

平成 2 1 年度

修士学位論文

題目 ロボットの認識能力の変化が学習に及ぼす影響について-センサの能力の違いが強化学習の学習効率に及ぼす影響-

提出者 室蘭工業大学大学院工学研究科

情報電子工学系

専攻

平成 22年 4月 入学

氏名 中南 義典

提出年月日 平成 24年 1月 31日

室蘭工業大学大学院

目次

1.	はじめに.....	4
1.1.	研究背景.....	4
1.2.	本論文の構成.....	8
2.	調査に用いるセンサのモデル化.....	9
2.1.	センサのモデル化の必要性.....	9
2.2.	学習を行うロボットにおけるセンサの役割.....	10
2.3.	センサのモデル化.....	11
3.	センサが学習効率に及ぼす影響の検証方法.....	13
3.1.	調査の対象.....	13
3.1.1.	強化学習概要.....	13
3.1.2.	センサの出力数と強化学習の関係.....	14
3.1.3.	センサの過不足と学習の関係.....	17
3.2.	センサの不足が学習効率に及ぼす影響の検証方法.....	19
3.3.	不要センサが学習に及ぼす影響の検証方法.....	21
4.	センサの不足が学習効率に及ぼす影響の実験.....	23
4.1.	実験概要.....	23
4.2.	実験に用いる学習手法.....	23
4.3.	実験に用いるタスク.....	25
4.4.	結果の比較方法.....	27
4.5.	実験パラメータ.....	29
4.6.	実験結果.....	29
4.7.	実験結果の考察.....	31
5.	過剰なセンサが学習に及ぼす影響の実験.....	33
5.1.	実験概要.....	33
5.2.	実験に用いるタスク.....	33
5.3.	結果の比較方法.....	35
5.4.	実験パラメータ.....	37
5.5.	実験結果.....	38
5.6.	実験結果の考察.....	39
6.	タスク及び所持センサの違いに対する調査.....	42
6.1.	調査対象の拡張.....	42
6.2.	センサの縮退による調査の単純化.....	44
6.3.	状態数の異なるタスクにおける検証方法.....	48
6.3.1.	調査対象のセンサが必要センサの場合.....	48
6.3.2.	調査対象のセンサが不要センサの場合.....	49

6.4.	ロボットの所持センサの違いに関する検証方法	51
6.4.1.	具体的な調査対象	51
6.4.2.	検証方法	52
7.	タスク及び所持センサの違いに対する実験方法	57
7.1.	タスクの違いに対する必要センサの影響の実験方法	57
7.1.1.	実験の流れ.....	57
7.1.2.	実験に用いるタスク	58
7.1.3.	結果の比較方法.....	59
7.2.	タスクの違いに対する不要センサの影響の実験方法	60
7.2.1.	実験の流れ.....	60
7.2.2.	実験に用いるタスク	61
7.2.3.	結果の比較方法.....	62
7.2.4.	実験パラメータ	63
7.3.	ロボットの所持センサの違いに関する実験方法	63
7.3.1.	実験の流れ.....	63
7.3.2.	実験に用いるタスク	65
7.3.3.	結果の比較方法.....	65
7.3.4.	実験パラメータ	69
8.	タスク及び所持センサの違いに対する実験結果	70
8.1.	タスクの違いに対する必要センサの影響の実験	70
8.1.1.	実験結果.....	70
8.1.2.	考察	73
8.2.	タスクの違いに対する不要センサの影響の実験	73
8.2.1.	実験結果.....	73
8.2.2.	実験結果の考察.....	77
8.3.	ロボットの所持センサの違いに関する実験結果	78
8.3.1.	センサが不足しているロボットに対する必要センサの影響の実験結果.....	78
8.3.2.	センサが不足しているロボットに対する不要センサの影響の実験結果.....	79
8.3.3.	センサを過剰に所持するロボットに関する実験結果	81
8.3.4.	実験結果の考察.....	82
8.3.5.	調査結果のまとめ	83
9.	まとめ	84

謝辞

参考文献

研究業績

1. はじめに

1.1. 研究背景

近年では技術の発達により，ロボットの活躍する場面が徐々に広がっている．ロボットが登場した当時は，工場や特定の研究室などの限定された空間で単純な作業を行うのみであった．しかし現在では技術の発達に伴いロボットの行える仕事が増えている[1][2]．それと共に，ロボットの使用環境を家庭や公共施設などに拡大していくことが期待されている[3]．また，ごく少数ではあるが，実際に家庭などで利用されるロボットも見かけることができる．

工場ではロボットが使用される環境がかっちりと決まっており，ロボットの使用される状況を容易に想定することができる．しかし家庭や施設などは使用される環境が一定ではない．例えばリビングの広さや，家具の位置や種類などは家庭によって異なる．過程でロボットを使用するには，こうした環境の違いに対してもロボットが適切に行動できなくてはならない．しかし，家庭のように使用環境が多様に存在する場合にはロボットの使用される状況全てを想定することは困難である．そこで，ロボット自身が自律的に，未知の環境に対して適応的な行動を獲得することが望まれるようになった．

未知環境に適応するための手法として注目されているのが学習である[3]-[5]．学習という手法自体はロボットに限られた技術ではなく，様々なソフトウェアに利用されている[6][7]．学習を行うソフトウェアは，外部からの入力に対して適切な出力を自律的に学習していく．たとえば日本語の文字列の変換ソフトであれば，人間からある文字列がソフトウェアに対して入力される．ソフトウェアは入力された文字列に対して，変換後の文字列の候補を人間に対して出力する．この行動を繰り返して学習が進めば，人間にとってより望ましい変換候補を，より優先的に表示するようになる．この学習という手法はロボットにとっても非常に有用である．ロボットが現在置かれている環境の状態を入力とし，環境の状態に対して適切な出力を学習することで，未知環境に対しても適切に学習を行うことができるようになる．このような理由から，現在多くのロボットに学習が利用されている．

ロボットに学習を適用する際には，ロボット特有の問題が生じる．それは，ロボットが身体を持つ学習者であるという点である[8]-[10]．ロボットの学習の様子を Fig.1.1 に示す．ロボットが学習を行う際，ロボットに対する入力はセンサを通して行われる．そのため，ロボットが認識する環境の状態はロボットが持つセンサの数やセンサの性能にゆ

だねられる。また、ロボットはセンサからの入力に対して学習結果を元に行動を出力するが、実際の行動はアクチュエータを用いて行われる。そのため学習結果に対して適切な行動を行えるかどうかは、アクチュエータによっても変わるといえる[11]-[14]。

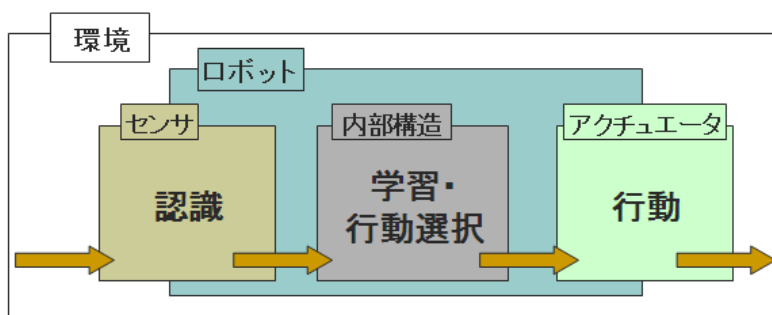


Fig.1.1 ロボットの学習の流れ

本論文ではロボットの身体の中でも特にセンサの重要性に注目する。ロボットはセンサからの入力に対して適切な出力を学習していく。すなわち、ロボットにとってはセンサを通して認識した環境の状態が学習の対象となるのである。タスクによって構成される状態空間と、センサによって構成される状態空間の関係を Fig.1.2 に示す。ロボットが学習を行うときには、まず初めに人間によって学習すべきタスクが与えられる。タスクが決まると、実環境の中から学習の対象となる状態空間が決まる。このときロボットの学習の対象となる空間は、タスクによって構成される空間となる。一方、ロボットが認識できる状態はロボット自身が持つセンサによって構成される状態空間のみである。そのためロボットから見たときに学習の対象となる空間は、センサによって構成される状態空間となる。ロボットが持つセンサが異なればセンサによって構成される状態空間も変化する。このときロボットから見た学習の対象となる状態空間も変化するため、ロボットの学習の結果も変化してしまう。ロボットに与えるタスクが決まればタスクの状態空間も決まるが、タスクの状態空間を人間が完全に把握できるとは限らない。ロボットに搭載するセンサは人間によって決められる。そのため人間が想定する状態空間とタスクの状態空間にずれが生じていけば、ロボットの認識する状態空間とタスクの状態空間の間にもずれが生じてしまう。このときタスクを構成する状態空間の一部が認識できない、あるいはタスクとは関係のない状態を認識するなどの事態が生じる。このとき、ロボットは本来の学習効率を發揮できなくなってしまう可能性がある。そのため、ロボットの学習においてはタスクに対して適切にセンサを搭載することが非常に重要であるといえる。

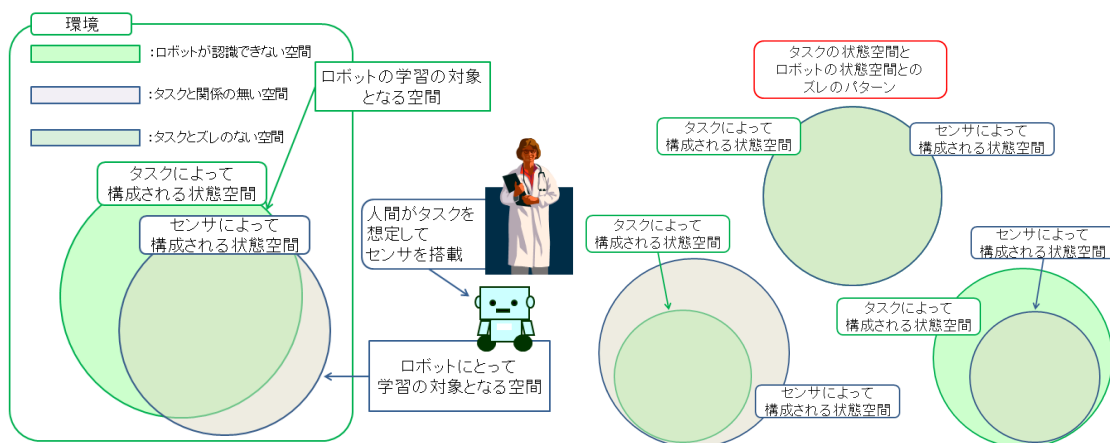


Fig.1.2 タスクの状態空間とセンサの状態空間の関係

ロボットの学習に関する従来研究として、ロボットの学習におけるセンサの重要性に注目したものがいくつか存在する[15]-[19]。これらの研究では従来の学習手法を利用し、ロボットに搭載するセンサのみを工夫することでロボットの学習効率を向上させている。山村らの研究[15]では追跡問題というタスクに対して、ロボットの視覚センサの認識範囲を段階的に調整することで学習にかかる時間の短縮することに成功している。また杉浦らの研究[16]では、ラインレースというタスクに対して適切なセンサの配置を、強化学習[20]や GA[]を利用してロボットに学習させた。学習の結果得られたセンサの組み合わせを用いてロボットに学習を行わせた結果、ラインレースの精度を向上させることに成功している。こうした従来研究を見ると、ロボットの学習においては学習手法だけでなく搭載するセンサも学習結果に関わっている。すなわち、ロボットの学習におけるセンサの重要性を確認することができる。

ただしこれらの従来研究は特定のタスクや特定のセンサを対象とした研究である。特定のタスクに対して特定のセンサをより適切に搭載するための方策を示しているが、タスクやセンサが異なる場合には同じ方法を利用できるとは限らない。ロボットが行うタスクや使用するセンサの種類に限定されずに、適切なセンサを搭載するための方策を示すことができれば、より多様なロボットに対して適切にセンサを搭載し、学習の効率を改善することができる。そのためには、ロボットのタスクや使用するセンサの種類に限定されず、センサが学習効率に及ぼす影響を調査する必要がある。

センサが学習効率に及ぼす影響を調査した先行研究 [21]がある。センサが学習効率に影響を及ぼす可能性のある要素としては、ロボットが持つセンサの数や、ロボットが持つ個々のセンサの能力が挙げられる。この研究では特に、ロボットが持つセンサの数の

違いに注目して調査を行っている。調査には実機のセンサを使うのではなく、モデル化したセンサを利用して調査を行っている。センサには全てのセンサに共通する基本的な性質と、具体的な個々のセンサ特有の性質がある。この研究ではセンサをモデル化する際に個々のセンサ特有の性質を排除し、全てのセンサに共通する基本的な性質のみを持つセンサとしてモデル化している。これにより、全てのセンサに共通して起こる学習効率への影響を調査している。そして調査の結果として、タスクの状態空間に対してセンサが不足している場合とセンサを過剰に所持している場合について、それぞれ学習効率に異なる影響を及ぼすことを示している。

この先行研究ではセンサのモデル化を行う際に、センサの能力についても定義を行っている。センサは基本的に出力しうる値の数が決まっており、その値の中からどれか一つのみをロボットに対して出力する。先行研究では、センサが出力しうる値の数によってセンサの能力を区別している。しかし先行研究では調査を単純化するために、ロボットが持つセンサの能力をすべて均一化して調査を行っている。そのため、センサの能力の違いに対する学習への影響の違いについては示されていない。ロボットが認識する状態空間はロボットが持つセンサからの入力によって決まる。センサから得られる入力の数が多いほど、ロボットが認識する状態空間も大きくなる。センサから得られる入力数は、ロボットが持つセンサの数や、個々のセンサの能力（出力しうる値の数）によって決まる。そのため、センサの数だけでなく個々のセンサの能力が異なる場合にも、センサが学習効率に及ぼす影響は異なると考えられる。センサの能力の違いに対する学習効率への影響を示すことができれば、センサの数だけでなくセンサの能力からも、ロボットの学習効率の変化を判断することができる。これにより、学習を行うロボットに対してより適切なセンサを搭載するための指針を示すこともできる。

そこで本論文では、センサの能力の違いに対する学習効率への影響について調査を行う。ただし、ロボットの学習手法にもさまざまな種類があり、全ての学習手法に対して調査を行うことは非常に困難である。そこで本論文では学習手法の中でも強化学習[]に注目する。強化学習は、ロボットに学習を適用する際に最も多く使われている学習手法である。そのため調査の対象を強化学習に絞り、センサの能力の違いに対する強化学習の学習効率への影響について調査を行う。

1.2.本論文の構成

本論文の構成について述べる. 1章ではロボットの学習とセンサとの関わりに関する背景について述べた. そして, センサの能力の違いが学習効率に及ぼす影響を調査する必要性について述べた.

2章では実験を行うために, センサのモデル化を行う. また, 本論文で扱うセンサの能力について明確に定義を行う. 3章では, 2章でモデル化したセンサを用いた検証方法について述べる. このとき, 調査に用いる学習手法である強化学習について述べた上で, 調査の対象のセンサをタスクに必要なセンサと不要なセンサに分類する. 4章では出力集の異なる必要センサが学習に及ぼす影響, 5章では出力数の異なる不要センサが学習に及ぼす影響について, 3章で述べた検証方法を元にしてそれぞれ実験を行い, 調査結果の考察を行う.

