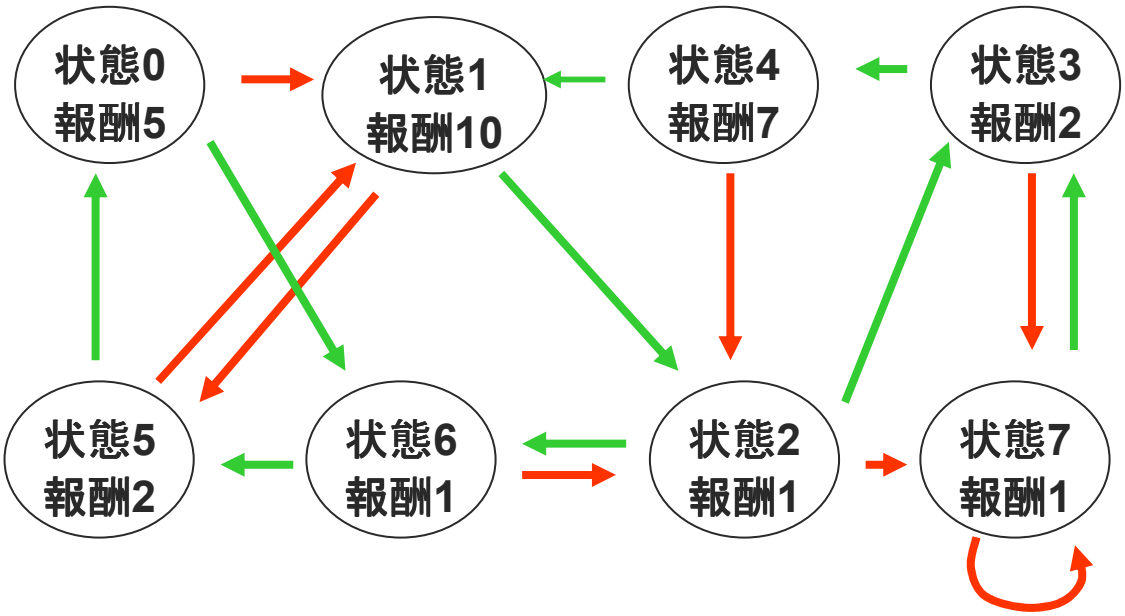

図 4.2　状態数 8 の場合のタスクの例

ここで，環境の各状態における報酬の設定は，次のように行う．現実世界において，高い報酬が得られるような状態を考えたとき，その状態に留まり続けることは一般的に難しい場合が多い．学校での成績を例にすると，高い成績を維持するには日々の予習復習が欠かせない．高い成績であるほど必要な勉強量は増え，ミスも許されなくなる．そのような現実世界をモデルとして報酬設定を行った．状態 i への移動に対して与えられる報酬 Rwd_i を式（4.1）に従って定義する．

$$Rwd_i = \exp\lfloor \alpha_{\exp} \cdot D_{ii} \rfloor \tag{4.1}$$

式(4.1)において，D_{ii} は i 番目の状態からスタートして，再び i 番目の状態へ戻ってくるまでの最短距離である． $\alpha_{\exp}$ は exp の値を調整するための係数である．

4.2.4　実験設定

本実験におけるその他の設定を以下に示す．

・環境に関する設定

本実験において，エージェントが学習を行う環境は，N 種類の要素から構成される環境とし，各要素の値 E_i(i=1～N)は，E_i={0,1}の 2 値をとるものとした．また，環境の状態は，各要素の組み合わせによって決定される．そのため，環境のとりうる状態の数 S(E)は式(4.2)で表される．

$$S(E) = 2^N \tag{4.2}$$

今回，各状態における遷移先は以下の条件に従ってランダムに設定した．

- 環境の各状態は，Tr 種類の遷移先を持つ．
- 環境のどの状態も，他の全ての状態へ遷移可能なルートを持つ

以上のような環境下で，エージェントは学習を行う．

・エージェントに関する設定

本実験では，エージェントの設定を次のようにして，学習を行わせる．エージェントの選択可能な行動は，遷移先の数と同数とする．0～N 種類のセンサを持ったエージェントを各一体ずつ，計 N+1 体のエージェントに学習を行わせる．また，今回の環境では，学習が行われない場合でもある程度の報酬が得られることが予想される．そのため，学習を行わずランダムに行動して報酬を得るエージェントを同時に行動させる．以上より，本実験では N+2 体のエージェントが同一環境内で行動する．

・学習手法に関する設定

本実験では，第 3 章で述べたように強化学習を用いて実験を行う．強化学習の手法には，代表的な手法である Q 学習を用いる．行動選択手法は，代表的な手法として ε -greedy 法・Softmax 法・追跡手法などが挙げられるが，その中でも今回は Softmax 法を用いる．