6章では, まず初めに5章までの実験に関する問題点を考察する. そして, 5章での調査では不足している点として, ロボットのタスクの違いやロボットの持つセンサの違いについて述べる. また, これらの調査を行うための検証方法についても述べる. 7章では6章で述べた検証方法を元にして, 詳細な実験の方法, 実験の設定について述べる. そして8章では調査実験の結果を示し, 調査結果の考察を行う. 最後に9章にて本論文全体のまとめを行い, 本論文で得られた成果と今後の研究課題について述べる.

2. 調査に用いるセンサのモデル化

2.1. センサのモデル化の必要性

センサが学習に及ぼす影響は、大きく分けて二つ考えることができる。一つは特定のセンサにのみ生じる学習への影響であり、もう一つは全てのセンサに共通して生じる学習への影響である。

センサは測定対象と測定方法に応じて様々な種類が存在し、その種類ごとに様々な特性を持っている。例えば距離を測るセンサである超音波センサは、周囲の温度や測定の対象物の硬さによって測定の精度が変化するという特性を持っている。一方、距離を測るセンサでもレーザーセンサの場合、測定の対象物の反射率や透明度によって測定の精度がするという特性を持っている。また、測定対象の違いを考えると、温度センサや距離センサは1次元の物理量を認識するが、カメラやマイクなどのセンサは2次元以上の物理量を認識する。こうした個々のセンサ特有の性質は、ロボットの学習に対しても特有の影響を及ぼすことが予想される。

一方で、全てのセンサに共通する性質も存在する。それは、環境の状態を認識し、認識した状態を何らかの形で出力するというものである。この性質がロボットの学習に及ぼす影響はいかなるセンサにも共通するものであると考えることができる。

センサが学習に及ぼす影響を考えると、これら二つの影響はどちらもロボットに対して重要である。しかし、これら全ての影響を一度に調査することは不可能である。そこで、より優先度の高い影響として、全てのセンサに共通して生じる学習への影響を考える。仮に特定のセンサに生じる学習への影響を調査することを考えた場合、全てのセンサに共通して生じる学習への影響が不明なまま調査を行うことになる。この場合、個々のセンサ特有の影響と全てのセンサに共通の影響との切り分けが難しくなる。また、膨大に存在するセンサ全てに対して調査を行うことは現実的に難しいという側面もある。一方全てのセンサに共通して生じる学習への影響を調査する場合には、モデル化したセンサを用いることで、個々のセンサ特有の影響を排除する形で調査を行うことができる。また、調査の結果は全てのセンサに対して適用可能であるので、応用の幅も広がる。そのため本論文ではモデル化したセンサを用いて、全てのセンサに共通して生じる学習への影響を調査する。

2.2.学習を行うロボットにおけるセンサの役割

センサのモデル化を行うにあたり，学習を行うロボットにおけるセンサの役割について述べる．まず初めに，ロボットの用いるセンサが出力する値の性質について述べる．センサは一般的に環境の状態を認識したあと，認識した状態を何らかの形に変換して出力する．例えば温度計であれば，認識した値に応じて体温計内の液体の体積を変えることで温度を示している．人間は，体温計内の液体とメモリの位置を見ることで，温度を一目でわかりやすく認識することができる．ロボットがセンサを扱う場合にも，センサが認識した値をロボットが利用しやすい形に変換する必要がある．ロボットは基本的に連続値を扱うことが難しいので，センサから入力される情報は離散値であることが望ましい．そのため本論文では，認識した状態を離散値で出力するセンサを考える．このとき，センサにノイズや誤認識などが生じていなければ，環境の状態に対してセンサが出力する値は一意に決まる．ノイズや誤認識などが生じる条件はセンサ個々の特性によって変化するため，本論文では無視して考える．

次に，学習を行うロボットにおけるセンサの役割について述べる．ロボットの学習は，センサから得られた入力に対して適切な出力を学習するという形で行われる．すなわち学習を行うロボットにとってセンサとは，ロボットの学習対象を認識するものである．ここで，ロボットにとってのセンサの役割を Fig.2.1 に示す．ロボットは単数，もしくは複数のセンサを用いて学習対象の状態を認識する．ロボットの持つ各センサが環境の状態を認識すると，認識した状態に応じてロボットに対して何らかの値を出力する．ロボットは各センサから入力された値の組み合わせを用いて状態を認識する．このことから，学習を行うロボットにおけるセンサの役割とは，ロボットが環境の状態を認識するための値をロボットに対して出力するものであるといえる．

次に，学習を行うロボットにおけるセンサの能力について考える．センサの能力を決定する要素は，測定範囲や分解能，測定の精度など様々なものが考えられる．その中でも本論文では，センサの能力を決める最も単純な要素として，センサが出力する値の数を考える．

ロボットにセンサを搭載するとき，そのロボットに認識させたい状態に応じてセンサを選択する．一つのセンサが認識できる物理量は，センサの測定対象や，センサの測定範囲や分解能などの能力に応じて決まっている．そのため，必要に応じて複数のセンサを利用することになる．このとき，各センサが出力する値の数，すなわちセンサが認識できる状態の数はセンサによって異なる．そして，認識できる状態の数が多きセンサ

ほど、ロボットの状態空間の大きさに及ぼす影響も大きい。たとえば Fig.2.1 であれば、センサ v_1 は 4 種類の状態をロボットに対して出力することができるが、センサ v_3 は 2 種類の状態しか出力できない。センサ v_1 が失われるとロボットの状態数は 4 分の 1 にまで減少するが、センサ v_3 が失われてもロボットの状態数は 2 分の 1 にしか減少しない。ロボットの状態空間の変化はロボットの学習の結果に直結しているため、センサが出力する値の数もロボットの学習に直結する、センサの重要な能力であるといえる。

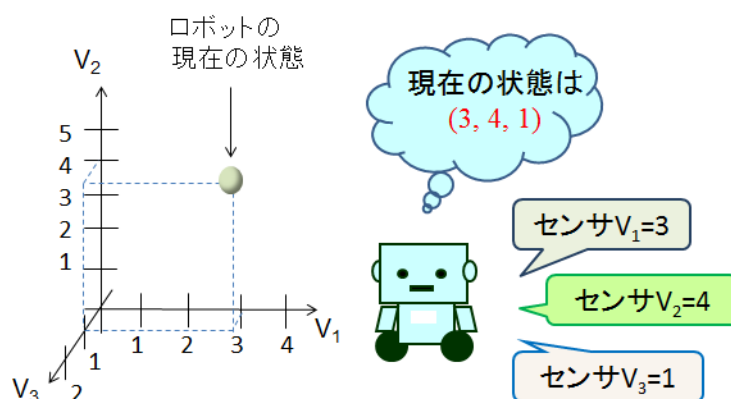


Fig.2.1 学習を行うロボットにおけるセンサの役割

2.3.センサのモデル化

2.2 で述べたセンサの役割を考慮すると、モデル化すべきセンサの要素は次のようになる。

- ・環境の物理量を認識する
- ・認識した物理量を離散値に変換して出力する
- ・センサの出力する値の数が、そのセンサの能力となる

これらの要素を踏まえた上で、センサを次のようにモデル化する。 n 個のセンサを持つエージェントについて、 i 個目のセンサが実際に出力する離散値 v_i を式 (2.1) でモデル化する。ここで V_{\max} はセンサが出力する値の最大値であり、 $V_{\max} > 1$ である。センサは環境の状態を認識すると、認識した値に応じて式 (2.1) に示された値のいずれか一つをロボットに対して出力する。このとき、センサが出力する値は環境の物理量をそのまま出力するのではなく、ロボットが扱いやすい形で出力する。本論文では物理量の変換後の値として、最も単純な 1 から V_{\max} の 1 刻みの離散値を用いる。そのため (2.1) の定義においては、 V_{\max} の値がそのままセンサの出力する値の種類数となる。以下、本論文ではセ

ンサが出力する値の種類数を，そのセンサの出力数と呼ぶ．

$$v_i = \{1, 2, \dots, V_{\max}\} \quad (1)$$

モデル化したセンサを利用した場合にロボットが認識する状態 s を式 (2.2) に示す．ロボットはセンサを一つ以上用いることで環境の状態を認識する．まず初めに，ロボットの持つ各センサが環境の状態を認識する．次に，各センサはロボットに対して，認識した状態に応じた離散値をロボットに対して出力する．最後にロボットは，センサから入力された値の組み合わせに応じて環境の状態を決定する．

$$s = (v_1, v_2, \dots, v_n) \quad (2.2)$$

3. センサが学習効率に及ぼす影響の検証方法

3.1. 調査の対象

本章では 2 章で定義したセンサを用いて、センサの能力の違いに対する学習効率の変化の検証方法を述べる。そのためには調査の対象を明確にする必要がある。まず初めに、本論文の調査に用いる学習手法として強化学習の概要を述べる。次に、強化学習を行うロボットにとって重要となるセンサの能力について述べる。本論文では特にセンサの出力数に注目し、強化学習におけるセンサの出力数の重要性と調査の必要性について述べる。また先行研究において、センサがタスクに対して必要なセンサである場合と不要なセンサである場合では、同じセンサでも学習に及ぼす影響が異なることが示されている。そこで最後に、タスクに対して必要なセンサ、不要なセンサと学習との関係について述べ、センサを必要センサと不要センサに分類して調査を行うことの必要性を示す。

3.1.1. 強化学習概要

機械学習には様々な種類があるが、その中でも最も多くロボットに使用されている手法のひとつが強化学習である。強化学習の特徴として、教師信号を必要とせず、代わりに報酬と呼ばれるスカラ値を利用して学習を行っている点が挙げられる。報酬とは、ロボットの目的を表すものである。教師信号を必要とする学習手法は基本的に、目的を達成するための手順をロボットに教える必要がある。しかし強化学習では目的を達成する手段を与えずとも、ロボットに対して報酬という形で目的を与えるだけで、目的を達成する手段を自律的に獲得できる。この、目的達成までの手段を自律的に獲得できるという点が強化学習の最大の強みである。ロボットの場合、目的が同じであっても使用される環境が異なる場合が多い。例えば掃除ロボットは「床を掃除する」という一つの目的を持っている。しかし使用される家庭によって掃除を行う環境は異なる。たとえばテーブルやタンス、棚などは家庭によって配置されているものの種類や位置、数等が異なるため、それぞれの家庭に適した掃除の経路（手段）を学習する必要がある。このような、環境によって異なる「手段」を与える必要がなく、「目的」のみを与えることで学習を行える強化学習は、ロボットに適した学習手法として重宝されているのである。

強化学習は大きく分けて次の四つの要素から構成されている。

- エージェント（ロボット）

- ・環境
- ・報酬関数
- ・価値関数

エージェントとは、学習を行う主体である。本論文ではロボットがこれに該当する。環境とは、エージェントが学習を行う対象となる状態の集合である。ロボットの場合、センサを通して認識する状態の集合が環境となる。報酬関数とは、環境に対して報酬の割り当て方を決める関数である。報酬関数によってロボットの目的、すなわちロボットが行うタスクが決定する。価値関数とは、ロボットが行った行動に対して、その行動の良し悪しを自律的に評価するための関数である。ロボットの行動は、この価値関数による評価に基づいて決定される。

強化学習の概念図を Fig.3.1 に示す。強化学習では、ロボットがある状態に対して取った行動の組、すなわち状態行動対に対して学習を行う。まず初めにロボットは現在の時刻 t において、自身の置かれている状態 s_t を認識する。次に、自身の状態に対する価値関数を参照し、価値関数を元にするべき行動 a_t を選択し実行する。そしてロボットの行動の結果に応じて次の状態 s_{t+1} に遷移し、報酬関数により定められた報酬 r_{t+1} がロボットに与えられる。ここでロボットは得られた報酬 r_{t+1} を元に、状態 s_t と行動 a_t の組（状態行動対）を評価（学習）する。このように状態認識、行動選択、報酬獲得、状態行動対の評価というサイクルを繰り返して学習を行う。

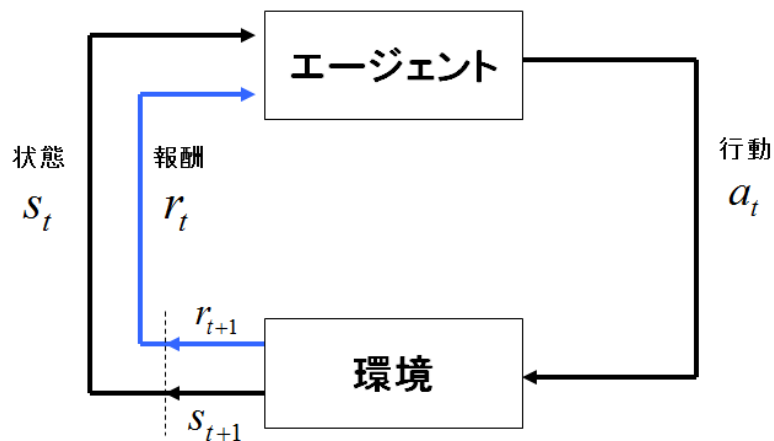


Fig.3.1 強化学習の概念図

3.1.2. センサの出力数と強化学習の関係

本論文で調査対象とするセンサの能力として、センサの出力数と強化学習の関係につ

いて述べる。センサと強化学習の関係を考察するために、まず初めに強化学習の学習対象となる状態空間について考える。

強化学習は状態行動対に対して適切な行動を学習していく学習手法である。そのため強化学習においてはロボットが認識する状態の集合、すなわち状態空間がロボットの学習の対象となる。ロボットの認識する状態空間が変化すれば学習の対象となる状態空間も変化し、学習の内容も変化する。ロボットの認識可能な状態が増加すると状態空間も大きくなる。このときロボットは、増加した状態全てに対して適切な行動を学習していく。そのため、ロボットの状態空間が大きくなるほど様々な状態に対して適応的な行動を獲得することができる。例として、状態空間の大きさが異なる二体のロボットが学習可能な行動を Fig.3.2 に示す。この二体はどちらも「障害物を回避しながら目的地に向かう」というタスクが与えられている。ロボット A は自身が持つセンサによって「障害物がある」「障害物がない」という二つの状態を認識できる。一方ロボット B が持つセンサは「障害物がない」に加え「障害物があり、近づいている」「障害物があるが、遠ざかっている」という、合計三つの状態を認識できる。ロボット A は障害物の有無しか認識できないため「障害物がある場合」には、障害物の状態にかかわらず回避するという行動しか学習できない。一方ロボット B は「障害物がある」場合でも、障害物が「近づいている」場合には回避行動を取り、「遠ざかっている場合」には回避しないというように、それぞれに対して適応的な行動を学習することが出来る。このように、ロボットの状態空間が大きくなるほど、ロボットの学習により適応できる場面が増える。しかし、強化学習では一般的に一回の行動につき一つの状態行動対しか学習することができない。そのため、ロボットの状態空間が大きくなると、その分だけ学習にかかる時間も増加する恐れがある。このように強化学習においては、ロボットの認識する状態空間の大きさが学習に深く関わっている。

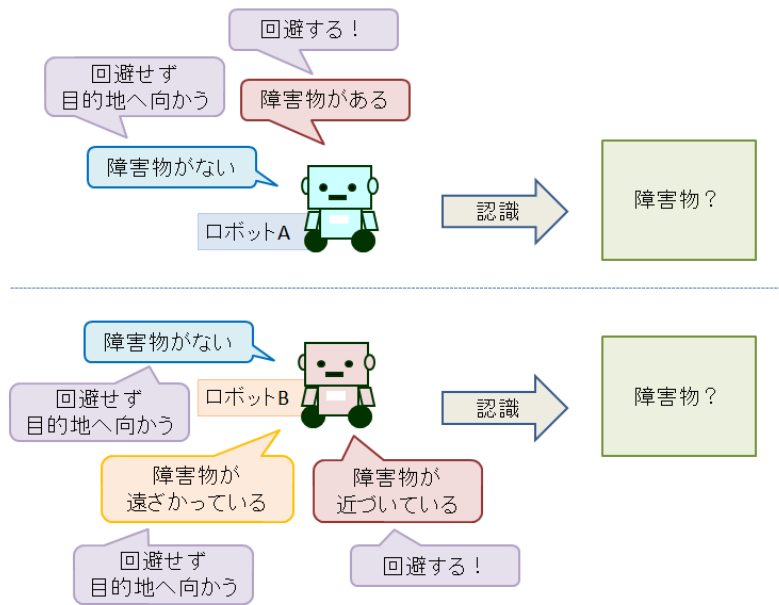


Fig.3.2 状態空間の大きさの異なる二体のロボットが学習可能な行動

ロボットが認識する状態空間は、ロボットが所持するセンサによって決まる。そのためロボットが所持するセンサが異なれば、ロボットが認識する状態の数も変化する。ロボットはセンサを用いて状態を認識するとき、自身の所持する各センサから入力された値の組み合わせによって状態を認識している。そのためロボットが認識する状態空間の大きさは、ロボットが持つセンサの数と、ロボットが持つ各センサの出力数によって決まる。ロボットが所持するセンサの数が多いほど、ロボットは多くの状態を認識できる。また、ロボットが持つ各センサの出力数が多いほど、ロボットは多くの状態を認識できる。すなわち強化学習を行うロボットにとっては、ロボットが持つセンサの数と各センサの出力数が非常に重要であるといえる。

ロボットが持つセンサ数の違いが学習に及ぼす影響については先行研究[]にて既に示されている。この研究ではロボットが持つセンサについて、タスクに対して必要なセンサと不要なセンサに分類している。その上で、タスクに対して必要なセンサが不足した場合とタスクに対して不要なセンサを過剰に所持した場合にセンサが学習に及ぼす影響をそれぞれ調査している。この調査は強化学習を対象としているため、ロボットの獲得報酬量の違いからセンサが学習に及ぼす影響を示している。そして調査の結果として、ロボットの持つセンサがタスクに対して不足すると獲得報酬量が低い状態で学習が収束してしまうことが示されている。また、ロボットがタスクに対して過剰にセンサを所持すると、獲得報酬の増加速度が遅くなることが示されている。そしてこれらの結果から、

ロボットに搭載するセンサの数を増減させることで学習速度と学習収束後の獲得報酬とのトレードオフを調整し、より効率的に学習を行えることを示している。具体的には、ロボットに搭載するセンサを増やすと過剰なセンサが生じやすく、学習の速度が低下しやすいとしている。逆に、ロボットに搭載するセンサを減らすとセンサの不足が生じやすく、学習収束後の獲得報酬量が減少しやすいとしている。その上で、ロボットが行うタスクの時間的制約に応じてセンサ数によるトレードオフを調整することが、ロボットの学習において有効であるとしている。

ただし実際にトレードオフの関係を調整するには、学習収束にかかる時間の変化量や学習収束時点での獲得報酬量の違いなど、センサが学習に及ぼす影響の大きさを考慮する必要がある。しかし先行研究では調査対象のセンサの能力をすべて均一に設定して調査を行っている。そのためセンサ一つ一つが学習に及ぼす影響の大きさは示されておらず、具体的にどのセンサを増やす、あるいは減らすことが有効であるのかを判断することができない。

そこで調査対象として出てくるのが、センサの出力数の違いによる、センサが学習に及ぼす影響の変化である。前述したように、センサの出力数はロボットの学習と深いかわりがある。センサの出力数が異なれば、ロボットが認識する状態の大きさも変化する。出力数が多いセンサほどロボットが認識する状態を大きく変化させ、出力数が少ないセンサほどロボットが認識する空間の変化量は小さいと考えられる。強化学習ではロボットが認識する状態数が学習の結果に大きく反映されるため、センサの出力数が異なれば学習に及ぼす影響の大きさも異なると考えられる。センサの出力数の違いによる学習に及ぼす影響の大きさの違いを示すことができれば、センサ一つ一つが学習に及ぼす影響の大きさがわかる。これにより、センサ数によるトレードオフの調整を行う際に出力数の大きさを参考にすることで、どのセンサを増やす、あるいは減らすことが有効であるのかを判断することができる。

そこで本論文では、センサの出力数の違いに対する学習効率の変化を調査する。これにより、ロボットに搭載するセンサを選択する際の指針として、出力数の大きさから学習への影響を予測できるようにする。

3.1.3. センサの過不足と学習の関係

3.1.2 節で述べたように、先行研究ではロボットが持つセンサについて、タスクに対して必要なセンサと不要なセンサに分類している。その上で、タスクに対して必要なセン

サが不足した場合とタスクに対して不要なセンサを過剰に所持した場合にセンサが学習に及ぼす影響をそれぞれ調査している。その結果として、センサがタスクに対して必要な場合と不要な場合では、学習に及ぼす影響が全く異なることが示されている。そのため本論文においても、センサがタスクに対して必要な場合と不要な場合それぞれについて、センサの出力数の違いによる学習への影響の変化を調査する必要がある。そこで本節ではタスクに対するセンサの過不足について述べ、強化学習におけるセンサの過不足についての定義を行う。

まず初めに、強化学習においてタスク達成のために認識する必要のある状態とは何かを考える。強化学習は、ロボットの状態行動対に対して学習を行う手法である。強化学習ではロボットの目的を示すものとして、タスクに対して報酬が与えられる。そして、ロボットはより多くの報酬が得られる方法を獲得することで学習を行う。この時ロボットは式 (3.1) に示すように、時刻 t におけるロボットの状態 s_t に対して行動 a_t を行った場合に次の時刻 $t+1$ で得られる報酬量 r_{t+1} の期待値を学習していく。このときロボットは、状態行動対 (s_t, a_t) に対して得られる報酬 r_{t+1} が予測可能である必要がある。状態行動対に対して得られる報酬が全く予測できない場合には、ロボットは学習を行うことができない。そのため強化学習においては、ロボットが獲得報酬を予測するために必要な状態を認識することが重要となる。

$$r_{t+1} = (s_t, a_t) \quad (3.1)$$

ロボットが認識する状態は、ロボットが持つセンサから出力された値の組み合わせによって決定する。そのため式 (3.1) は、ロボットが持つ各センサから出力された値と行動との組み合わせとして式 (3.2) のように書き換えることが出来る。ここで v_{it} は、時刻 t においてロボットが持つ i 番目のセンサが出力した値である。

$$r_{t+1} = (v_{1t}, v_{2t}, \dots, v_{nt}, a_t) \quad (3.2)$$

このことから、ロボットが獲得報酬を予測するために必要な状態を認識できるかどうかは、ロボットが持つセンサによって決まるといえる。そのため、任意のセンサ v_i がロボットの獲得報酬を予測可能にするために必要か不要かによって、センサがタスクに対して必要であるか不要であるかを定義する。 v_i から出力された値がロボットの獲得報酬を予測可能にするために必要である場合、 v_i は必要センサであるとする。一方、 v_i から出力された値がロボットの獲得報酬と関係なく、獲得報酬の予測に役に立たない場合、 v_i は不要センサであるとする。以下本論文では、必要センサの一部を所持していないロボットはセンサが不足していると言う。また、タスクに対して不要なセンサを所持しているロ

ロボットはセンサを過剰に所持していると言う。センサに不足が無く不要センサも持たないロボットは、過不足なくセンサを所持するロボットと呼ぶ。

具体例として、二つのセンサ v_1, v_2 が必要センサ、センサ v_3 が不要センサであるタスクを考える。このタスクにおける報酬とセンサの関係を Fig.3.3 に示す。ただし Fig.3.3 では簡単のため、ロボットの行動による状態遷移と報酬の関係は考えていない。このタスクでは二つのセンサ v_1, v_2 を用いて状態認識することで、各状態に対して得られる報酬が確定する。また、 v_1, v_2 のどちらか一方しか持たない場合でも、確率的ではあるが各状態における獲得報酬を予測することができる。そのため二つのセンサ v_1, v_2 はこのタスクに対して必要センサであると言える。しかしセンサ v_3 を用いた場合には、得られる報酬を全く予測することができない。そのためセンサ v_3 は不要センサであると言える。

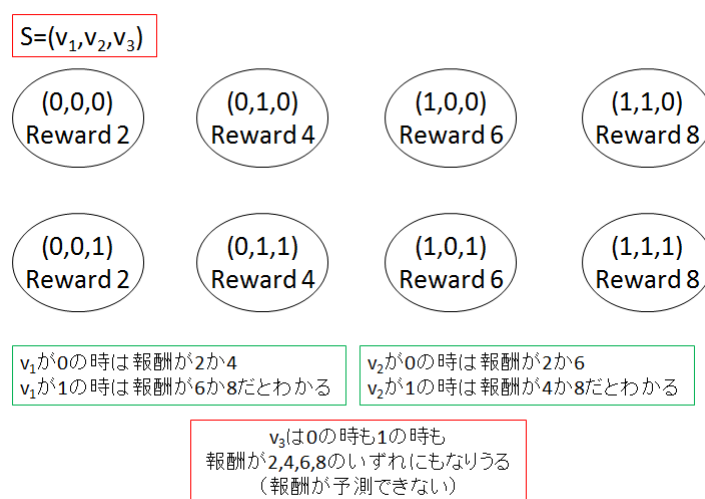


Fig.3.3 報酬と必要センサ・不要センサの関係

本論文ではこれらの定義に基づいてセンサを必要センサと不要センサに分類する。そして、調査対象のセンサが必要センサである場合と不要センサである場合それぞれに対して、センサの出力数の違いによる学習への影響の変化を調査する。

3.2. センサの不足が学習効率に及ぼす影響の検証方法

調査対象のセンサが必要センサである場合に、センサの不足が学習効率に及ぼす影響の検証方法を述べる。この調査では不要センサを扱わないため、以下では必要センサという言葉を省略し、単にセンサと呼ぶ。検証方法の概要図を Fig.3.4 に示す。この調査では、任意の出力数 i のセンサがロボットから失われた時に、ロボットの学習効率に及ぼす

影響を調査する。ここでロボットの学習効率とは、ロボットの学習収束後の獲得報酬の大きさと、ロボットの学習の速度の二点を指す。強化学習では獲得報酬量の多いロボットほど、より最適解に近い学習結果を得ていることを示す。そのため、学習収束した時点で得られる獲得報酬の大きさは、ロボットの学習の良し悪しを判断するための重要な指標の一つとなる。また、学習をロボットに適用する場合には、学習を行うことのできる時間が限られている場合も多い。そのためロボットの学習においては学習収束後の獲得報酬量だけでなく、学習にかかる時間も重要となる。以上のことから本論文では、センサがロボットの学習に及ぼす影響として、ロボットの学習速度と学習収束後の獲得報酬量の二点に注目する。ただし、ロボットの学習速度と学習収束後の獲得報酬とでは、どちらがより重要であるかはタスク毎に異なるため、この二つの影響を総合的に評価することは難しい。そのため実際の調査の際にはこれら二つの影響は個別に評価を行う。本論文ではこれら二つの影響をまとめて学習効率と呼ぶ。

出力数 i のセンサが学習効率に及ぼす影響は、二体のロボットの学習効率の差を利用して算出する。ここで算出に用いるロボットは、タスクに対してセンサを過不足なく所持するロボットと、タスクに対して出力数 i のセンサのみを持たないロボットである。この二体に共通の学習手法を適用し、共通のタスクに対して学習を行わせる。このとき二体のロボットの違いは出力数 i のセンサの有無のみである。そのため、この二体の学習速度、および学習収束後の獲得報酬量の差を算出することで、出力数 i のセンサが学習効率に及ぼす影響を示すことができる。これを出力数の異なる複数のセンサに対して行う。その上で、出力数の違いに対するセンサの学習効率への影響を比較する。

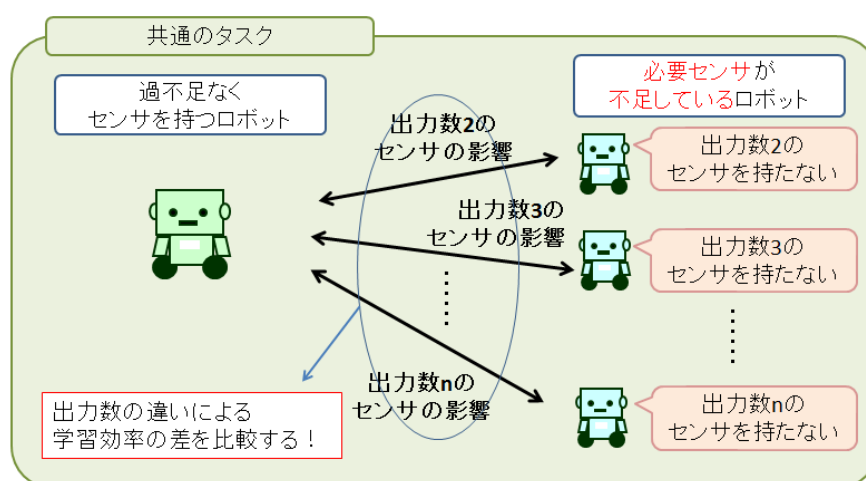


Fig.3.4 出力数の異なるセンサが不足した際に学習に及ぼす影響の検証方法

検証の手順を述べる。まず、ロボットが行うためのタスクを用意する。このとき、調査の対象となる出力数 $2,3,\dots,n$ のセンサがすべて必要センサに含まれるタスクを用意する。次に、このタスクに対して過不足なくセンサを所持するロボットを一体用意する。さらにタスクに必要なセンサのうち、出力数が $2,3,\dots,n$ であるセンサのいずれか一つのみを持たないロボットをそれぞれ一体ずつ用意する。つまり、ロボットの数は全部で n 体となる。これら全てのロボットに対して共通の学習手法として強化学習を適用し、学習を行わせる。そして各ロボットの獲得報酬の遷移から、各ロボットの学習速度、および学習収束後の獲得報酬の大きさをそれぞれ算出する。そして、前述した方法で各出力数のセンサが学習速度、および学習収束後の獲得報酬に及ぼす影響をそれぞれ算出する。最後に、算出した各センサの学習速度、および学習収束後の獲得報酬への影響が、センサの出力数によってどのように変化するかを比較する。これにより出力数の違いに対するセンサの学習効率への影響を示す。