・パラメータ設定

各パラメータ設定を表 4.1～4.3 に示す．

表 4.1　環境の設定に関するパラメータ

N（環境を構成する要素の数）	10
Tr（環境の各状態における遷移先の数）	2
S(E)　（環境の取り得る状態の数）	1024
α_{exp}	0.65

表 4.2　学習手法に関するパラメータ

α	0.7
γ	0.6
τ	1000

表 4.3　学習回数に関するパラメータ

R（学習回数）	50000
S（学習を行う環境の種類）	20

4.2.5　実験結果

以上の設定で実験を行った．その結果を図 4.3～4.7 に示す．

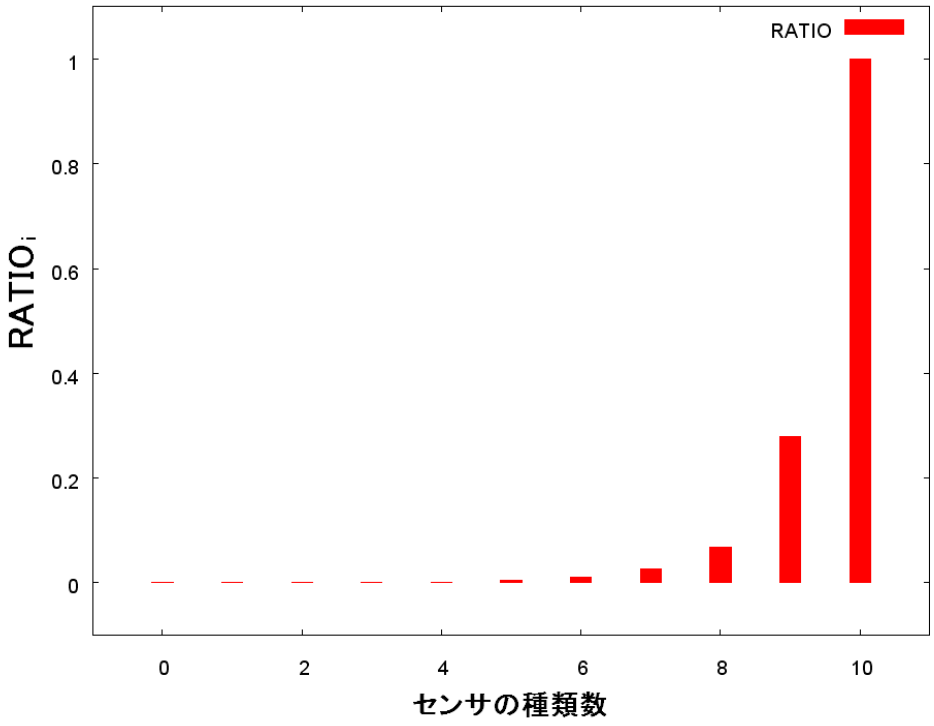

図 4.3　獲得報酬の総和の比較

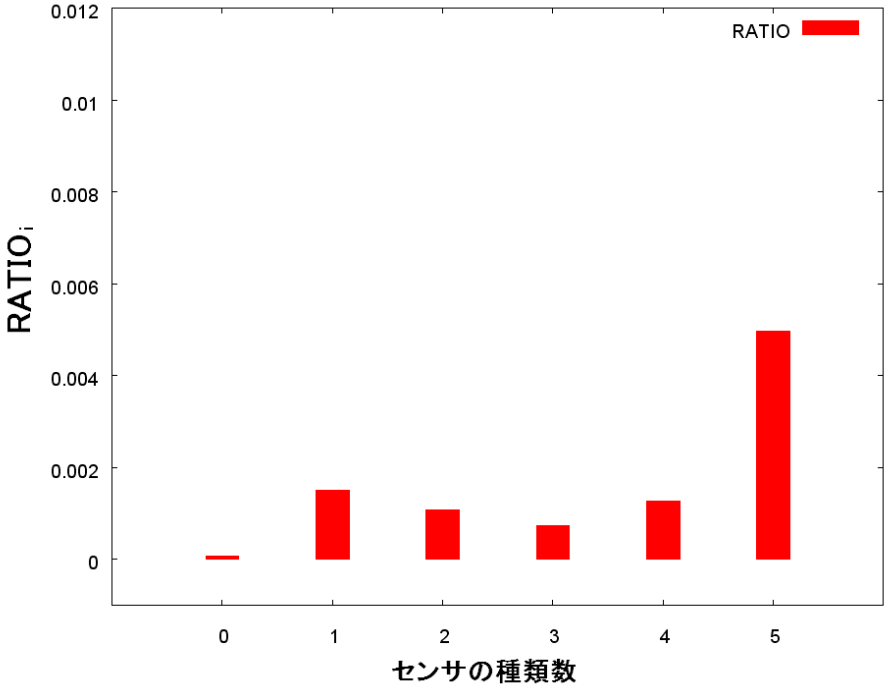

図 4.4　図 4.3 をセンサ数 5 以下のエージェントについて拡大したもの

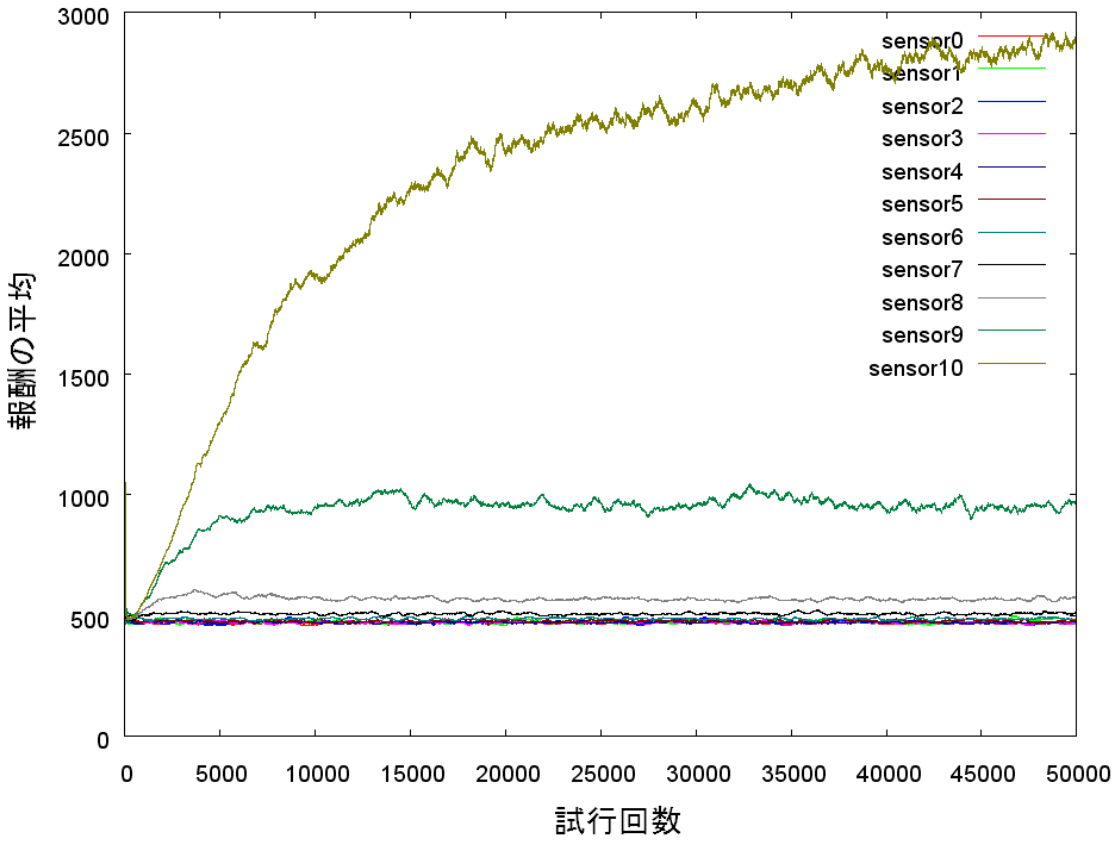

図 4.5　L 試行回数で平均をとった報酬の推移

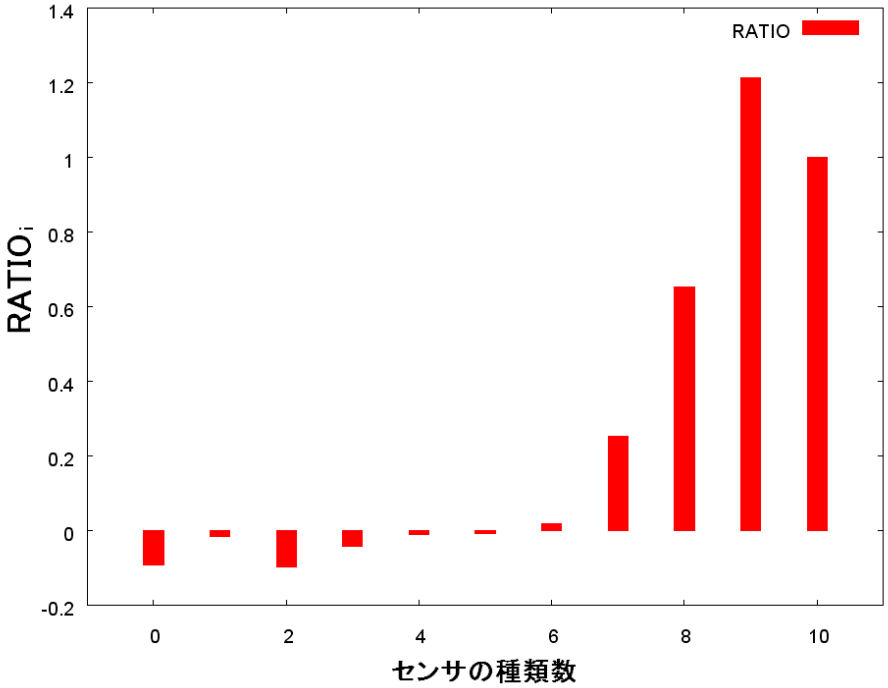

図 4.6　試行回数 0～2000 回での獲得報酬の比較

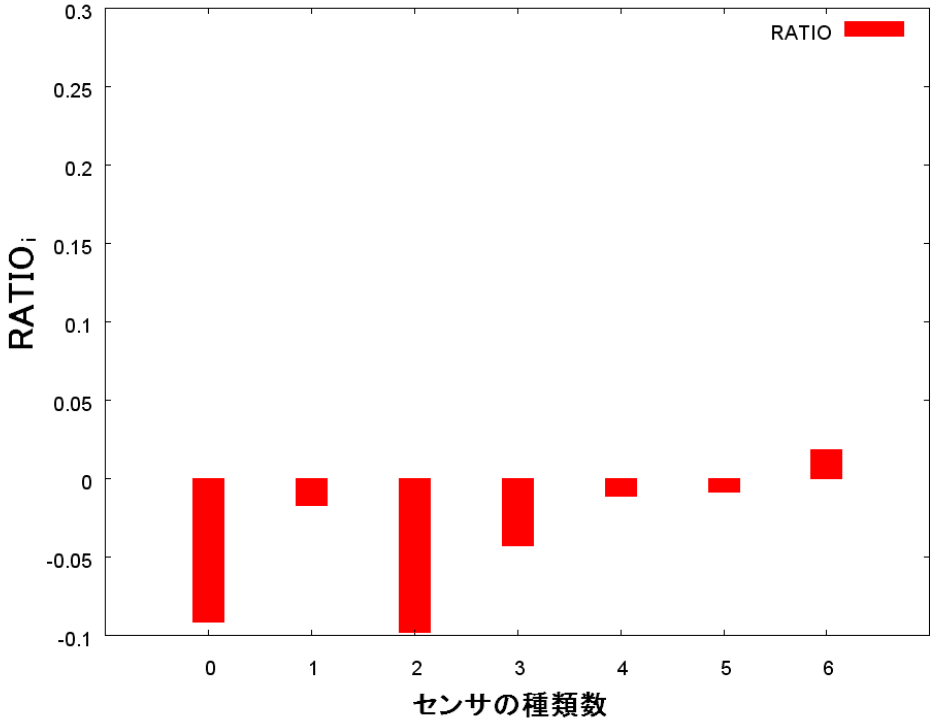

図 4.7　図 4.6 をセンサ数 6 以下のエージェントについて拡大したもの

まず図 4.3 を見ると, センサ数が多いエージェントほど最終的な学習効率が高いという結果になった．学習効率の増え方はセンサが増えるごとに指数関数的に増えている．このことは，センサが一つ増える度に 2 倍の状態を認識可能なため，学習効率も 2 倍ずつ増えているためだと考えられる．

次に，図 4.3 のうちセンサ数の少ないエージェントに注目する．図 4.3 をセンサ数の少ないエージェントについて拡大したのが図 4.4 である．この図を見ると，センサ数が 4 種類以下のエージェントについてはセンサ数に関係なくバラバラの学習効果となっている．このため，センサ数が一定数以下になると学習の効果が期待できなくなるといえる．

次に図 4.5 を見ると, センサ数が最大のエージェントの獲得報酬は 15000 回程度の試行回数でようやく安定し始め，その後も緩やかに獲得報酬が増大している．一方，センサ数が 9 種類のエージェントを見ると, 試行回数が 5000 回の付近で早くも獲得報酬が安定している．センサ数が 8 種類に減ると，獲得報酬が安定するまでの試行回数はさらに少なく，試行回数 2500 回程度で安定している．これは，センサ数が多い場合には認識可能な状態数が多くなるため，学習のために経験しなければならない状態数も増え，学習に時間がかかるのだと考えられる．このことから，試行回数が少ない場合，センサ数が最大よりも少ないエージェントでも高い学習効果が期待できる可能性がある．

学習回数が少ない場合の獲得報酬の総和を比較したのが図 4.6 である．この図を見ると，センサ数 9 のエージェントがセンサ数 10 のエージェントより高い学習効率を示している．さらにセンサ数 8 以下エージェントも，学習回数が多い場合の結果に比べてセンサ数 10 に近い学習効率を示している．このことから，学習回数が少ない場合にはセンサ数が少ない方が高い学習効率を示したり，センサ数が多い場合と大きな差の無い学習効率を示す場合がある．そのため，学習回数をあまり多く取れないような環境では，センサ数はある程度少ない方が良いといえる．

最後に，図 4.6 のうちセンサ数の少ないエージェントに注目する．図 4.6 をセンサ数の少ないエージェントについて拡大したのが図 4.7 である．センサ数が 6 個を下回ると，学習回数が多い場合と同じようにほとんど学習効果が出ていない．そのためセンサ数が少なすぎる場合は，どんな環境に対しても上手く学習を行えないことがわかる.