3.3.不要センサが学習に及ぼす影響の検証方法

調査対象のセンサが不要センサである場合に学習効率に及ぼす影響の検証方法を述べる。検証方法の概要図を Fig.3.5 に示す。この調査では、任意の出力数 i の不要センサをロボットに搭載したときに、ロボットの学習効率に及ぼす影響を調査する。ここでロボットの学習効率とは、ロボットの学習収束後の獲得報酬の大きさと、ロボットの学習の速度の二点を指す。ただし先行研究において、不要センサは学習収束後の獲得報酬には影響せず、学習速度に特に大きく影響を及ぼすことが示されている。そのため実際の比較の際には特に学習速度に及ぼす影響に注目する。ただし学習収束後の獲得報酬に関しても、変化が生じないことを確認する。

出力数 i のセンサが学習効率に及ぼす影響は、二体のロボットの学習効率の差を利用して算出する。ここで算出に用いるロボットは、タスクに対してセンサを過不足なく所持するロボットと、全ての必要センサと出力数 i の不要センサ一つを持つロボットである。この二体に共通の学習手法を適用し、共通のタスクに対して学習を行わせる。このとき二体のロボットの違いは出力数 i の不要センサの有無のみである。そのため、この二体の学習速度、および学習収束後の獲得報酬量の差を算出することで、出力数 i の不要センサが学習効率に及ぼす影響を示すことができる。これを出力数の異なる複数の不要センサに対して行う。その上で、出力数の違いに対する不要センサの学習効率への影響を比較する。

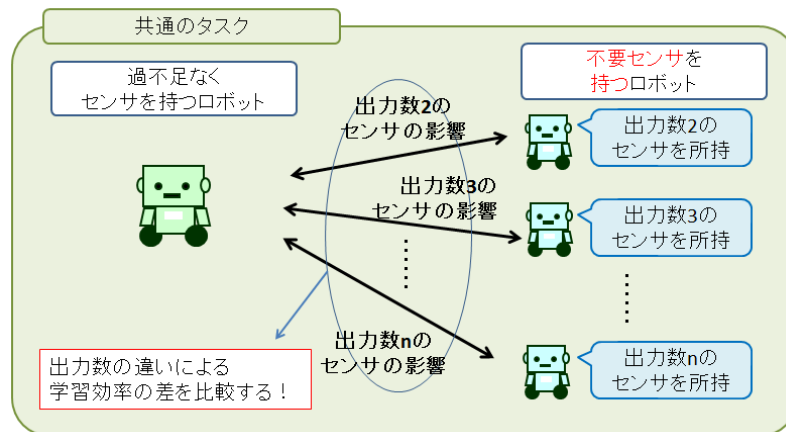


Fig.3.5 B 値認識センサを過剰に所持した際に学習に及ぼす影響の検証方法

検証の手順を述べる。まず、ロボットが行うためのタスクを用意する。このとき用意するタスクに特に制限は無いが、必要センサの実験と同じタスクを用いるのが望ましい。次に、このタスクに対して過不足なくセンサを所持するロボットを一体用意する。さらにタスクに必要なセンサを全て持ち、出力数が2,3,...,nである不要センサのいずれか一つのみを持たないロボットをそれぞれ一体ずつ用意する。つまり、ロボットの数は全部でn体となる。これら全てのロボットに対して共通の学習手法として強化学習を適用し、学習を行わせる。そして各ロボットの獲得報酬の遷移から、各ロボットの学習速度、および学習収束後の獲得報酬の大きさをそれぞれ算出する。そして、前述した方法で各出力数の不要センサが学習速度、および学習収束後の獲得報酬に及ぼす影響をそれぞれ算出する。最後に、算出した各センサの学習速度、および学習収束後の獲得報酬への影響が、センサの出力数によってどのように変化するかを比較する。これにより出力数の違いに対して不要センサが学習効率に及ぼす影響を示す。

4. センサの不足が学習効率に及ぼす影響の実験

4.1. 実験概要

出力数が異なるセンサが学習に及ぼす影響の実験概要について述べる。実験の概要図を Fig.4.1 に示す。まず、ロボットの学習対象となるタスクを用意する。次に、このタスクに対して過不足なくセンサを所持するロボットを一体用意する。さらに、タスクに必要なセンサのうち出力数が $2, 3, \dots, n$ であるセンサのいずれか一つのみを持たないロボットをそれぞれ一体ずつ用意する。つまり、ここまででロボットの数は全部で n 体となる。これらのロボットに対して共通のタスクを与えて学習を行わせる。ただし、一つのタスクのみで実験を行うとランダムな要素によって特定のロボットに対して有利な状況が生まれる可能性がある。そこで、必要センサが同じであるタスクを E 個生成し、 E 個のタスクに対して学習を行わせる。また、学習を行わずランダムに行動するロボットを用意し、他のロボットと同様に同じタスクにて行動させる。このロボットの行動結果は、センサがロボットの学習効率に及ぼす影響を算出する際に利用する。以上の実験を行い、各タスクにおける各ロボットの獲得報酬の平均を用いて、出力数の異なる必要センサが学習効率に及ぼす影響を調査する。

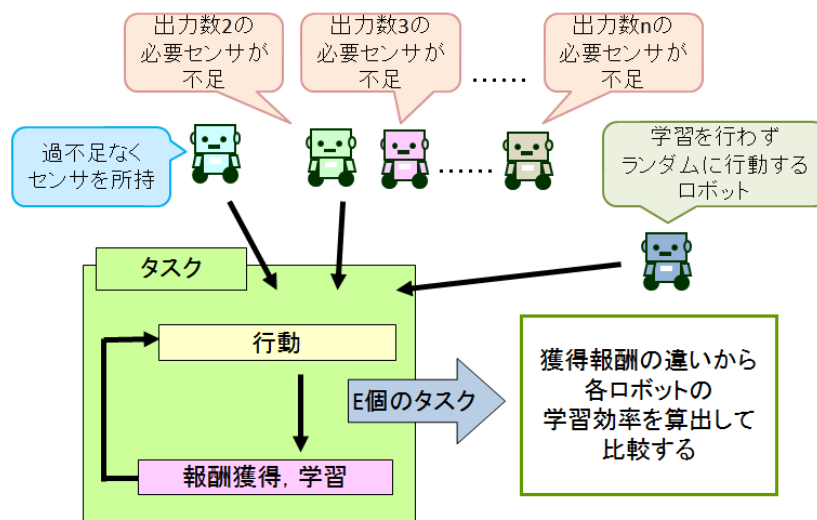


Fig.4.1 必要センサを対象にした実験の概要図

4.2. 実験に用いる学習手法

全てのロボットは共通の学習手法として強化学習を用いる。3.1でも述べたように、強

化学習はタスクに与えられる報酬を元に適切な状態行動対を獲得する学習手法である。強化学習のアルゴリズムはさらに、状態行動対の良し悪しを評価するための行動学習手法と、学習した結果を元に次の行動を選択するための行動選択手法に分けることができる。ロボットはタスクから報酬を受け取ると、その報酬を元に状態行動対の良し悪しを評価する。この評価を行うのが行動学習手法である。また、ロボットは現在置かれている状態に対して、それまでの学習結果を元にして、次にとるべき行動を選択する。ロボットが次にとるべき行動を選択するのが行動選択手法である。本論文では行動学習手法として Q 学習、行動選択手法として ϵ -greedy 法を用いる。

Q 学習について説明する。Q 学習ではロボットの全ての状態行動対に対して Q 値と呼ばれるスカラ値が割り当てられる。Q 値とはその状態行動対に対して得られる報酬の期待値を表す値である。Q 学習において、Q 値を更新する式を式 (4.1) に示す。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (4.1)$$

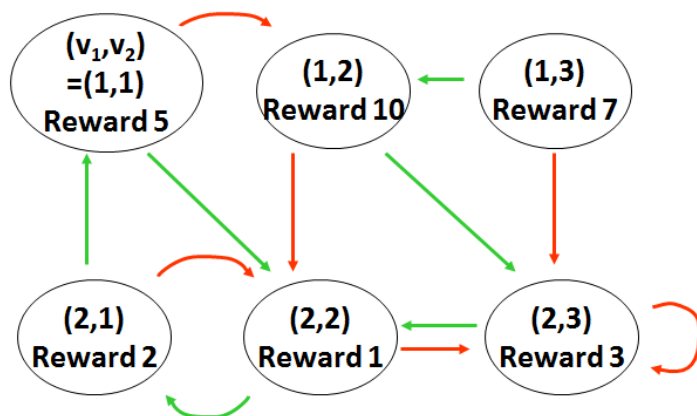
$Q(s_t, a_t)$ 状態行動対 (s_t, a_t) に与えられる Q 値である。 α は学習率と言い $0 < \alpha < 1$ の値を取る。また、 γ は割引率と言い $0 < \gamma < 1$ の値を取る。強化学習では、ある状態で得られる報酬は低くても、その先の状態に遷移すると高い報酬が得られるというような、報酬に遅延が生じるタスクが多く存在する。Q 学習ではある状態に遷移したその時に得られた報酬だけでなく、さらにその先の行動に対して得られると予想される報酬量まで考慮して Q 値が決定される。そのため、報酬に遅延が生じるようなタスクであっても適切に状態行動対の評価を行うことができる。また、遅延報酬に対応した行動学習手法は他にも存在するが、多くの場合は状態行動対の評価に複数ステップを要する。これに対し Q 学習では、状態行動対の評価を 1 ステップ毎に行えるという特徴がある。

ϵ -greedy 法について説明する。 ϵ -greedy 法は、 $1 - \epsilon$ の確率で最も Q 値の高い行動を選択し、 ϵ の確率で Q 値と関係なくランダムに行動を選択する手法である。強化学習における各状態行動対の評価は、ロボットの経験を元に決定される。状態行動対の評価が正しく行われている場合には、評価が最も高い状態行動対を選択していけば良い。しかし多くの場合、状態行動対に対して適切な評価を行うには、その状態行動対を複数回経験する必要がある。そのためロボットの経験数が少ない場合には Q 値による状態行動対の評価が適切でない可能性があり、Q 値に縛られずある程度探索的な行動を行う必要がある。これを ϵ によって実現するのが ϵ -greedy 法である。

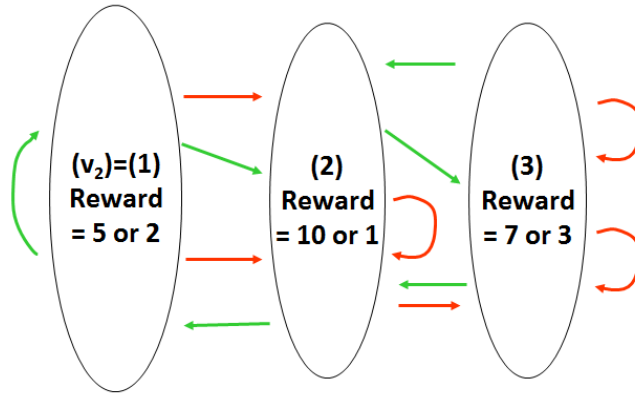
4.3. 実験に用いるタスク

ロボットが行うタスクについて述べる．ロボットが行うタスクはルート探索のタスクである．このタスクはいくつかの状態から構成されており，ロボットは出力数 $2,3,\dots,n$ の必要センサを一つずつ，合計 $n-1$ 個のセンサを用いることですべての状態を過不足なく認識することが出来る．そのため，タスクの状態数は $n!$ 個となる．このとき，ロボットのセンサが不足している場合，ロボットの認識できる状態数が減少することになる．タスクの各状態はそれぞれ一つの報酬と二つの遷移先を持つ．ロボットは各状態において行動を選択することで，選択した行動に対応した遷移先の状態へ確定的に遷移する．そして，遷移先の状態に応じて報酬を獲得する．ロボットの目的は，このタスクにおいて報酬の獲得量が最大になるルートを探ることである．

2 個のセンサ (v_1, v_2) を用いて認識できるタスクの例を Fig. 4.2 に示す．ロボットは各状態において二つの行動を選択することができ，行動 1 を選択した場合は赤い矢印，行動 2 を選択した場合は緑の矢印の遷移先にそれぞれ遷移する．Fig. 4.2 (a) は，ロボットが過不足なくセンサを所持している場合にロボットが認識する状態である．これに対し Fig. 4.2 (b) は，センサ v_1 が不足しているロボットが認識する状態である．センサ v_1 が不足したことによっていくつかの状態に対して状態の切り分けが出来なくなり，認識できる状態の数が減少している．また，それに伴い各状態における遷移先や獲得報酬が確定的ではなく確率的になってしまっている．例えば Fig. 4.2 (b) の状態 (2) は本来二つに切り分けるべき状態である．そのため行動 2 (赤い矢印) を選んだ時に，状態 (2) に遷移するのか (3) に遷移するのか判別することができず，確率的に遷移する状態となっている．



(a) 過不足なくセンサを持つロボットが認識するタスク



(b) センサ v_1 が不足しているロボットの認識するタスク

Fig.4.2 2個の必要センサ (v_1, v_2) を用いて認識できるタスクの具体例

次に、タスクの報酬設定について述べる．今回のタスクは状態の遷移が双方向ではなく一方である．そのため、ある状態 i から再び状態 i へ戻るためにはいくつかの状態を経由する必要がある．そこで、ある状態 i について、再び状態 i へ戻るまでに経路する必要がある最小の状態数が多いほど、高い報酬が得られる状態として報酬を設定した．状態 i に与えられる報酬 Rwd_i を決定する式を式 (4.2) に示す．

$$Rwd_i = e^{\beta \cdot D_{ii}} \quad (4.2)$$

ここで D_{ii} は状態 i から遷移して再び状態 i へ戻るまでに経路する必要がある最小の状態数である． D_{ii} が大きい状態ほど高い報酬が設定される．また、 β は \exp の値が大きくなりすぎないように調整するための D_{ii} の係数である．

実験では、このタスクを生成する際に各状態における遷移先をランダムに設定する．しかしこのタスクは、各状態における遷移先の設定によっては特定のセンサを持つロボットに有利になってしまう可能性がある．そのためこのタスクは、各状態における遷移先をランダムに変更したものを E 個生成する． E 個のタスクにおける学習結果の平均を用いることで、タスクによって特定のロボットにのみ生じる有利、不利な条件を打ち消す．ロボットが行動を選択し、状態遷移して学習を行うまでを 1 ステップとする．各ロボットは同じ状態からタスクを開始し、一つのタスクにつき R ステップの学習を行う．そして、各タスクでの獲得報酬の平均を用いてロボット毎に学習効率を算出する．

4.4.結果の比較方法

結果の比較方法について述べる．結果の比較の手順を以下に示す．

- 1.ロボットの学習効率として，各ロボットの獲得報酬の単純移動平均を算出する
- 2.手順1で算出した各ロボットの単純移動平均を利用して，任意のセンサがロボットの学習速度，および学習収束後の獲得報酬に及ぼす影響を算出する
- 3.出力数 2,3,...,n のセンサについて手順2を行う
- 4.手順2, 3で算出したセンサの影響が，出力数によってどのように変化するのか比較を行う．