4.2.6 考察

今回の実験では，学習時間が十分にとれる場合にはセンサ数が多いエージェントほど高い学習効率を示す事がわかった．このとき学習効率は，センサ数に応じて指数関数的に高

くなる．特にセンサ数が最大のエージェントは，ある程度獲得報酬が安定してからも学習回数を重ねる毎にさらに獲得報酬が増えていく．そのため，試行回数が十分に取れるような環境ではセンサ数をなるべく多く設置することは非常に重要であることがわかった．

一方，センサの数が多いエージェントは学習完了までに時間がかかる事もわかった．そのため学習回数が少ない場合には，センサ数が少し少ないエージェントの方が高い，もしくはあまり差のない学習効率を示すこともわかった．このことから，十分な学習回数をとれないような場合には，センサ数を少し減らした方が良いと言える．

最後に，センサ数が少なすぎる場合には，学習回数にかかわらず学習が上手く行われないことがわかった．そのため，どんな環境で学習を行うにせよ，ある程度のセンサ数を設置することは必須であるといえる．

4.3 冗長センサを持たないエージェントを用いた実験 2

4.3.1 実験の目的

本節では，エージェントが冗長なセンサを持たない場合について，センサの種類数の違いが学習に及ぼす影響の検証を行う．また，本節では非定常環境を用いて検証することを目的とする．

4.3.2 実験方法

実験方法は，3 章で述べた検証方法を用いる．N 種類の要素からなる環境上で，センサの種類数が 0〜N 種類のエージェントに同時に学習を行わせる．また，学習効果の比較のため，学習を行わずランダムに行動するエージェントも行動させることとする．

今回の実験では，環境変動が起こる非定常環境を用いて行う．環境変動は，エージェントが Ch 回試行する度に起こる．

実験は，次のような流れで行った．1 個の非定常環境に対し，各エージェントに R ステップの学習を行わせる．これを S 種類の非定常環境に対して行い，その結果獲得した報酬の平均を用いて学習効率の比較を行う．今回用いる環境はランダムに生成される．そのため，環境の作り方によっては特定のエージェントに対して有利な状況が生まれる可能性がある，そこで S 種類の環境での平均を用いることで，環境によって特定のエージェントに対して有利な状況が生まれる，ということを防ぐ．

4.3.3　実験に用いるタスク

エージェントの行うタスクは，最大の報酬が得られるルートの探索である．エージェントは環境の各状態において行動を選択し, 次の状態へと遷移することができる. その結果, 遷移先の状態に応じてエージェントは報酬を得る．このような環境内で R 回行動を行い, その中で得られる報酬の総和を最大にすることがエージェントの目的となる.

ここで，環境の各状態における報酬の設定は，4.2 での実験と同様，4.2.3 の式（4.1）を用いる.

4.3.4　実験設定

本実験におけるその他の設定を以下に示す.

・環境に関する設定

本実験において，エージェントが学習を行う環境は，N 種類の要素から構成される環境とし，各要素の値 E_i(i=1～N)は，E_i={0,1}の 2 値をとるものとした．また，環境の状態は，各要素の組み合わせによって決定される.

今回，各状態における遷移先は以下の条件に従ってランダムに設定した.

- 環境の各状態は，Tr 種類の遷移先を持つ.
- 環境のどの状態も，他の全ての状態へ遷移可能なルートを持つ

以上のように環境は生成される.

また，今回の実験で用いる環境は，環境変動の起こる非定常環境である．環境変動の設定は次のように行った．エージェントが Ch 回試行を行うと，環境は変動を起こす．環境変動は，ある状態から次の状態への遷移先を別の状態へ遷移するように変更することで行う．遷移先の変更は，各遷移先について P%の確率で行われる（図 4.8）．このとき遷移先の変更は，初期の遷移先の設定と同じ条件でランダムに行われる.

本実験では，各状態の距離に応じて報酬の設定が行われている．そのため，遷移先が変更されると状態間の距離にも変化が生じる．そこで，環境変動が行われたときには，その距離に応じて各状態の報酬を再設定する. このときの報酬設定も, 4.2 での実験と同様, 4.2.3 の式（4.1）を用いる.

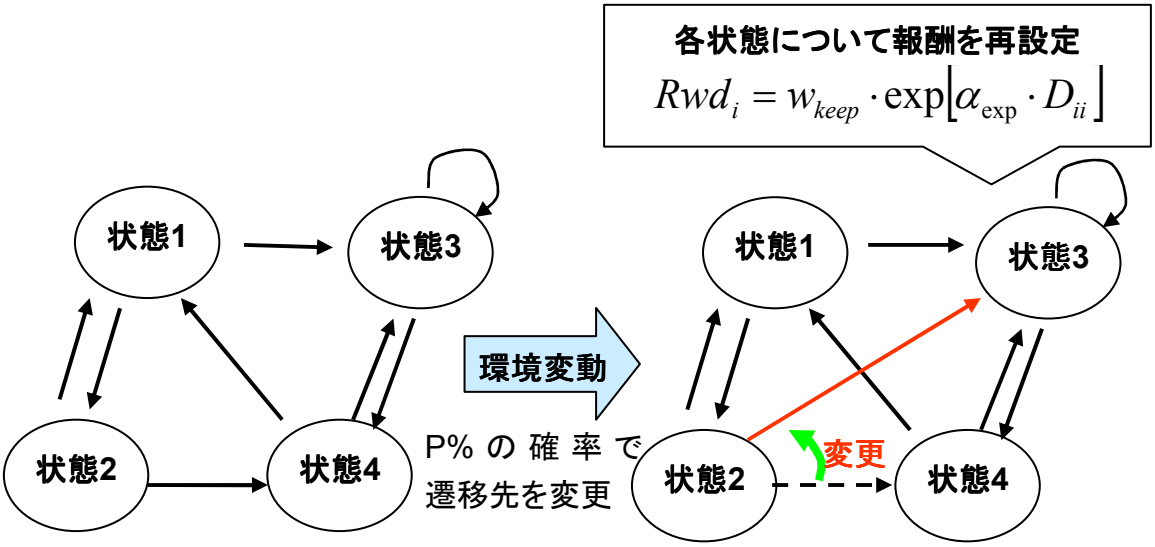

図 4.8　環境変動の例

以上のような環境下で，エージェントは学習を行う．

・エージェントに関する設定

本実験では，エージェントの設定を次のようにして，学習を行わせる．エージェントの選択可能な行動は，遷移先の数と同数とする．0～N 種類のセンサを持ったエージェントを各一体ずつ，計 N+1 体のエージェントに学習を行わせる．また，今回の環境では，学習が行われない場合でもある程度の報酬が得られることが予想される．そのため，学習を行わずランダムに行動して報酬を得るエージェントを同時に行動させる．以上より，本実験では N+2 体のエージェントが同一環境内で行動する．

・学習手法に関する設定

本実験では，第 3 章で述べたように強化学習を用いて実験を行う．強化学習の手法には，代表的な手法である Q 学習を用いる．行動選択手法は，代表的な手法として ε -greedy 法・Softmax 法・追跡手法などが挙げられるが，その中でも今回は Softmax 法を用いる．

・パラメータ設定

各パラメータ設定を表 4.4～4.7 に示す．

表 4.4　環境の設定に関するパラメータ

N（環境を構成する要素の数）	10
Tr（環境の各状態における遷移先の数）	2
S(E)　（環境の取り得る状態の数）	1024
α_{exp}	0.65

表 4.5　環境変動に関するパラメータ

Ch	2500
P	10

表 4.6　学習手法に関するパラメータ

α	0.7
γ	0.6
τ	1000

表 4.7　学習回数に関するパラメータ

R（学習回数）	50000
S（学習を行う環境の種類）	20

4.3.5　実験結果

以上の設定で実験を行った．その結果を図 4.9～4.11 に示す．

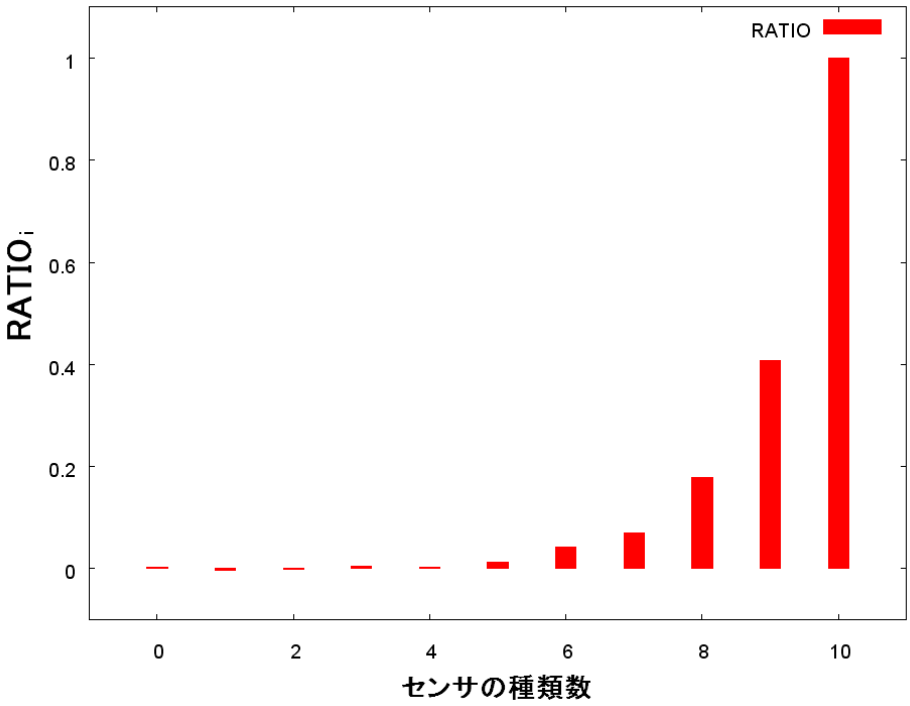

図 4.9　獲得報酬の総和の比較

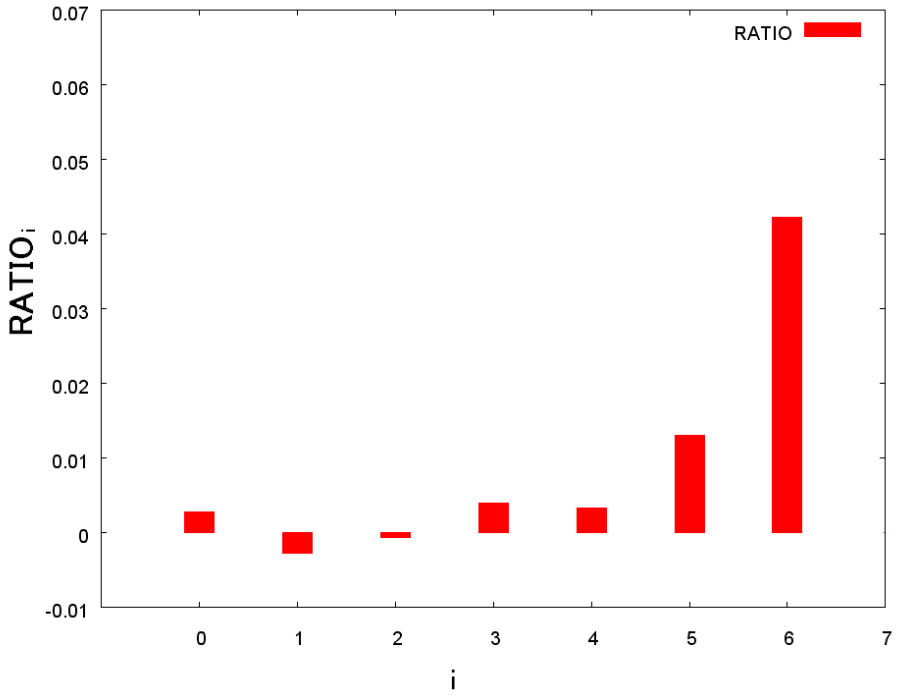

図 4.10　図 4.9 をセンサ数 6 以下のエージェントについて拡大したもの

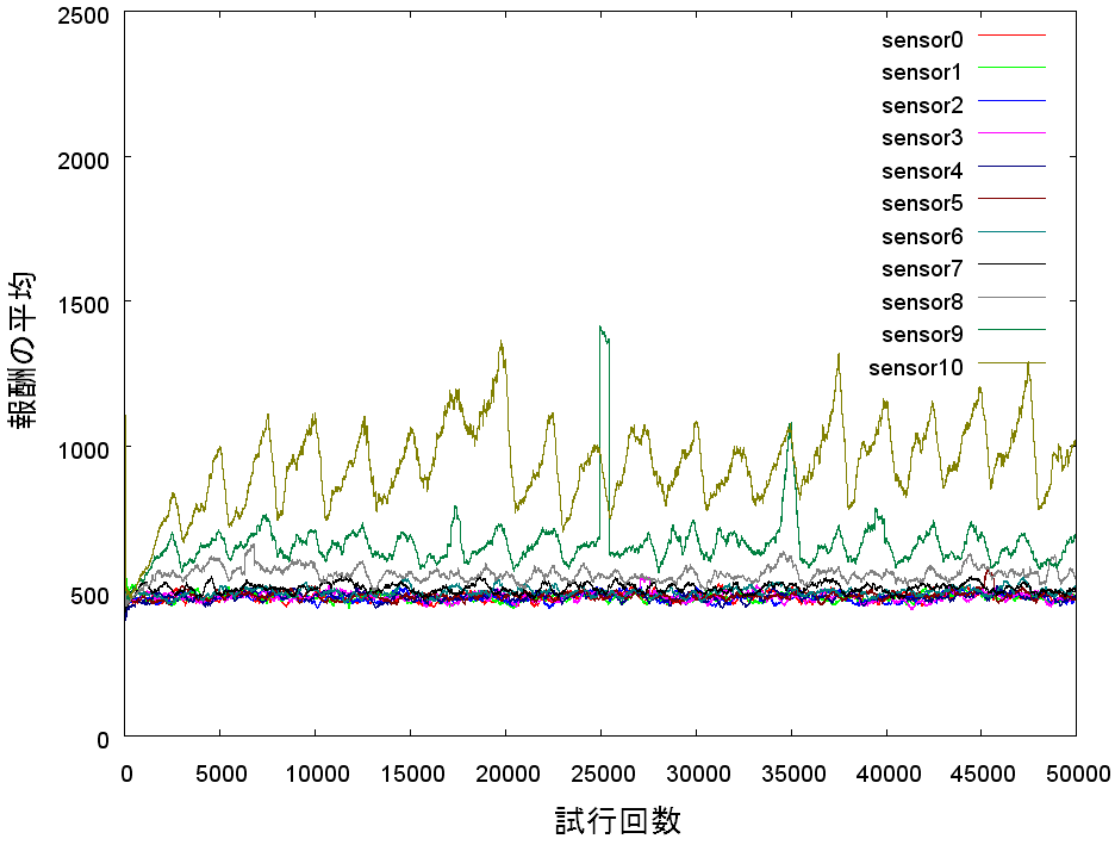

図 4.11　L 試行回数で平均をとった報酬の推移

まず図 4.9 を見ると，定常環境での実験同様，センサ数が多いエージェントほど最終的な学習効率が高いという結果になった．しかし学習効率の増え方は，4.2 での実験に比べて緩やかになっている．この結果の原因は図 4.4.11 に現れている

図 4.11 は獲得報酬の推移のグラフである．センサ数の多いエージェントに注目すると，環境変動の起こる度に獲得報酬が減っていることがわかる．これは，環境が変化した部分について再学習が必要であることが原因と考えられる．また，4.2 の実験結果から，センサ数が少ないエージェントほど学習速度が早いことがわかっている．そのため，何度も再学習する必要がある非定常環境では，センサ数の多いエージェントが不利となったと考えられる．

しかし，学習効率がある程度落ちたとはいえ，センサ数が増える毎に学習効率が格段に良くなることに変わりはない．これは，環境変動の仕方に原因があると考えられる．今回の環境変動はあくまでも環境の一部が変化するものである．そのため，環境変動が起きても，一部の状態については変動前の学習内容を引き継いで利用できる．このため，再学習

の際には一から学習し直すわけではなく，ある程度学習が進んだ地点からの再学習となる．そのため，センサ数が多いエージェントが依然高い学習効率を示していると考えられる．

最後に，図 4.10 のうちセンサ数の少ないエージェントに注目する．図 4.10 をセンサ数の少ないエージェントについて拡大したのが図 4.11 である．センサ数の少ないエージェントは，定常環境と同じようにセンサ数と関係なくバラバラの学習効果を示している．このことから，センサ数が少なすぎる場合は非定常環境であっても学習効果が期待できないといえる．

4.3.6 考察

今回の実験では，環境変動がある場合でも，センサ数の多いエージェントが高い学習効率を示すことがわかった．しかし，定常環境に比べると，センサ数による学習効率の差は小さいものであった．この時の学習効率の違いは次の二つの要素によって決まると考えられる

- 環境変動までの試行回数
- 環境変動時に変化する状態の範囲

環境変動までの試行回数が少ない場合，環境変動が頻繁に起き，そのたびに再学習が必要となる．そのため，センサ数が多く学習速度の遅いエージェントは学習効率が落ちることになる．

環境変動時に変化する状態の範囲が広い場合，再学習のときにより多くの状態について学習を行う必要がある．そのため，センサ数が多く学習速度の遅いエージェントは学習効率が落ちることになる．

以上より，非定常環境下では学習速度が非常に重要であることがわかる．今回の実験では，環境変動の範囲がある程度狭かったため，センサ数が多いエージェントの学習効率が高かった．しかし，環境変動の範囲がもっと広い場合や環境変動の頻度がもっと多い場合には，センサ数が少ないエージェントの学習効率が高くなる可能性がある．そのため非定常環境下では，環境変化の度合いによってセンサ数を調整する必要があるといえる．

また，定常環境での実験と同じく，センサ数が少なすぎる場合には学習効果が期待できないこともわかった．このことから，センサ数が少なすぎる場合にはどんな環境であれ学習効果が期待できないと考えられる．そのため，最低限のセンサを設置することは必須条件であるといえる．

4.4 重複センサを持つエージェントでの実験

4.4.1 実験の目的

本実験では，エージェントが冗長なセンサを持つ場合，センサの種類数の違いが学習に及ぼす影響について，定常環境を用いて検証することを目的とする．本節では，冗長センサの中でも重複センサに注目して実験を行う．

4.4.2 実験方法

実験方法は，4 章で述べた検証方法を元にする．N 種類の要素からなる環境上で，各センサ数のエージェントに同時に学習を行わせる．エージェントは，センサ数が 0〜N 種類の各エージェントに対し，冗長センサ数が 0〜M 種類のエージェントを用意する．ただし，重複センサの性質上，センサ数が 0 のときには重複センサは生まれない．そのため，センサ数 0 で冗長センサ数 1 以上のエージェントは考えないものとする．以上より，(N*M)+1 体のエージェントに学習を行わせる．また，学習効果の比較のため，学習を行わず，ランダムに行動するエージェントも，同時に行動させることとする．

実験は次のような流れで行った．1 個の環境に対し，各エージェントに R ステップの学習を行わせる．これを S 種類の環境に対して行い，その結果獲得した報酬の平均を用いて学習効率の比較を行う．ここで S 種類の環境での平均を用いる理由は，環境によって特定のエージェントに対して学習に有利な状況が生まれる，ということを防ぐためである．

4.4.3 実験に用いるタスク

エージェントの行うタスクは，最大の報酬が得られるルートの探索である．エージェントは環境の各状態において行動を選択し, 次の状態へと遷移することができる. その結果, 遷移先の状態に応じてエージェントは報酬を得る．このような環境内で R 回行動を行い，その中で得られる報酬の総和を最大にすることがエージェントの目的となる．

ここで，環境の各状態における報酬の設定は，4.2 での実験と同様，4.2.3 の式（4.1）を用いる．

4.4.4 実験設定

本実験におけるその他の設定を以下に示す．

・環境に関する設定

・環境に関する設定

本実験において，エージェントが学習を行う環境は，N 種類の要素から構成される環境とし，各要素の値 E_i(i=1～N)は，E_i={0,1}の 2 値をとるものとした．また，環境の状態は，各要素の組み合わせによって決定される．

今回，各状態における遷移先は以下の条件に従ってランダムに設定した．

・ 環境の各状態は，Tr 種類の遷移先を持つ.

・ 環境のどの状態も，他の全ての状態へ遷移可能なルートを持つ

以上のような環境下で，エージェントは学習を行う．

・エージェントに関する設定

本実験では，エージェントの設定を次のようにして学習を行わせる．エージェントの選択可能な行動は, 遷移先の数と同数とする. 1～N 種類のセンサを持った各エージェントに，冗長センサを 0～M 種類持たせたものを用意する．さらにセンサ数 0 のエージェントを加え，計(N*M)+1 体のエージェントに学習を行わせる．また，今回の環境では，学習が行われない場合でも，ある程度の報酬が得られることが予想される．そのため，学習を行わずランダムに行動して報酬を得るエージェントを，同時に行動させ，比較対象とする．以上より，本実験では(N*M)+2 体のエージェントが同一環境内で行動する．

・学習手法に関する設定

本実験では，第 3 章で述べたように強化学習を用いて実験を行う．強化学習の手法には，代表的な手法である Q 学習を用いる．行動選択手法は，代表的な手法として ε -greedy 法・Softmax 法・追跡手法などが挙げられるが，その中でも今回は Softmax 法を用いる．

・パラメータ設定

各パラメータ設定を表 4.8～4.11 に示す.