以下では，各手順における学習効率の算出方法や比較方法について述べていく．

手順1における各ロボットの学習効率の算出方法について述べる．本論文ではセンサがロボットの学習効率に及ぼす影響として，学習速度および学習収束後の獲得報酬量に注目する．この二つはいずれも，ロボットが獲得した報酬の推移によって示すことができる．学習収束後により高い報酬を獲得しているロボットは，より最適解に近い行動を獲得していると言える．また，少ないステップ数で高い報酬を獲得しているロボットは学習の速度が速いと言える．以上の事から，各ロボットの獲得報酬の推移を用いて学習効率を算出する．今回のタスクはルート探索のタスクなので，ロボットが獲得したルート全体，数ステップを通して得た報酬の量が重要となる．そこで時刻 t において，ロボットが直近の L ステップで得た報酬の平均 (SMA) を算出し，SMA の推移を用いて各ロボットの学習効率の比較を行う． t ステップ目における SMA の式を式 (4.3) に示す． $r_{k,t}$ は， k 個目のタスクにおいてロボットが t ステップ目に得た報酬の値である．

$$SMA(t) = \sum_{k=1}^E \frac{r_{k,t} + r_{k,t-1} + \dots + r_{k,t-L+1}}{L} \quad (4.3)$$

手順2における，任意のセンサが学習効率に及ぼす影響の算出方法について述べる．任意の出力数 i のセンサが学習効率に及ぼす影響とは，出力数 i のセンサを持つロボットと持たないロボットとの SMA の差であると考えられる (Fig.4.3)．そこで出力値 i のセンサが学習に及ぼす影響として，センサを過不足なく所持する (出力値 i のセンサを持つ) ロボットと出力値 i のセンサを持たないロボットの SMA (獲得報酬) の差を算出する．このとき SMA の差は絶対量ではなく，センサを過不足なく所持するロボットが学習により獲得した SMA を 1 としたときの変化割合によって算出する．

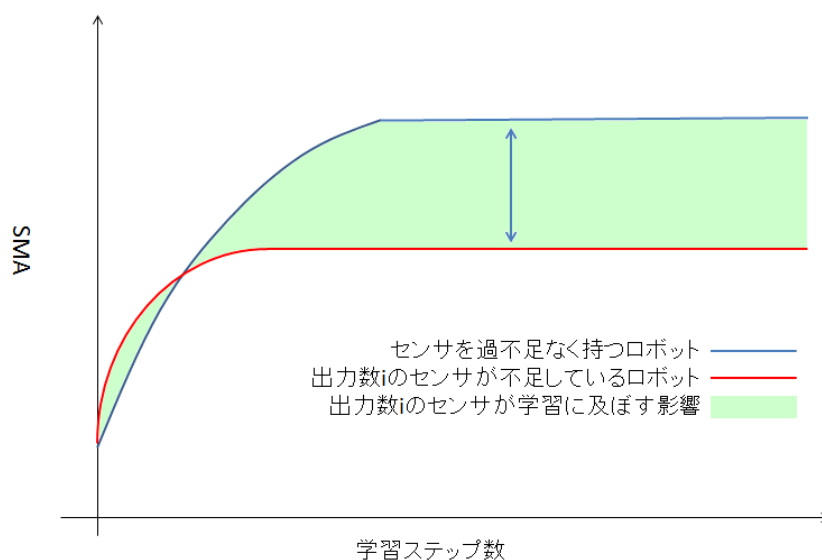


Fig.4.3 出力値 *i* のセンサが学習に及ぼす影響

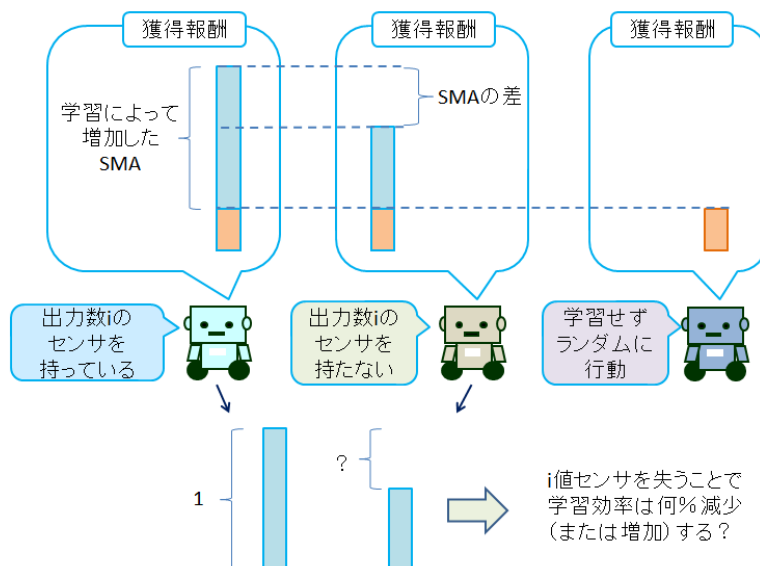


Fig.4.4 出力数 *i* のセンサによる学習効果の変化の割合の比較方法

ロボットが学習により獲得した SMA について述べる. 本実験で用いるタスクは全ての状態に報酬が割り当てられているため, ランダムに行動していてもある程度の報酬を得ることができる. そのため, あるロボットが獲得した報酬量のうち学習によって増加した報酬量は, そのロボットの SMA からランダムに行動するロボットの SMA を引いたものであると考えることができる (Fig.4.4). t ステップ目において出力数 *i* のセンサが学習効率に及ぼす影響 (RSMA) を式 (4.4) によって算出する.

$$RSMA_i(t) = \frac{SMA_1(t) - SMA_i(t)}{SMA_1(t) - SMA_{random}(t)} \quad (4.4)$$

SMA_i はセンサを過不足なく持つロボットの SMA である。 SMA_i は出力値 i のセンサを持たないエージェントの SMA である。 SMA_{random} は学習を行わずランダムに行動するエージェントの SMA である。 $RSMA$ が正の値であれば、ロボットから出力数 i のセンサを取り除いたときの獲得報酬は減少することを示す。 逆に $RSMA$ が負の値であれば、獲得報酬が上昇することを示す。 ロボットの学習収束前の $RSMA$ を参照することで、センサがロボットの学習速度に及ぼす影響を見ることができる。 また、ロボットの学習収束後の $RSMA$ を参照することで、ロボットの学習収束後の獲得報酬に及ぼす影響を見ることができる。

4.5. 実験パラメータ

実験パラメータを Table 4.1 に示す。今回は出力数が 2,3,4,5 のセンサを調査対象とした。そのため、タスクは出力数 2,3,4,5 のセンサを一つずつ用いることで過不足なく認識可能なタスクを用いた。これにより、タスクの状態数は 120 となる。学習ステップ数については、全てのロボットの獲得報酬量がおおむね安定するステップ数として 10 万とした。また、今回のタスクは比較的探索的な行動を多く行わなければ最適解を見つけるのが難しいタスクであるので、 Q の初期値はある程度大きめの 500 とした。

Table 4.1 必要センサに対する実験の実験パラメータ

n (調査対象のセンサの出力数の最大値)	5	β	0.65
タスクの状態数	120	α	0.7
R(学習ステップ数)	100000	γ	0.6
E(実験を行うタスクの数)	300	ϵ	0.1
Qの初期値	500	L	500

4.6. 実験結果

出力数の異なるセンサが不足した場合に学習効率に及ぼす影響の実験結果について述べる。実験結果を Fig.4.5 に示す。Fig.4.5 は各ロボットの SMA の推移のグラフである。Fig.4.5 は 10 万ステップまでに全てのロボットの学習が収束していることを確認するために表示している。全てのロボットについて、学習終了前数万ステップに渡って同程度の SMA を獲得している。そのため全てのロボットが学習収束しているといえる。

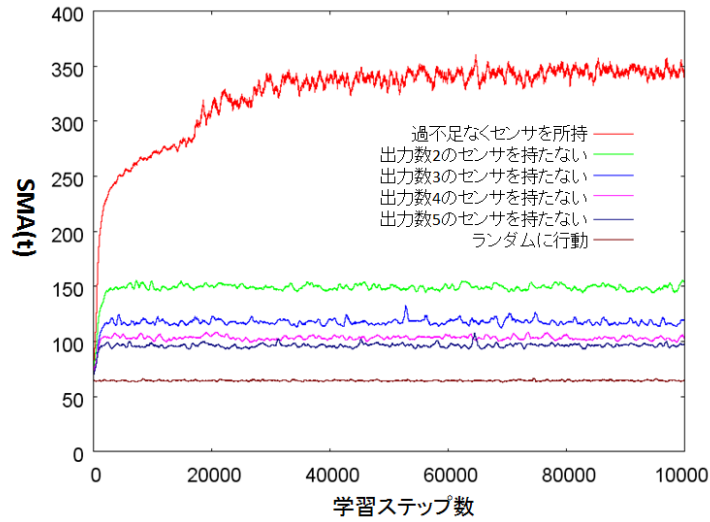


Fig. 4.5 各ロボットの SMA の推移

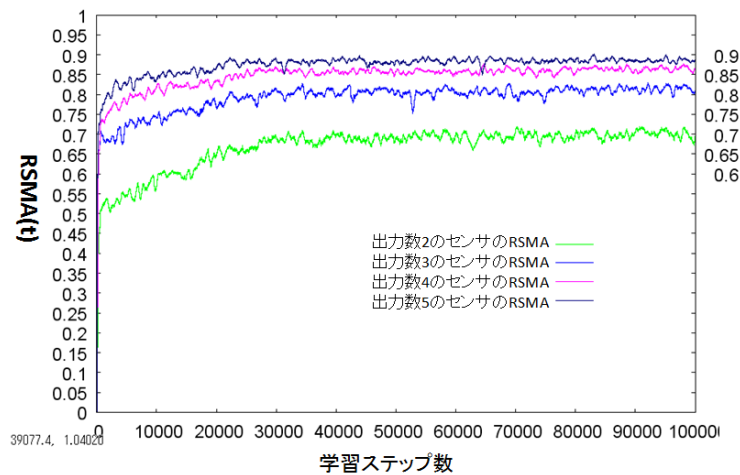


Fig. 4.6 センサの出力数の違いによる RSMA の推移の違い

センサが学習に及ぼす影響を示すグラフとして、Fig.4.6 に注目する。Fig.4.6 は Fig.4.5 の結果を元に各センサの RSMA を算出したものである。このグラフは出力数 i のセンサが不足したときの、ロボットの獲得報酬の減少量の推移を表している。RSMA の値が大きいほど獲得報酬の減少量が大きく、センサの学習効率への影響が大きいといえる。例えば RSMA が 0.7 であれば、センサの不足によって獲得報酬量が 70%減少したことを示している。このグラフを見ると、いずれのセンサが不足した場合でも、学習初期から学習終了時まで獲得報酬が減少していることがわかる。また、出力数が大きいセンサほど RSMA が大きい。特に学習終了時の RSMA に注目すると、最も影響の小さい出力数 2 のセンサが不足した場合であっても、獲得報酬量が 70%ほど減少している。ただし RSMA

の増加量には特徴があり，出力数の大きいセンサほど **RSMA** の値の差が小さい．出力数 2 のセンサと出力数 3 のセンサでは **RSMA** の差が 0.1 程度である．これに対して出力数が 4 と 5 のセンサでは **RSMA** の差は 0.03 程度しかなく，**RSMA** の差が非常に小さくなっていることがわかる．このことから，ある程度センサの出力数が増えていくと，学習への影響は次第に増加しなくなっていくと言える．

学習の速度への影響についてもほぼ同様のことがいえる．学習ステップ数が少ないうちは **RSMA** も小さい．しかし，最も出力数の小さい出力数 2 のセンサであっても学習初期段階で **RSMA** が 0.5 を超えている．また，学習初期から後期のどの部分を見ても，出力数が大きくなるにつれて **RSMA** の増加量は小さくなっていく．このことから学習速度についても，センサの出力数が増えていくと学習速度への影響は次第に増加しなくなっていくと言える．

4.7. 実験結果の考察

今回の実験により，ロボットのセンサが不足した場合には，ロボットの獲得報酬が学習開始時から終了時に渡って減少することがわかった．また，獲得報酬の減少量はそのセンサの出力数に応じて決まり，出力数が大きいほど獲得報酬の減少量が大きいことがわかった．ただし，出力数が大きいセンサになると，獲得報酬の減少量はあまり変わらなくなっていくこともわかった．

こうした結果が生じた理由は，ロボットの状態数の減少量に関係していると考えられる．ロボットのセンサが不足すると，ロボットは認識できる状態の数が減少する．センサの不足によりタスクの状態数よりもロボットの認識可能な状態数が少なくなると，タスクの状態のうちいくつかの状態を切り分けることができなくなる．このとき，タスクの状態としては異なる状態であるにも関わらず，ロボットは同じ状態として認識してしまうという事態が生じる．ロボットは自分が認識している状態に対しては適切に学習を行うことができる．しかし，そもそもロボットが認識している状態がタスクの状態空間に対して適切でないため，結果的に適切な学習を行うことができなくなり，獲得報酬が減少してしまう．ロボットが認識可能な状態数が少ないほどタスクの状態空間との差異も大きくなり，獲得報酬の減少量も増えていくと考えられる．

ここで，各センサが失われた時の状態数の減少量を考える．出力数 2,3,4,5 のセンサが不足したときに減少する状態数は以下の通りである．

- ・出力数 2 のセンサが失われたとき：認識できる状態数は 50%減少

- ・出力数 3 のセンサが失われたとき：認識できる状態数は約 67%減少
- ・出力数 4 のセンサが失われたとき：認識できる状態数は約 75%減少
- ・出力数 5 のセンサが失われたとき：認識できる状態数は 80%減少

まず，出力数 2 と 3 のセンサが不足した場合の状態数の減少量に注目する．出力数 2 のセンサが不足した場合には状態数が 50%減少するのに対し，出力数 3 のセンサが不足した場合の状態数の減少量は 67%であり，状態数の減少量の差は約 17%である．同様に，出力数 3 と 4 のセンサでは約 8%，出力数 4 と 5 のセンサではわずか 5%となり，出力数の大きいセンサほど，状態数の減少量の差が小さくなっていく．以上の事から，センサが不足した場合の獲得報酬の減少量について，出力数の多いセンサになると獲得報酬の減少量があまり変わらなくなったのだと考えられる．