表 4.8　環境の設定に関するパラメータ

N（環境を構成する要素の数）	10
Tr（環境の各状態における遷移先の数）	2
S(E)　（環境の取り得る状態の数）	1024

表 4.9　エージェントやタスクに関するパラメータ

M	9
α_{exp}	0.65

表 4.10　学習手法に関するパラメータ

α	0.7
γ	0.6
τ	1000

表 4.11　学習回数に関するパラメータ

R（学習回数）	50000
S（学習を行う環境の種類）	20

4.4.5　実験結果

以上の設定で実験を行った．その結果を図 4.12，図 4.13 に示す．

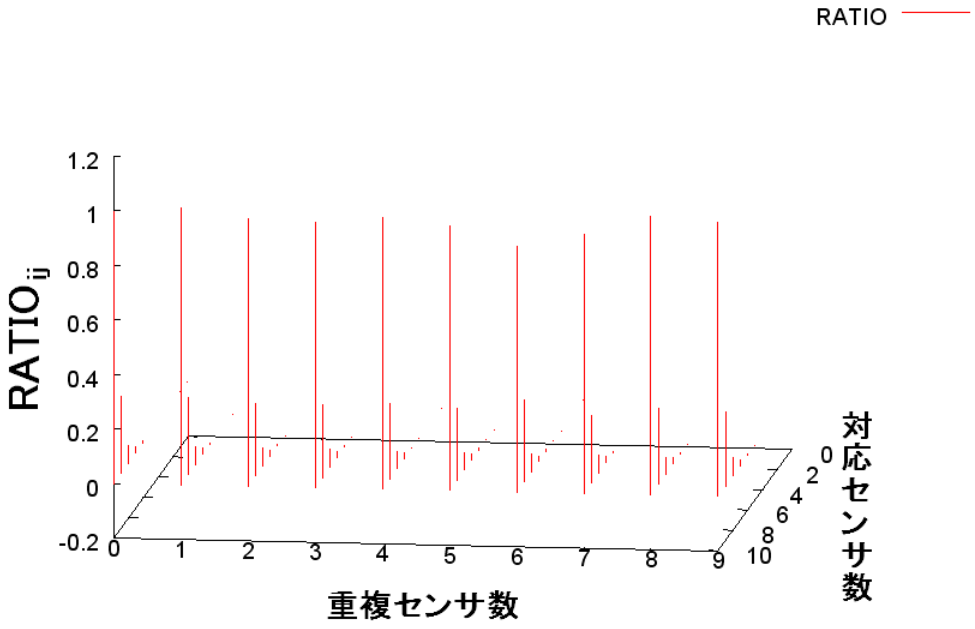

図 4.12　各エージェントの獲得報酬の総和の比較

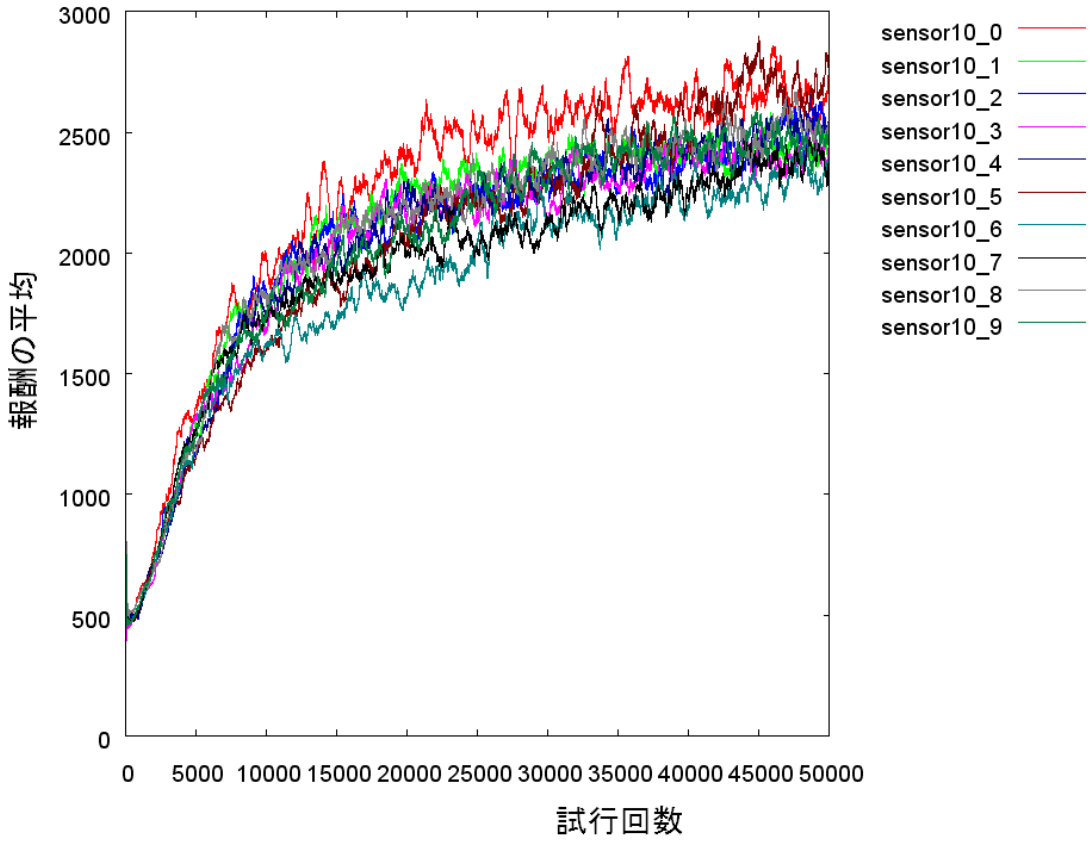

図 4.13　対応センサ数が 10 のエージェントの獲得報酬の推移

どちらの結果を見ても，重複センサ数にかかわらず，対応センサの数が同じエージェント同士はほぼ同じ学習効率となっている．つまり，重複センサは学習効率に全く影響を与えていないということがわかる．

重複センサが増えた場合，エージェントが保持する状態数が増えることは前章で述べた．しかし，重複センサは一つの状態に対し，どのセンサも同じタイミングで認識を行う．そのため実際に要素の値を認識するとき，重複センサ同士はすべて同じ値を認識することになる．つまり，重複センサ同士が別々の値を認識することはなく，実際に認識される状態数は重複センサが増加しても変わらない．以上より，重複センサは環境認識に影響を与えることはなく，学習にも影響がないといえる．

4.4.6　考察

今回の実験では，重複センサの有無は学習に関係がないことがわかった．そのため，学習効率の検証の際には，重複センサの有無について考える必要はない．ただし前章で述べ

たように，重複センサが増加するとエージェントが保持する状態数も増加する．実際に認識する状態数は変わらなくても，メモリ上で保持する状態数は変わらないと考えられる．このように，学習以外の部分に影響を与える可能性はあるので，注意が必要である．

また，重複センサが学習に影響を及ぼさないということは，あえて重複センサを増やしても問題が無いともいえる．この特徴を利用できる例として，実機のロボットへのセンサ設置を考える．実機のロボットのセンサは故障する可能性を考えなくてはいけない．このとき，あえて重複センサを設置することで，一つのセンサが故障しても他の重複センサによって故障をカバーすることが可能である．このような場合には，学習効率への影響を考えずに重複センサを設置することができる．

4.5 冗長なセンサを用いた実験 2

4.5.1 実験の目的

本実験では，エージェントが冗長なセンサを持つ場合，センサの種類数の違いが学習に及ぼす影響について，定常環境を用いて検証することを目的とする．本節では，冗長センサの中でもノイズセンサに注目して実験を行う．

4.5.2 実験方法

実験方法は，4 章で述べた検証方法を元にする．N 種類の要素からなる環境上で，各センサ数のエージェントに同時に学習を行わせる．エージェントは，センサ数が 0～N の各エージェントに対し，冗長センサ数が 0～M のエージェントを用意する．以上より，((N+1)*M) 体のエージェントに学習を行わせる．また，学習効果の比較のため，学習を行わず，ランダムに行動するエージェントも，同時に行動させることとする．

実験は次のような流れで行った．1 個の環境に対し，各エージェントに R ステップの学習を行わせる．これを S 種類の環境に対して行い，その結果獲得した報酬の平均を用いて学習効率の比較を行う．今回用いる環境はランダムに生成される．そのため，環境の作り方によっては特定のエージェントに対して有利な状況が生まれる可能性がある，そこで S 種類の環境での平均を用いることで，環境によって特定のエージェントに対して有利な状況が生まれる，ということを防ぐ．

4.5.3　実験に用いるタスク

エージェントの行うタスクは，最大の報酬が得られるルートの探索である．エージェントは環境の各状態において行動を選択し, 次の状態へと遷移することができる. その結果, 遷移先の状態に応じてエージェントは報酬を得る．このような環境内で R 回行動を行い, その中で得られる報酬の総和を最大にすることがエージェントの目的となる.

ここで，環境の各状態における報酬の設定は，4.2 での実験と同様，4.2.3 の式（4.1）を用いる.

4.5.4　実験設定

本実験におけるその他の設定を以下に示す.

・環境に関する設定

本実験において，エージェントが学習を行う環境は，N 種類の要素から構成される環境とし，各要素の値 E_i(i=1～N)は，E_i={0,1}の 2 値をとるものとした．また，環境の状態は，各要素の組み合わせによって決定される.

今回，各状態における遷移先は以下の条件に従ってランダムに設定した.

- 環境の各状態は，Tr 種類の遷移先を持つ.
- 環境のどの状態も，他の全ての状態へ遷移可能なルートを持つ

以上のような環境下で，エージェントは学習を行う.