最後に，実験結果を元にセンサ搭載の指針について述べる．基本的には，センサに不足が生じた場合にはロボットの学習に対して悪影響しか与えない．そのためセンサに不足が生じないようにロボットに対してセンサを搭載することが最も重要である．しかしロボットにセンサを搭載する際には，必ずしもタスクに適したセンサを搭載できるとは限らない．複数タスクを行うロボットであれば，タスクの切り替えによって必要センサ・不要センサが変化しセンサに不足が生じるという場合がある．また，タスクが複雑である場合には，センサ自体がタスクに対して必要か不要かわからない場合もある．このように搭載するセンサの選択が難しい場合には，そのセンサの出力数に応じて搭載するセンサを選択することが重要となる．基本的には出力数の大きいセンサほど不足したときの獲得報酬の減少量が多い．そのため，出力数の大きいセンサはなるべく不足が生じないように，優先的にセンサを搭載することが重要である．ただし，出力数が大きくなっていくと，獲得報酬の減少量の差はあまり変わらなくなっていく．例えば出力数 10 の必要センサと出力数 11 の必要センサがあり，片方が不足したときの影響を考えたとき，どちらが不足しても獲得報酬への影響はほとんど変わらない．そのため，この場合どちらを優先して搭載するかは，センサの不足以外の要素を参考にするとよい．例えば複数のタスクを行うロボットについて，出力数 10 のセンサが一つのタスクに対して必要なセンサであり，出力数 11 のセンサは二つ以上のタスクに対して必要なセンサであるとする．このような場合には，出力数 10 のセンサと出力数 11 のセンサが不足したときの影響はほぼ変化しないため，より多くのタスクに対して必要となる出力数 11 のセンサを優先して搭載するのが有効である．

5. 過剰なセンサが学習に及ぼす影響の実験

5.1. 実験概要

出力数が異なるセンサが学習に及ぼす影響の実験概要について述べる．実験の概要図を Fig.5.1 に示す．まず，ロボットの学習対象となるタスクを用意する．次に，このタスクに対して過不足なくセンサを所持するロボットを一体用意する．さらに，タスクに必要なセンサを全て持ち，かつ出力数が $2, 3, \dots, n$ である不要センサのいずれか一つのみを持つロボットをそれぞれ一体ずつ用意する．つまり，ロボットの数は全部で n 体となる．これらのロボットに対して共通のタスクを与えて学習を行わせる．ただし，一つのタスクのみで実験を行うとランダムな要素によって特定のロボットに対して有利な状況が生まれる可能性がある．そこで，同じ条件でタスクをランダムに E 個生成し， E 個のタスクに対して学習を行わせる．その上で各タスクにおける各ロボットの獲得報酬の平均を用いて，出力数の異なる必要センサが学習効率に及ぼす影響を調査する．

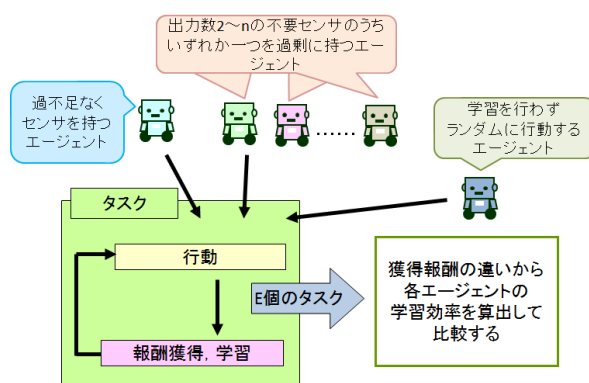


Fig.5.1 不要センサを対象とした実験の概要図

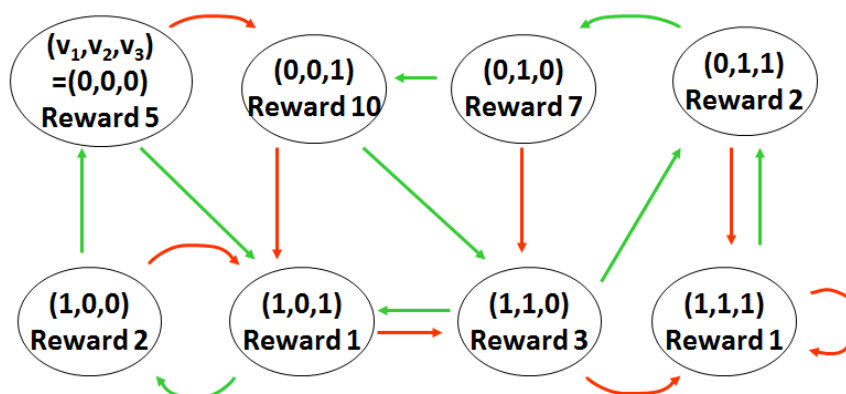
実験の際，全てのロボットは共通の学習手法として強化学習を用いる．このとき必要センサの調査と同様に，強化学習の行動学習手法として Q 学習，行動選択手法として ϵ -greedy 法を用いる．Q 学習や ϵ -greedy 法については 4.1.2 で述べたため省略する．

5.2. 実験に用いるタスク

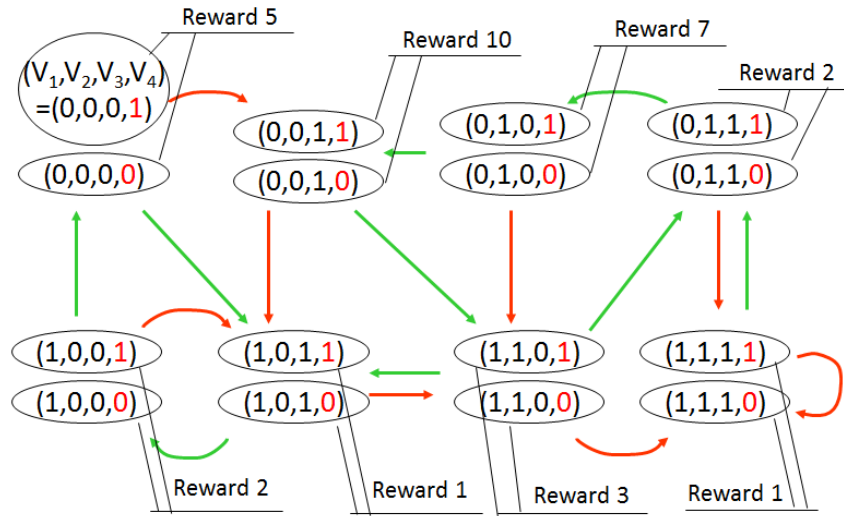
ロボットが行うタスクについて述べる．必要センサでの実験と同様に，ロボットが行うタスクはルート探索のタスクである．このタスクはいくつかの状態から構成されており，ロボットは出力数 $2, 3, \dots, n$ の必要センサを一つずつ，合計 $n-1$ 個のセンサを用いるこ

とですべての状態を過不足なく認識することが出来る．そのため，タスクの状態数は $n!$ 個となる．このとき，ロボットがセンサを過剰に所持している場合，ロボットの認識できる状態数が増加する．タスクの各状態はそれぞれ一つの報酬と二つの遷移先を持つ．ロボットは各状態において行動を選択することで，選択した行動に対応した遷移先の状態へ確定的に遷移する．そして，遷移先の状態に応じて報酬を獲得する．ロボットの目的は，このタスクにおいて報酬の獲得量が最大になるルートの探索である．

3 個の必要センサ (v_1, v_2, v_3) を用いて認識できるタスクの例を Fig.5.2 に示す．ロボットは各状態において二つの行動を選択することができ，行動 1 を選択した場合は赤い矢印，行動 2 を選択した場合は緑の矢印の遷移先にそれぞれ遷移する．Fig. 5.2 (a) は，ロボットが過不足なくセンサを所持している場合にロボットが認識する状態である．これに対し Fig. 5.2 (b) は，必要センサ v_4 を過剰に所持しているロボットが認識する状態である．必要センサ v_4 からの出力値に応じて各状態が切り分けられており，認識できる状態の数が増加している．



(a) 過不足なくセンサを持つロボットが認識するタスク



(b) センサ v_4 を過剰に所持しているロボットが認識するタスク

Fig.5.2 3 個の必要センサ (v_1, v_2, v_3) を用いて認識できるタスクの具体例

次に、タスクの報酬設定について述べる．報酬設定についても必要センサでの実験と同様に、式 (4.2) によって設定する．

$$Rwd_i = e^{\beta \cdot D_{ii}} \quad (4.2)$$

実験ではこのタスクをランダムに E 個生成する． E 個のタスクにおける学習結果の平均を用いることでタスクによって生じる特定のロボットにのみ生じる有利、不利な条件を打ち消す．ロボットが行動を選択し、状態遷移して学習を行うまでを 1 ステップとする．各ロボットは同じ状態からタスクを開始し、一つのタスクにつき R ステップの学習を行う．そして、各タスクでの獲得報酬の平均を用いてロボット毎に学習効率を算出する．

5.3.結果の比較方法

不要センサについては学習の速度にのみ影響することが、先行研究において示されている．そのため、学習効率の比較は特に学習速度に注目して行う．ただし、学習収束後の獲得報酬が変化しないことも確認する．結果の比較の手順を以下に示す．

1. ロボットの学習効率として、各ロボットの獲得報酬の単純移動平均を算出する
2. 各ロボットの学習収束後の獲得報酬を比較し、学習収束後の獲得報酬に変化が無いことを確認する

3.手順 1 で算出した各ロボットの獲得報酬の単純移動平均を利用して、任意のセンサが学習速度に及ぼす影響を算出する

4.出力数 2,3,...,n のセンサについて手順 3 を行う

5.手順 3, 4 で算出した学習速度への影響が、センサの出力数によってどのように変化するか比較を行う。

以下では、各手順における学習効率の算出方法や比較方法について述べていく。

手順 1 における各ロボットの学習効率の算出方法について述べる。必要センサでの実験と同様に、時刻 t において、ロボットが直近の L ステップで得た報酬の平均 (SMA) を算出し、SMA の推移を用いて各ロボットの学習効率の比較を行う。 t ステップ目における SMA の式を式 (5.1) に示す。 $r_{k,t}$ は、 k 個目のタスクにおいてエージェントが t ステップ目に得た報酬の値である。

$$SMA(t) = \sum_{k=1}^E \frac{r_{k,t} + r_{k,t-1} + \dots + r_{k,t-L+1}}{L} \quad (5.1)$$

不要センサの実験では、各ロボットの学習収束後の SMA に注目する。学習収束後の各ロボットの SMA に違いが無ければ、センサの出力数にかかわらず不要センサは学習収束後の獲得報酬に影響を与えないといえる。

学習速度を比較する手法として、今回は時定数を利用する。時定数とは、システムのとる値の開始地点 (最低値) を 0%、システムが最終的に取る値を 100%としたときに、システムのとる値が 63%に達するまでの時間である (Fig5.3)。時定数を用いることでシステムの即応性を示すことができる。時定数を利用した、ロボットの学習速度の比較方法を Fig.5.4 に示す。この実験では、エージェントが獲得した SMA の最低値を 0%、ロボットが獲得した SMA の最大値 (最終的に取る値) を 100%としたときに、ロボットの SMA が 63%に到達するまでのステップ数を時定数とする。各ロボットについて時定数を算出する。その上で、センサを過不足なく持つロボットの時定数を 1 としたときに出力数 i の不要センサを持つロボットの時定数が何%増加したのか、その増加割合を比較する。これにより、センサの出力数の違いによる学習速度への影響の違いを示す。

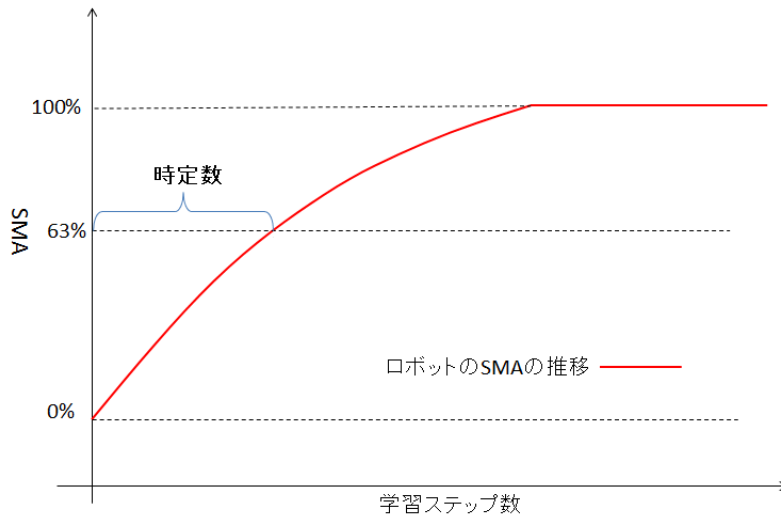


Fig.5.3 時定数の算出方法

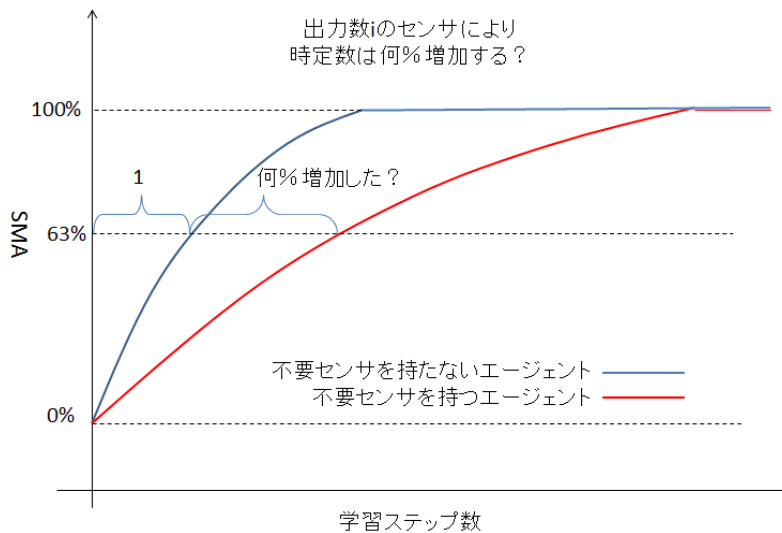


Fig.5.4 時定数を用いた学習速度の比較方法

5.4. 実験パラメータ

実験パラメータを Table 5.1 に示す. 今回は出力数が 2,3,4,5 のセンサを調査対象とした. 使用するタスクには特に制限はないが, 必要センサの実験に合わせて状態数 120 のタスクを用いた. また, 今回は不要センサとして, 出力する可能性のある値をランダムにひとつ出力するものを用いた.