・エージェントに関する設定

本実験では，エージェントの設定を次のようにして学習を行わせる．エージェントの選択可能な行動は, 遷移先の数と同数とする. 0～N 種類のセンサを持った各エージェントに, 冗長センサを 0～M 種類持たせたものを用意し， 計((N+1)*M)+1 体のエージェントに学習を行わせる．また，今回の環境では，学習が行われない場合でもある程度の報酬が得られることが予想される．そのため，学習を行わずランダムに行動して報酬を得るエージェントを同時に行動させる．以上より，本実験では((N+1)*M)+2 体のエージェントが同一環境内で行動する.

・学習手法に関する設定

本実験では，第 3 章で述べたように強化学習を用いて実験を行う．強化学習の手法には,

代表的な手法である Q 学習を用いる．行動選択手法は，代表的な手法として ε -greedy 法・Softmax 法・追跡手法などが挙げられるが，その中でも今回は Softmax 法を用いる．

・パラメータ設定

各パラメータ設定を表 4.12～4.15 に示す．

表 4.12　環境の設定に関するパラメータ

N（環境を構成する要素の数）	10
Tr（環境の各状態における遷移先の数）	2
S(E)（環境の取り得る状態の数）	1024

表 4.13　エージェントやタスクに関するパラメータ

M	9
α_{exp}	0.65

表 4.14　学習手法に関するパラメータ

α	0.7
γ	0.6
τ	1000

表 4.15　環境の設定に関するパラメータ

R（学習回数）	50000
S（学習を行う環境の種類）	20

4.5.5　実験結果

以上の設定で実験を行った．その結果を図 4.14～4.16 に示す．

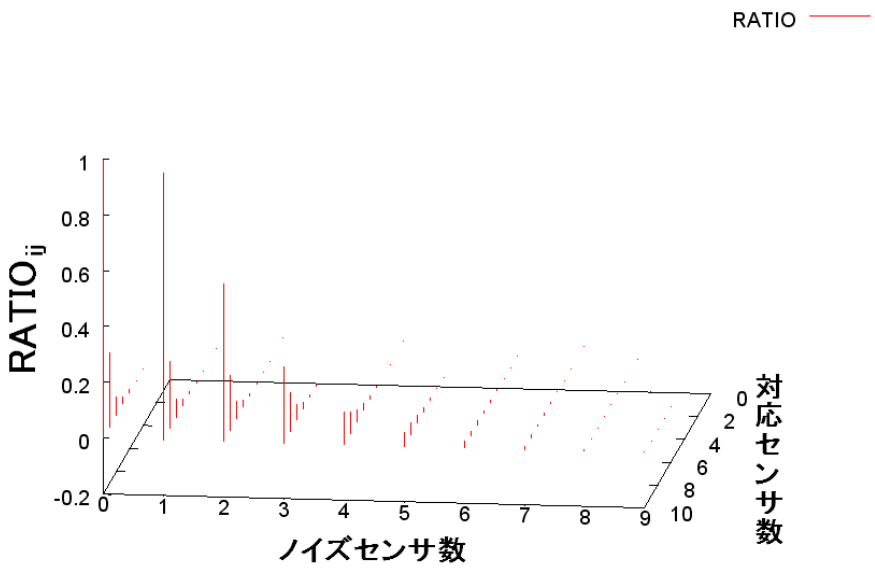

図 4.14　獲得報酬の総和の比較

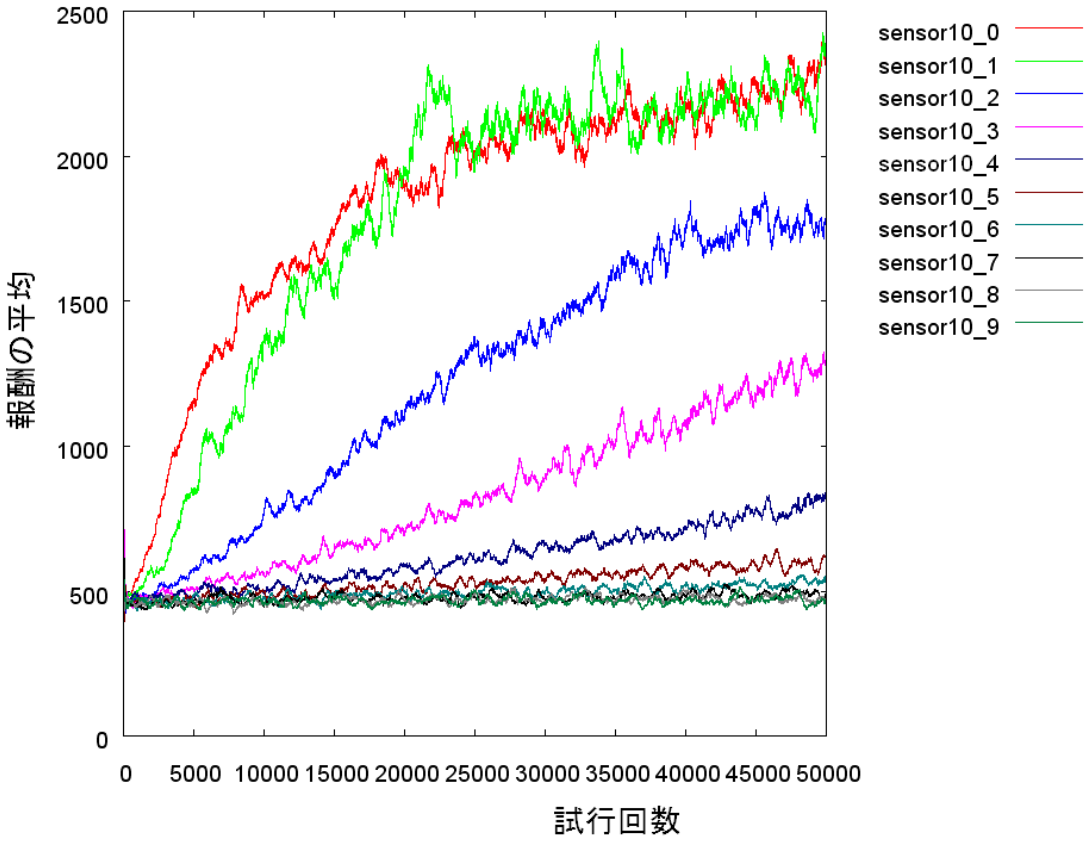

図 4.15　対応センサ数が 10 のエージェントの獲得報酬の推移

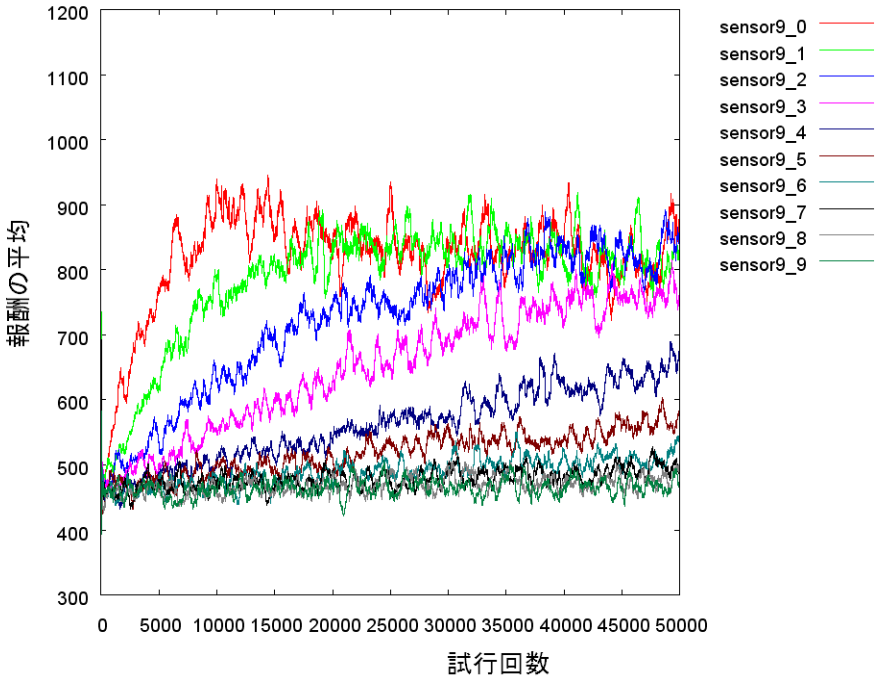

図 4.16　対応センサ数が 9 のエージェントの獲得報酬の推移

まず, 図 4.14 をみると, ノイズセンサが増える毎に学習効率が落ちているようにみえる. しかし，学習効率の変化の仕方は冗長でないセンサの数によっても変わっているようにみえる．図 4.14 の結果の傾向や原因を探るため，対応センサ毎の獲得報酬の推移の結果を図 4.15，図 4.16 に示す.

まずはセンサ数 10 のエージェントに注目する．センサ数 10 のエージェントの獲得報酬の推移を図 4.15 に示す．図 4.15 をみると，ノイズセンサが増える毎に獲得報酬の増え方が緩やかになっていることがわかる．つまり，ノイズセンサが増える毎に学習の速度が落ちているといえる．また，学習の速度は落ちても，ある程度学習が進めば，冗長センサを持つエージェントの獲得報酬が冗長センサを持たないエージェントに追いつく傾向も見える. しかし 4.2 の結果より，センサ数 10 のエージェントが完全に学習が完了するまでには非常に時間がかかるため，この実験結果では確認できない．そこで次に，センサ数 9 のエージェントに注目する.