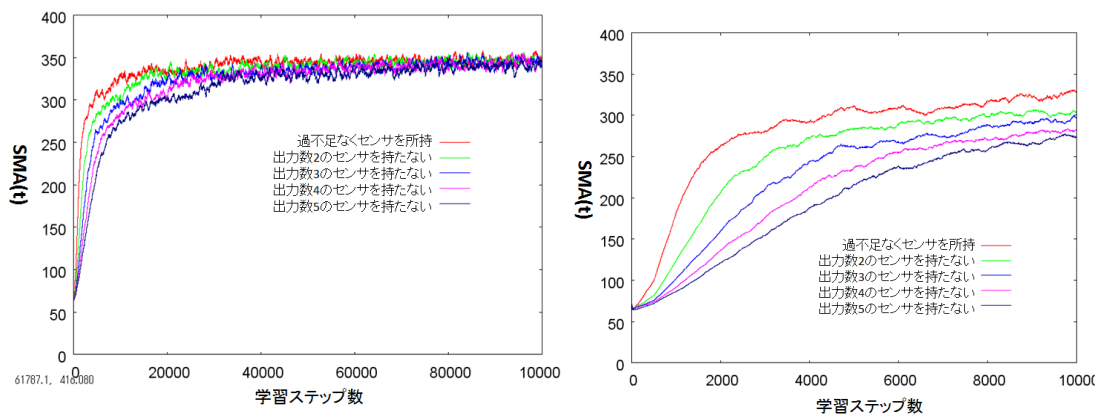
Table 5.1 必要センサに対する実験の実験パラメータ

n (調査対象のセンサの出力数の最大値)	5	β	0.65
------------------------	---	---------	------

タスクの状態数	120	α	0.7
R(学習ステップ数)	100000	γ	0.6
E(実験を行うタスクの数)	300	ε	0.1
Qの初期値	500	L	500

5.5.実験結果

実験結果を Fig.5.5, Fig.5.6, Table 5.2 に示す. Fig.5.5 は各ロボットの SMA の推移のグラフである. Fig.5.5 (a)が 100000 ステップまで, Fig.5.5 (b)が 10000 ステップまでの SMA の推移のグラフである. Fig.5.5 の 80000 ステップ以降の SMA に注目すると, いずれのロボットも SMA が 350 付近で安定して推移していることがわかる. このことから, センサの出力数に関わらず, 不要センサは学習収束後の獲得報酬に影響を及ぼさないことがわかる.



(a) 10 万ステップまでの SMA の推移

(b) 1 万ステップまでの SMA の推移

Fig.5.5 各ロボットの SMA の推移

Table 5.2 各ロボットの時定数

	過不足なし	出力数 2 の 不要センサを 所持	出力数 3	出力数 4	出力数 5
時定数 (ステップ数)	1750	3074	4403	5718	6994
時定数の割合	1.0	1.752566	2.510262	3.259977	3.987457
時定数の増加量	0	0.752566	1.510262	2.259977	2.987457

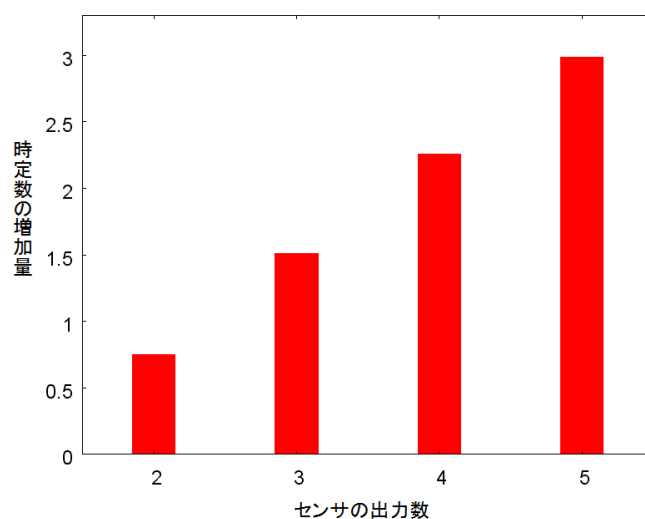


Fig.5.6 出力数の違いによる時定数の増加量

Fig.5.5 の結果を元に各ロボットの時定数を算出し、表にまとめたのが Table 5.2 である。1 行目は各ロボットの時定数を示している。2 行目は、過不足のないロボットの時定数を 1 としたときの各ロボットの時定数の割合を示している。3 行目は、過不足のないロボットの時定数を 1 としたときの、各ロボットの時定数の割合の増加量を示している。また、出力数の違いによる時定数の増加量を棒グラフで示したのが Fig.5.6 である。

Table 5.2 や Fig.5.6 における各ロボットの時定数の増加量に注目する。すると、センサの出力数に応じて時定数が 75% 程度ずつ増加している。増加量は出力数によって 73~76% ほどで多少ばらつきが見られるが、センサの出力数に対して時定数の増加量がほぼ比例の関係であることがわかる。今回は検定などを行っていないため、増加量のばらつきについて誤差の範囲内であること明確に示すことはできない。しかし今回用いているタスクはルート探索である上にルートの長さも不定なので、SMA を取っていてもばらつきが生じやすいタスクである。また、ロボットが用いている行動選択手法の ϵ -greedy 法も、ロボットの学習結果にかかわらず 10% の確率でランダムに移動するため、結果にある程度のばらつきを生む原因となっている。こうした要因を考えると、Table 5.2 に見られる時定数の増加量のばらつきは誤差の範囲内だと考えられる。以上のことから不要センサをロボットに搭載した場合、そのセンサの出力数に比例して学習速度が遅くなるといえる。

5.6.実験結果の考察

今回の実験により、ロボットが不要センサを所持した場合には、ロボットの学習速度が遅くなることがわかった。また、学習にかかる時間は、搭載した不要センサの出力数に比例して増加することもわかった。こうした結果が生じた原因は、ロボットの認識する状態数の増加量と関係していると考えられる。ロボットが不要センサを所持すると、ロボットの認識する状態の数が増加する。ロボットは1ステップにつき一つの状態しか学習を行うことができない。そのため、増加した状態数に応じて学習にかかるステップ数が増加したのだと考えられる。ロボットが出力数*i*の不要センサを所持すると、ロボットの状態数は*i*倍に増える。つまり、不要センサによる状態数の増加量はセンサの出力数に比例する。そのため、ロボットの学習にかかる時間についても出力数に比例して増加したのだと考えられる。

不要センサによって増加した状態は、タスクから見ると同じ状態である。例として、ある状態*A*が出力数3のセンサによって状態*A*'₁、*A*'₂、*A*'₃の三つの状態に切り分けられた場合を考える。これら三つの状態は、タスクにとってはすべて状態*A*と同じ状態である。ロボットは各状態に対してそれぞれ学習を行うが、十分に学習が進めば各状態に対する学習結果は同じになる。強化学習であれば、三つの状態*A*'₁、*A*'₂、*A*'₃の*Q*値は全て等しくなると考えられる。この結果、不要センサを持たないロボットが状態*A*に対して学習を行った結果に対して、不要センサを持つロボットが三つの状態*A*'₁、*A*'₂、*A*'₃に対して学習を行った結果は等しくなる。そのため、ロボットが不要センサを持つ場合でも学習収束後の獲得報酬量には影響が現れなかったのだと考えられる。

最後に、実験結果を元にセンサ搭載の指針について述べる。基本的には、不要センサが生じた場合にはロボットの学習に対して悪影響しか与えない。そのため不要センサが生じないようにロボットに対してセンサを搭載することが最も重要である。しかしロボットにセンサを搭載する際には、必ずしもタスクに適したセンサを搭載できるとは限らない。複数タスクを行うロボットであれば、タスクの切り替えによって必要センサ・不要センサが変化し不要センサが生じるという場合がある。また、タスクが複雑である場合には、センサ自体がタスクに対して必要か不要かわからない場合もある。このように搭載するセンサの選択が難しい場合には、そのセンサの出力数に応じて搭載するセンサを選択することが重要となる。まず、不要センサはロボットの学習速度に影響を及ぼすが、学習収束後の獲得報酬には影響を及ぼさない。そのため、ロボットの学習対象となるタスクに時間的な制約が無い場合であれば、不要センサが生じることを気にする必要は特にない。一方ロボットのタスクに時間的な制約である場合には、センサの出力数に

応じて搭載するセンサを選択することが有効である。不要センサの場合には、センサの出力数に対する学習速度の増加量が単純な比例関係にある。そのため不要なセンサとなりうるセンサが複数ある場合には、出力数の多いセンサほどロボットへの搭載を控えることが有効であると考えられる。また、センサの出力数と学習速度が比例という比較的単純な関係にあるため、ロボットの学習時間の変化から不要センサの発生、および特定を行うことも比較的容易であると考えられる。

6. タスク及び所持センサの違いに対する調査

6.1. 調査対象の拡張

本論文ではここまで、センサの能力の違いに対する学習効率の変化について調査を行ってきた。ロボットの学習においては、タスクの状態空間とロボットの認識する状態空間の大きさが非常に重要である。そのため、ロボットの認識する状態空間の大きさを決める要因としてセンサの出力数の違いに注目し、センサの出力数の違いに対する学習効率の変化について調査を行った。

しかしこれまでの調査はタスク依存、あるいはロボットの所持センサ数に依存した結果である可能性がある。本論文ではロボットの状態数を変化させる要因としてセンサの出力数の違いと学習との関係に注目して調査を行った。しかし出力数の同じセンサであっても、ロボットの所持センサやロボットの行うタスクによっては、学習に及ぼす影響が変化してしまう恐れがある。

ロボットのタスクが異なる場合の状態空間について考える。ロボットのタスク達成に必要なセンサはタスク毎に異なっている。タスクに必要なセンサの数や、各必要センサの出力数が多いタスクであれば、タスクの状態空間は大きくなる。逆にタスクに必要なセンサの数や、各必要センサの出力数が少ないタスクであれば、タスクの状態空間は小さくなる。ロボットから出力数 i のセンサが失われた場合、ロボットの状態数は i 分の 1 になる。逆に出力数 i のセンサを搭載した場合、ロボットの状態数は i 倍になる。そのため出力数が同じセンサであっても、タスクの状態数が異なる場合にはセンサによる状態数の変化量も異なる。

例として状態数 24 のタスク A と状態数 18 のタスク B を考える (Fig.6.1)。タスク A に対して過不足なくセンサを持つロボットから出力数 2 の必要センサを除去すると、ロボットの認識する状態数は 12 個減少する。一方タスク B から出力数 2 の必要センサを除去した場合には、ロボットの認識する状態数は 9 個しか減少しない。不用センサについても同様で、同じ出力数のセンサであっても増加する状態数がタスクによって異なる。このように、ロボットの行うタスクが異なると、出力数が同じセンサであっても状態空間の変化量が異なる。そのため、ロボットの学習に及ぼす影響の大きさも異なる可能性がある。

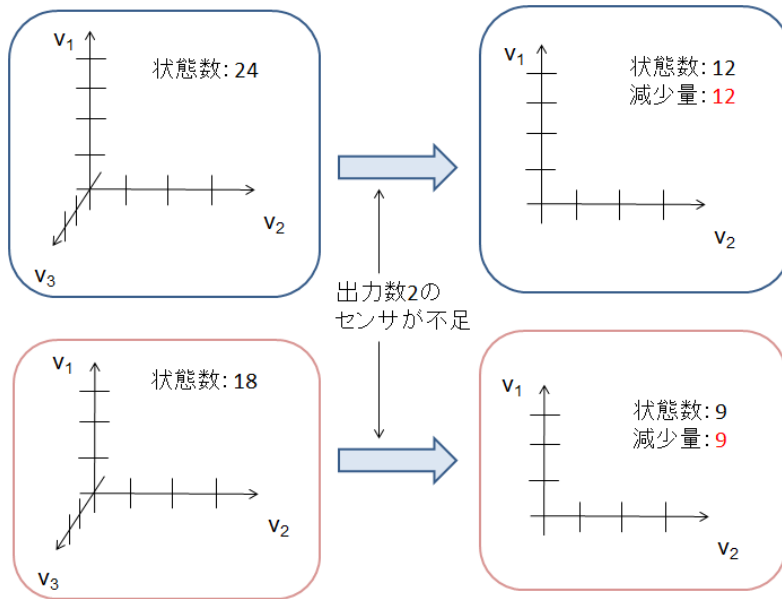


Fig.6.1 タスクの違いに対するセンサによる状態数の変化

ロボットの所持センサに過不足が生じている場合の状態空間について考える。5章までの実験では、過不足なくセンサを所持するセンサから一つだけ必要センサを除去、あるいは不要センサを搭載するという形で調査を行った。しかし実際にロボットを利用する際には、二つ以上のセンサが不足する、あるいは二つ以上のセンサを過剰に所持することも考えられる。さらにセンサが不足している上に不要センサを所持するなど、センサの過不足が複合的に生じる可能性も考えられる。出力数と同じセンサであっても所持センサの過不足が異なる場合には、ロボットの認識する状態数への影響は変化する。

例として状態数 24 のタスクを行うロボットを考える (Fig.6.2)。タスク A に対して過不足なくセンサを持つロボットから出力数 2 の必要センサを除去すると、ロボットの認識する状態数は 12 個減少する。ここでさらに出力数 2 のセンサを除去した場合、ロボットの認識する状態数は 6 個だけ減少する。同じ出力数 2 のセンサであっても、ロボットのセンサの過不足によって状態数の減少量が異なることがわかる。同様に、過不足なくセンサを持つロボットに出力数 2 の不要センサを搭載すると、状態数は 24 増加する。このロボットにさらに出力数 2 の不要センサを搭載すると、状態数は 48 増加する。ロボットのセンサの過不足が複合的に生じている場合も同様に、出力数と同じセンサであってもロボットの認識する状態数は変化する。このように、ロボットの所持センサが異なると、出力数と同じセンサであっても状態空間の変化量が異なる。そのため、ロボットの学習に及ぼす影響の大きさも異なる可能性がある。

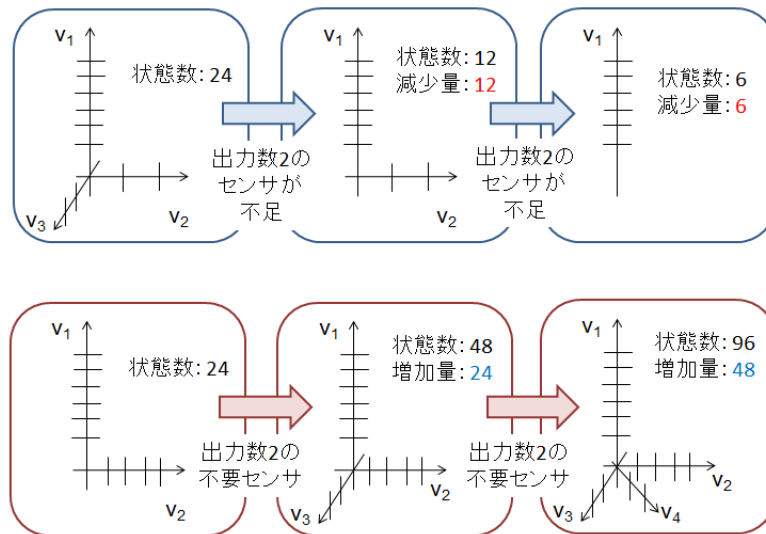


Fig.6.2 ロボットの所持センサが異なる場合のセンサによる状態数の変化

5 章までの実験結果だけでは特定のタスクにおいて特定のセンサを所持するロボットに対しての調査しか行っていない。そのためロボットの所持するセンサが 4 章の実験設定と異なる場合には、センサが学習に及ぼす影響を判断することができない。当然のことながら、一般的に利用されるロボットが行うタスクや所持するセンサはロボットによって異なる。本論文における調査結果を一般的なロボットに適用するためには、ロボットの行うタスクや所持するセンサが異なる場合でも、センサの出力数に対する学習効率の変化を判断できなければならない。そのためには、ロボットのタスクや所持センサが異なる場合に、センサの出力数と学習効率の関係がどのように変化するか調査する必要がある。