センサ数 9 のエージェントの獲得報酬の推移を図 4.16 に示す．図 4.16 を見ると，図 4.15 と同じく，冗長センサの増加と共に学習速度が遅くなっていることがわかる．しかし，ある程度学習が進んだエージェントに関しては，冗長センサを持たないエージェントと同じ学習効率を示している．今回の実験では学習回数を 50000 回としたが，全てのエージェントが学習完了するまで学習を行えば，冗長センサがいくつであっても，最終的な学習効果

は同じになると考えられる．

以上より，ノイズセンサが増えると学習の速度が遅くなることがわかった．しかし，十分な学習時間を取れる場合には，学習完了後の学習効率は，ノイズセンサの数によらないということがわかった．

4.5.6 考察

今回の実験で，ノイズセンサが増えると学習速度が落ちることがわかった．そのため，短時間での学習が必要な状況ではノイズセンサの有無が結果に大きく影響を与えるため，ノイズセンサを持たないよう注意して設計する必要がある．

しかし，ノイズセンサを持っていても，学習が完了すればノイズセンサを持たないエージェントと同程度の学習効果を期待できる．そのため，学習時間が十分にとれるような場合であれば，ノイズセンサについてはあまり考える必要はない．

4.6 冗長なセンサを持つエージェントを用いた実験 3

4.6.1 実験の目的

本実験では，エージェントが冗長なセンサを持つ場合，センサの種類数の違いが学習に及ぼす影響について，非定常環境を用いて検証することを目的とする．本節では，冗長センサの中でもノイズセンサに注目して実験を行う．

4.6.2 実験方法

実験方法は，4 章で述べた検証方法を元にする．N 種類の要素からなる環境上で，各センサ数のエージェントに同時に学習を行わせる．エージェントは，センサ数が 0〜N の各エージェントに対し，冗長センサ数が 0〜M のエージェントを用意する．以上より，((N+1)*M) 体のエージェントに学習を行わせる．また，学習効果の比較のため，学習を行わず，ランダムに行動するエージェントも，同時に行動させることとする．

実験は，次のような流れで行った．1 個の非定常環境に対し，各エージェントに R ステップの学習を行わせる．これを S 種類の非定常環境に対して行い，その結果獲得した報酬の平均を用いて学習効率の比較を行う．今回用いる環境はランダムに生成される．そのため，環境の作り方によっては特定のエージェントに対して有利な状況が生まれる可能性がある，そこで S 種類の環境での平均を用いることで，環境によって特定のエージェントに対して有利な状況が生まれる，ということを防ぐ．

4.6.3 実験に用いるタスク

エージェントの行うタスクは，最大の報酬が得られるルートの探索である．エージェントは環境の各状態において行動を選択し, 次の状態へと遷移することができる. その結果, 遷移先の状態に応じてエージェントは報酬を得る．このような環境内で R 回行動を行い, その中で得られる報酬の総和を最大にすることがエージェントの目的となる.

ここで，環境の各状態における報酬の設定は，4.2 での実験と同様，4.2.3 の式（4.1）を用いる.

4.6.4 実験設定

本実験におけるその他の設定を以下に示す.

・環境に関する設定

本実験において，エージェントが学習を行う環境は，N 種類の要素から構成される環境とし，各要素の値 E_i(i=1～N)は，E_i={0,1}の 2 値をとるものとした．また，環境の状態は，各要素の組み合わせによって決定される.

今回，各状態における遷移先は以下の条件に従ってランダムに設定した.

- 環境の各状態は，Tr 種類の遷移先を持つ.
- 環境のどの状態も，他の全ての状態へ遷移可能なルートを持つ

以上のように環境は生成される.

また，今回の実験で用いる環境は，環境変動の起こる非定常環境である．環境変動の設定は次のように行った．エージェントが Ch 回試行を行うと，環境は変動を起こす．環境変動は，ある状態から次の状態への遷移先を別の状態へ遷移するように変更することで行う．遷移先の変更は，各遷移先について P%の確率で行われる．このとき遷移先の変更は，初期の遷移先の設定と同じ条件でランダムに行われる.

本実験では，各状態の距離に応じて報酬の設定が行われている．そのため，遷移先が変更されると状態間の距離にも変化が生じる．そこで，環境変動が行われたときには，その距離に応じて各状態の報酬を再設定する. このときの報酬設定も, 4.2 での実験と同様, 4.2.3 の式（4.1）を用いる.

・エージェントに関する設定

本実験では，エージェントの設定を次のようにして学習を行わせる．エージェントの選択可能な行動は, 遷移先の数と同数とする. 0～N 種類のセンサを持った各エージェントに,

冗長センサを 0～M 種類持たせたものを用意し， 計((N+1)*M)+1 体のエージェントに学習を行わせる．また，今回の環境では，学習が行われない場合でもある程度の報酬が得られることが予想される．そのため，学習を行わずランダムに行動して報酬を得るエージェントを同時に行動させ比較対象とする．以上より，本実験では((N+1)*M)+2 体のエージェントが同一環境内で行動する．

・学習手法に関する設定

本実験では，第 3 章で述べたように強化学習を用いて実験を行う．強化学習の手法には，代表的な手法である Q 学習を用いる．行動選択手法は，代表的な手法として ε -greedy 法・Softmax 法・追跡手法などが挙げられるが，その中でも今回は Softmax 法を用いる．

・パラメータ設定

各パラメータ設定を表 4.16～4.20 に示す．

表 4.16　環境の設定に関するパラメータ

N（環境を構成する要素の数）	10
Tr（環境の各状態における遷移先の数）	2
S(E) （環境の取り得る状態の数）	1024

表 4.17　環境変動に関するパラメータ

Ch	2500
P	10

表 4.18　エージェントやタスクに関するパラメータ

M	9
α_{exp}	0.65

表 4.19　学習手法に関するパラメータ

α	0.7
γ	0.6
τ	1000

表 4.20　環境の設定に関するパラメータ

R（学習回数）	50000
S（学習を行う環境の種類）	20

4.6.5　実験結果

実験結果を図 4.17～4.19 に示す.

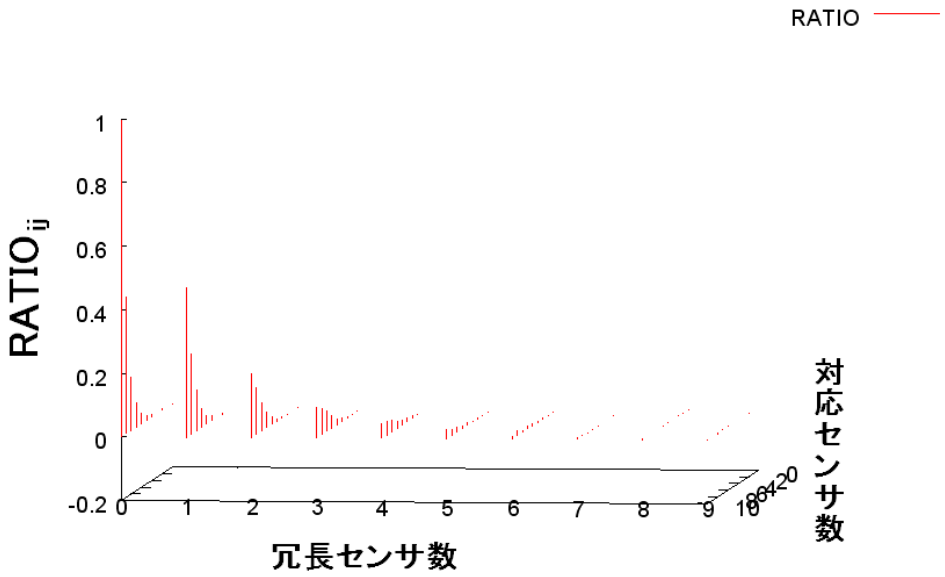

図 4.17　獲得報酬の総和の比較

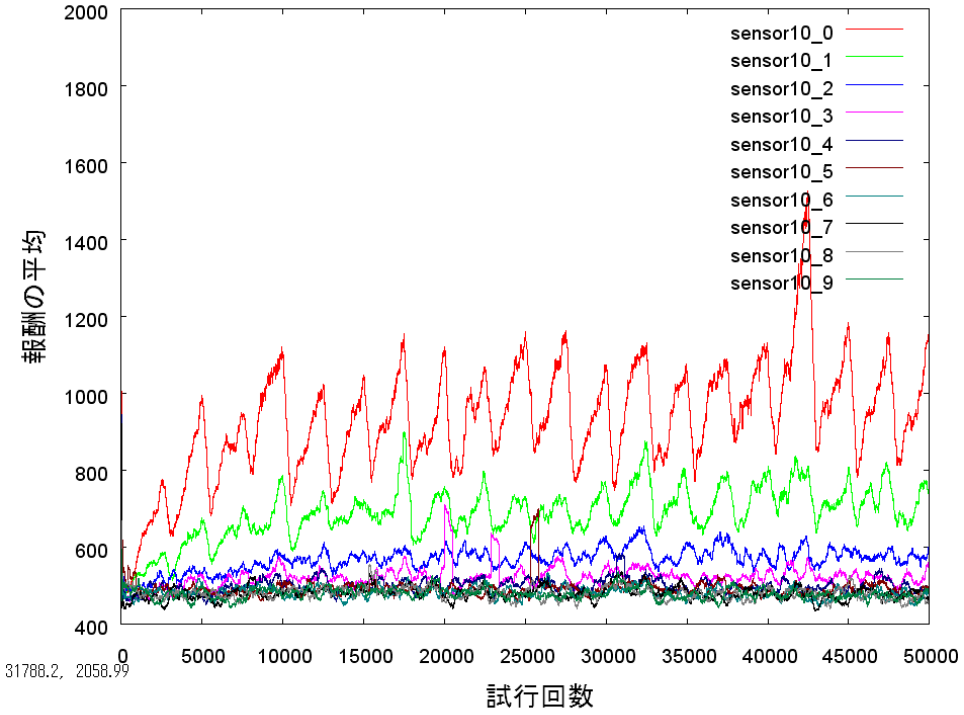

図 4.18　対応センサ数 10 のエージェントの獲得報酬の推移

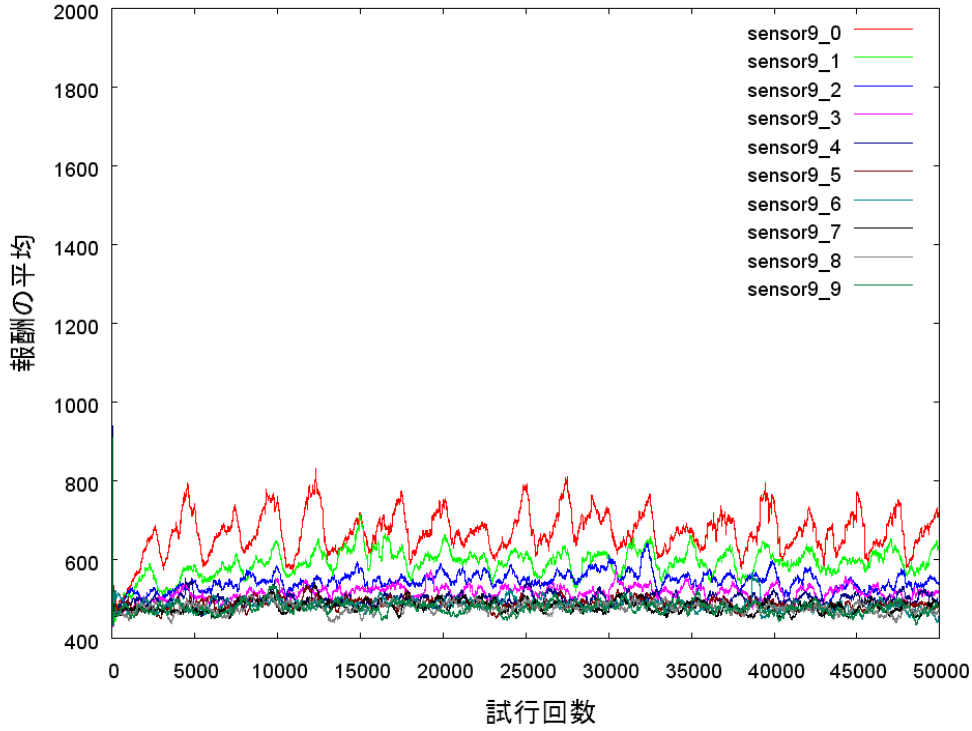

図 4.19　対応センサ数 9 のエージェントの獲得報酬の推移

まず図 4.17 を見ると，定常環境での実験に比べてノイズセンサの増加に対する学習効率の落ち方が大きくなっている．この原因は図 4.18 に現れている．図 4.18 は対応センサ数 10 のエージェントの獲得報酬の推移である．この図では，ノイズセンサを持つエージェントがノイズセンサを持たないエージェントの獲得報酬に追いつくことがなくなっている．これは，環境変動の度に再学習が必要となり，そのたびに学習に遅れが生じるエージェントは大きく学習効率が落ちるためだと考えられる．

同じように，対応センサが 9 個のエージェントについての結果を図 4.19 に示す．図 4.19 を見ると，やはり定常環境の場合と違い，試行回数が増えてもノイズセンサを持つエージェントがノイズセンサを持たないエージェントの獲得報酬に追いつくことは無い．以上より，学習の即応性が求められる非定常環境下では，学習速度を遅らせるノイズセンサは学習に大きく悪影響を与えるといえる．

4.6.6 考察

今回の実験で，非定常環境下ではノイズセンサによる学習効率の低下が大きくなることがわかった．また，その原因は学習速度の低下によるものであることもわかった．そのため，非定常環境のように素早い学習が求められるような環境では，ノイズセンサが存在しないように設計するよう十分に注意する必要がある．このことは，4.5.5 の実験結果の考察を実証する結果でもある．

4.7 考察

本節では，本実験全体の考察を行う．本実験では，センサの種類数の変化や冗長センサ数の変化が学習に与える影響について検証を行った．その結果判明した各センサの特徴と，その特徴に関する考察を以下に簡潔に示す．

1.冗長センサを持たず，対応センサ数のみが変化する場合

・十分に学習を行った場合，センサ数に応じて指数関数的に学習効率が高くなる

学習時間を十分に確保できる環境では，対応センサをなるべく多く設置することが非常に重要であるといえる．

・学習初期段階では，センサ数が少し少ないエージェントの方が良い結果を示す．もしくは，センサ数が多いエージェントに近い学習効率を示す．

学習時間が十分に確保できないような環境では，対応センサ数を少し少なく設置する

ことが有効であるといえる．

・センサ数があまりにも少ない場合は，どんな環境でも学習効果を期待できない
どんな環境に対しても，最低限のセンサを設置することは必須条件である．

.

2.重複センサを持つ場合

・重複センサは学習に何の影響も与えない
学習への影響のみを考える場合は，重複センサの有無を考慮する必要は無い．

3.ノイズセンサを持つ場合

・ノイズセンサが増加すると，学習速度が遅くなっていく
素早い学習が求められる環境では，ノイズセンサの設置は厳禁となる．

・学習完了後は，ノイズセンサを持っていないエージェントと同じ学習効率を示す
学習時間を十分に確保できるような環境であれば，重複センサと同じようにあまり気にする必要はない．

以上のように，各センサの設置の仕方によって，学習に様々な影響があることがわかった．影響を与える要素として特に多いのが，学習時間に関するものである．いつでも十分な学習時間を確保できるならば，センサを設置できるだけ設置すればよい．しかし実際にロボットに学習を行わせる際は，学習する時間が限られる場合や素早い学習が求められる場合が多い．そのため，学習手法の有効性の検証の際にセンサの種類数を考慮することは非常に重要であるといえる．

第 5 章　まとめと今後の課題

5.1　まとめ

本論文では，環境認識能力の違いが学習に及ぼす影響についての検証を目的とし，実験を行った．まず始めに，環境認識能力に変化を及ぼすセンサの要素として，センサの種類数・センサの分解能・センサのサンプリング周波数を挙げた．これらが環境認識に与える影響を考え，中でもセンサの種類数が学習に与える影響について検証を行った．また，センサが増えることによって生まれる冗長センサについても，環境認識能力に変化を及ぼす要素として挙げた．冗長センサは重複センサ・ノイズセンサ・不用センサの三種類に分けて考え，このうち重複センサとノイズセンサが学習に与える影響についての検証を行った．

検証方法として，センサ数の異なるエージェントを同一環境内で同時に学習を行わせ，その学習効果を比較するという方法で行った．また，定常環境と非定常環境の二種類の環境を用い，シミュレーションでの実験を行った．

実験の結果，センサによって学習に様々な影響を与えることが判明した．まず，対応センサのみの場合，学習回数が多い場合にはセンサ数が多いエージェントほど高い学習効率を示した．また，学習回数が少ない場合には，センサ数が少し少ないエージェントの方が高い学習効率を示した．しかし，センサ数があまりにも少ない場合には，学習効果が期待できないこともわかった．次に，重複センサについては，学習には全く影響が無いことがわかった．最後にノイズセンサについて，ノイズセンサが増えると学習の速度が遅くなっていくことがわかった．ただし，十分に学習を行えば，ノイズセンサがいくつであっても同じ学習効率を示すこともわかった．

以上のように，センサの種類数が変化すると，学習に様々な影響を与えることがわかった．そのため，学習手法の有効性を検証する際には，こうしたセンサの影響を加味して考察することが重要であるといえる．

5.2　今後の課題

・他のセンサ能力に関する実験

本論文では，センサの種類数の変化による学習への影響について検証を行った．しかし，学習に影響を与えると思われるセンサ能力として，センサの分解能とセンサのサンプリング周波数が挙げられる．これらのセンサ能力の変化が学習に与える影響について検証を行

う必要がある．また，本論文では冗長センサのうち重複センサとノイズセンサの影響についても検証を行った．しかし，冗長センサとしてもうひとつ不用センサが挙げられる．この不用センサについても今後検証を行う必要がある．

・別の環境を用いた実験

本論文では, 定常環境と非定常環境という二種類の環境について実験を行った. しかし, ロボットをとりまく環境は他にも様々なものが考えられる．例えば，今回用いた環境は確定的に遷移する環境であったが，遷移が確率的である環境も考えられる．また，今回用いた非定常環境は周期的に変動したが，不定期に変動する環境やロボットの行動によって変動する環境なども考えられる．こうした様々な環境に対しても検証を行う必要がある．

・センサ以外の身体構造が学習に与える影響について

本論文ではセンサが学習に与える影響を検証した．しかし，ロボットが持つ他の身体構造としてアクチュエータが挙げられる．そのため，アクチュエータの性能の変化が学習に与える影響についても検証を行う必要がある．

謝辞

本論分を結ぶにあたり，日ごろより懇切なるご指導を賜りました倉重太郎先生に深く感謝の意を表します．また，ご助言，ご指導頂いた畑中雅彦先生，渡辺修先生，渡邉真也先生に感謝の意を表します．そして，論文の査読や助言をしていただいた認知ロボティクス研究室の池田憲弘さん，木島康隆さん，宮崎愛央君に感謝いたします．

参考文献

[1]藤田義弘，“パーソナルロボット R100”，日本ロボット学会誌，Vol18，No2，pp?，1998

[2]山本大介，松日楽信人，土井美和子，“ユーザーと家電をつなぐロボットインタフェース”，日本ロボット学会誌，Vol26，No8，pp893-894，2008

[3]沢井邦仁，“” QRIO“におけるデザインプロセス”，日本ロボット学会誌，Vol22，No8，pp27-31，2004

[4]山田誠二，“HAL 研究のオリジナリティ”，人工知能学会誌，Vol24，No6，pp810-817，2009

[5]永谷圭司，“ロボット化社会”，日本ロボット学会誌，Vol26，No7，pp30-31，2008

[6]森川幸人，“マッチ箱の AI”，新紀元社，2000

[7]山口明彦，杉本徳和，川人光男，“回避行動の採用メカニズムを備えた強化学習手法と多関節ロボットの全身運動学習への応用”，日本ロボット学会誌，Vol27，No3，209-220，2009

[8]宮崎和光，木村元，小林重信，“Profit Sharing に基づく強化学習の理論と応用”，人工知能学会誌，Vol14，No5，pp800-807，1999

[9]新井幸代，宮崎和光，小林重信，“マルチエージェント強化学習の方法論-Q-Learning と Profit Sharing による接近-”，人工知能学会誌，Vol13，No4，pp105-114，1998

[10]山村忠義，馬野元秀，瀬田和久，“段階的な資格を持つエージェントにおける強化学習について-追跡問題を例にして-”，知能と情報，Vol18，No4，pp561-570，2006

[11]伊藤一之，高山明宏，“身体と環境の特性を利用した状態-行動空間の抽象化　-強化学習を用いた自立ヘビ型ロボットへの適用-”，知能と情報，Vol21，No3，pp402-403，2009

[12]石黒章夫，“知の基盤としてのしぶとさの創成”，日本ロボット学会誌，Vol24，No7，pp26-29，2006

[13]浅田稔，石黒浩，國吉康夫，“認知ロボティクスの目指すもの”，日本ロボット学会誌，Vol17，No1，pp2-6，1999

[14]Richard．S．Sutton，Andrew．G．Barto，“Reinforecement Learning”，The MIT Press，1998