そこで本章以降では、ロボットのタスクが異なる場合と、ロボットの所持するセンサが異なる場合のそれぞれについて、5 章までの調査結果を適用できるかどうかの調査を行う。これにより、5 章までの調査結果が適用可能な条件を明確に示す。ただし、ロボットのタスクや所持センサの違いによってセンサの学習効率への影響が全く異なってしまうなど、5 章までの調査結果が全く適用できない場合も考えられる。その場合には今後の研究課題として、センサが学習に及ぼす影響を調査するために必要な事項を洗い出して明確に示す。

6.2. センサの縮退による調査の単純化

ロボットのタスクが異なる場合、およびロボットの所持するセンサが異なる場合のセ

ンサの影響の調査を行うにあたり，調査を単純化することを考える．ロボットが行うタスクの違いを考える際には，タスクに必要なセンサについて考える必要がある．このときタスクに必要なセンサの数や各センサの出力数などを考慮すると，センサの組み合わせは非常に膨大になる．しかしこれまでも何度か述べたように，ロボットの学習，特に強化学習においてはタスクの状態空間の大きさ，すなわち状態数が重要となる．そのためタスクに必要なセンサの組み合わせにかかわらず，タスクの状態数のみから学習への影響を判断できる可能性がある (Fig.6.3)．これが可能であるならば，調査方法を単純化することができる．そこで本節ではタスクの状態数のみから学習への影響を判断する方法として，センサの縮退について述べる．センサの縮退を行うことにより，調査を単純化する．

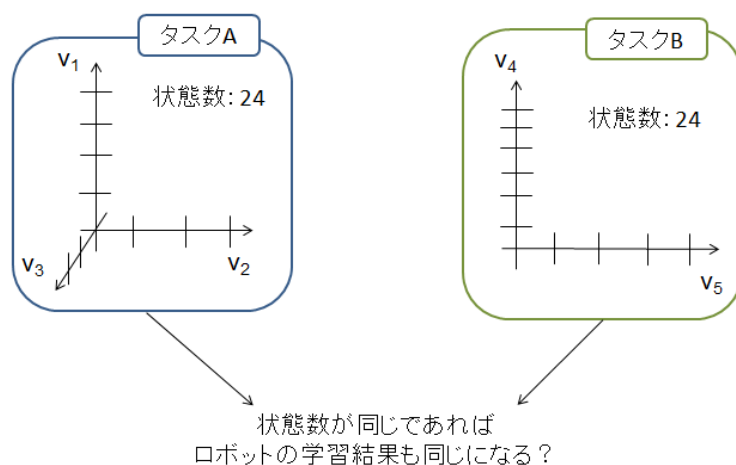


Fig.6.3 タスクと状態数について

本論文では調査の際に強化学習を用いるため，強化学習におけるセンサの縮退について考える．強化学習は，ロボットが持つ各状態行動対の価値を学習することで適切な行動を獲得する学習手法である．ある二体のロボットについて，二体のロボットの認識可能な状態空間と選択可能な行動が同じであれば，二体のロボットの学習内容は基本的に同じになる．このとき二体のロボットの持つセンサの種類や能力等は関係なく，認識可能な状態空間が全く同じであればこの関係は成立する．

例として，所持センサの異なる二体のロボットを考える．二体のロボットの関係を Fig.6.4 に示す．Fig.6.4 において，ロボット A は出力数 2 のセンサ A,B を持っている．ロボット A はセンサ A,B の両方を用いることで，4 つの状態 1,2,3,4 を認識することができる．一方，ロボット B は出力数 4 のセンサ C を一つ持っているのみで，ロボット A とは

所持しているセンサが異なる。しかしロボット B はセンサ C を用いることで、ロボット A と同じく 4 つの状態 1,2,3,4 を認識することができる。二体のロボットは所持するセンサが異なるが、どちらのロボットも全く同じ状態空間を持っている。そのため二体のロボットが行う学習は全く同じとなる。

本論文では、同じ状態を認識するセンサの組を、縮退の関係を持つと呼ぶ。Fig.6.4 の例では、センサ A,B の組とセンサ C のみからなる組がそれぞれ縮退の関係を持つセンサの組であると言える。以下では、あるセンサの組を仮想的なセンサに縮退する方法について述べる。

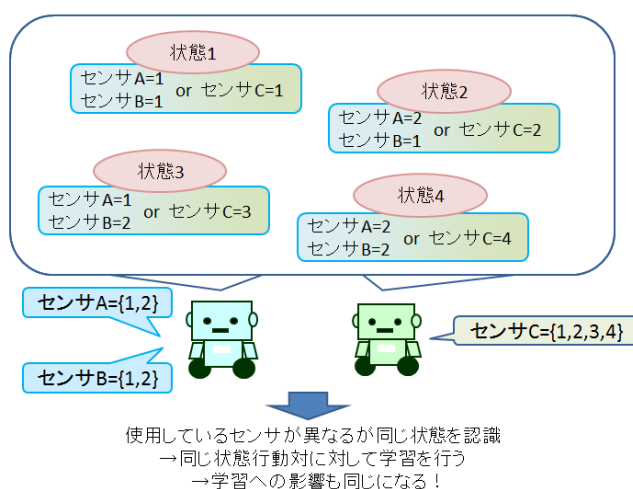


Fig.6.4 センサが縮退の関係を持つロボットの例

まず、 n 個のセンサの組 v_1, v_2, \dots, v_n を持つロボットについて考える。センサの縮退の様子を Fig6.5 に示す。ロボットが認識可能な全状態数を P とすると、ロボットが認識可能な状態それぞれに対して 1 から P までの番号を割り振ることができる。この P 個の状態に対して、各状態の状態番号に対応した値を出力する、出力数 P の仮想的なセンサを定義する。このとき出力数 P の仮想センサは、ロボットが持つ n 個のセンサの組 v_1, v_2, \dots, v_n を縮退したセンサとなる。そのため、出力数 P の仮想センサを持つロボットの学習は、 n 個のセンサの組 v_1, v_2, \dots, v_n を持つロボットの学習と等しくなる。

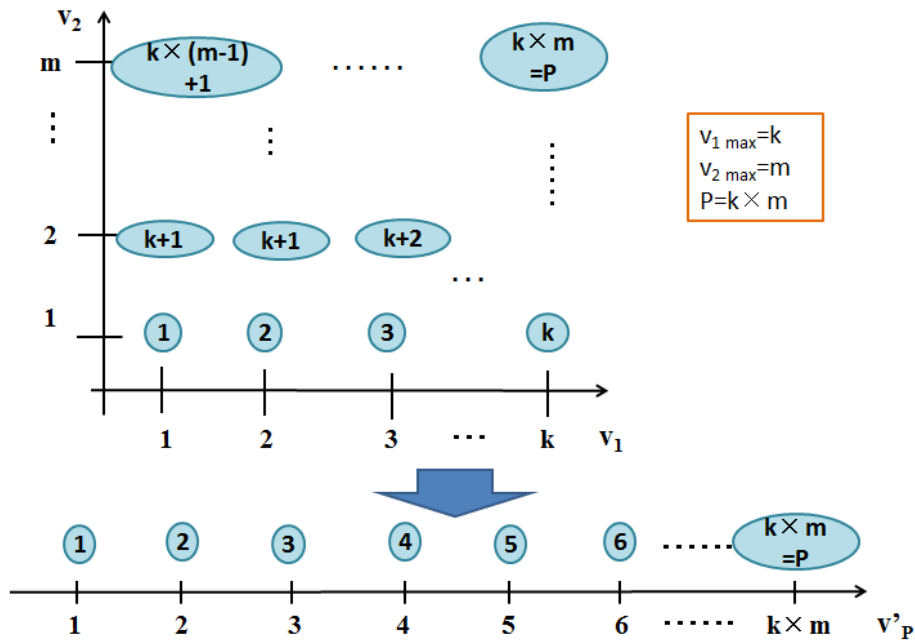


Fig.6.5 センサの縮退の様子

同様にいかなるセンサの組であっても，センサの組によって認識可能な状態数 P がわかれば，出力数 P の仮想的なセンサに縮退することができる．ロボットが認識できる状態数 P は，エージェントが持つ各センサの出力数がわかっているならば容易に算出することができる． n 個のセンサ v_1, v_2, \dots, v_n を持つエージェントについて，各センサの出力数を $V_{\max 1}, V_{\max 2}, \dots, V_{\max n}$ とすると，エージェントが認識可能な状態数 P は次の式 (6.1) を用いて算出できる．

$$P = V_{\max 1} \cdot V_{\max 2} \cdot \dots \cdot V_{\max n} \quad (6.1)$$

以上のことから，状態数が P となるような組み合わせのセンサを必要センサとするタスクは，出力数 P の仮想的なセンサひとつのみを必要センサとするタスクとして置き換えることができる．このようにセンサの縮退を利用することで，タスクに必要なセンサの組み合わせにかかわらず，タスクの状態数のみから学習への影響を判断できる．以上のことから本論文では出力数 P の仮想的なセンサを利用して，出力数の異なるセンサが学習効率に及ぼす影響の調査を行っていく．以下では縮退後の仮想的なセンサを v'_i と定義する．ここで i は，仮想的なセンサを識別するための番号である．

6.3.状態数の異なるタスクにおける検証方法

本節では、ロボットのタスクが異なる場合に、センサの出力数の違いが学習効率に及ぼす影響の検証方法について述べる。この調査ではロボットのタスクの違いとして、タスクの状態数のみに注目する。6.2で示したように、状態数が P となるセンサの組は全て、出力数 P の仮想的なセンサに縮退することができる。これはすなわち、状態数が P のタスクは全て、タスクに必要なセンサを出力数 P の仮想センサに縮退できることを示す。よって、タスクに必要なセンサの組は考えず、タスクの状態数の違いにのみ注目して調査を行う。

この調査は調査対象が必要センサである場合と不要センサである場合に分けて行う。以下では調査対象が必要センサである場合と不要センサである場合それぞれについて検証方法を述べる。

6.3.1.調査対象のセンサが必要センサの場合

調査対象のセンサが必要センサの場合の調査では、センサの出力数の違いが学習効率に及ぼす影響として以下の三点について注目する。

- ・センサが不足すると、学習開始時から終了時まで獲得報酬量が減少する
- ・センサの出力数の増加に伴い、学習収束後の獲得報酬の減少量が増加する
- ・センサの出力数が大きくなると、出力数の増加による獲得報酬の減少量が増えなくなる

これら三点の学習効率への影響が、タスクの違いによって変化するか否かを調査する。また、これらの学習効率への影響が変化する場合には、変化に明確な傾向があるのか、あるいは全く学習効率への影響が異なりタスク毎の比較は困難であるのかを明確にする。

状態数の異なるタスクにおける、センサの出力数の違いが学習効率に及ぼす影響の検証方法について述べる。検証方法の概要図を Fig.6.6 に示す。まず初めに状態数の異なる複数のタスクを用意する。次に、用意した各タスクに対してセンサの出力数の違いが学習効率に及ぼす影響の調査を行う。このとき各タスクにおける調査は、3.2と同じ方法で行う。3.2と同じ検証を異なるタスクで行うことで、任意の出力数のセンサがロボットの学習速度、およびロボットの学習収束後の獲得報酬量に及ぼす影響を調査することができる。また、それらの学習効率への影響が、センサの出力数の変化によってどのように変化するのかを調査することができる。

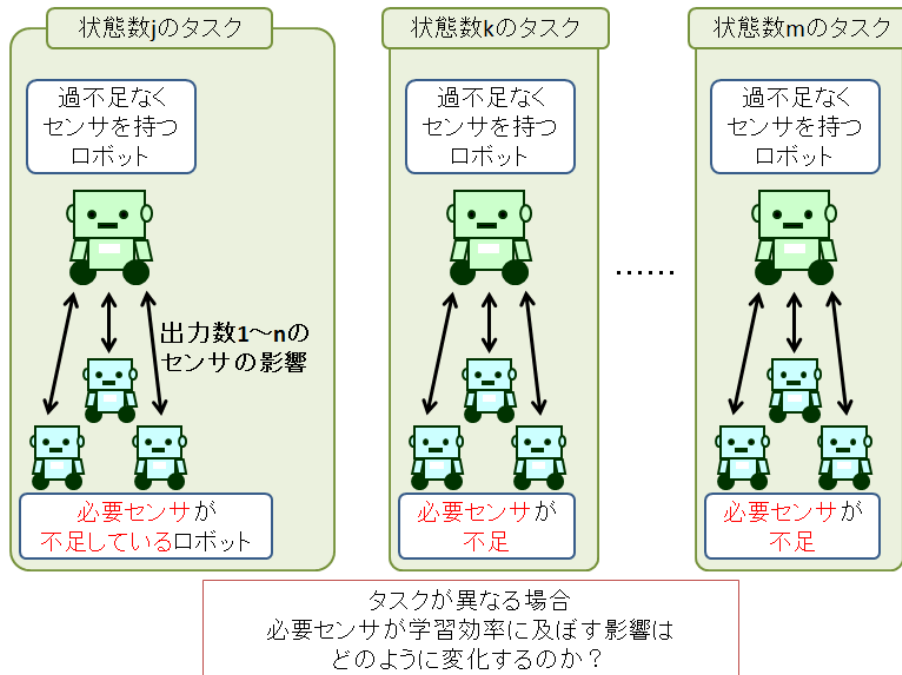


Fig.6.6 必要センサの出力数の違いが学習効率に及ぼす影響の検証方法概要図

各タスクでの実験が終了したら、全てのタスクに対して以下の三点を確認する。

- ・センサが不足すると、学習開始時から終了時まで獲得報酬量が減少する
- ・センサの出力数の増加に伴い、学習収束後の獲得報酬の減少量が増加する
- ・センサの出力数が大きくなると、出力数の増加による獲得報酬の減少量が増えなくなる

なる

タスクが異なる場合でもこれらの影響が変化しなければ、5章までの実験結果はタスクが異なる場合にも同じように適用することができる。一方、タスクの変化によってこれらの影響が変化する場合には、5章までの実験結果はタスクが変わると適用できないことになる。この場合は、タスクの違いに対してこれらの影響の変化に規則性があるかどうかを確認する。規則性やある種の傾向が見られる場合にはそれを示し、タスクの状態数に対してセンサが学習に及ぼす影響の変化を示す。タスクの違いに対してこれらの影響の変化に規則性などが見られない場合には、今後の研究課題としてタスクの違いに対するセンサの学習効率の変化について調査の必要性を示す。

6.3.2. 調査対象のセンサが不要センサの場合

調査対象のセンサが不要センサの場合の調査では、センサの出力数の違いが学習効率に及ぼす影響として以下の二点について注目する。

- ・センサの出力数にかかわらず，学習収束後の獲得報酬は変化しない
- ・センサの出力数が大きくなると，出力数の増加に比例して学習速度が遅くなる

これら二点の学習効率への影響が，タスクの違いによって変化するか否かを調査する。また，これらの学習効率への影響が変化する場合には，変化に明確な傾向があるのか，あるいは全く学習効率への影響が異なりタスク毎の比較は困難であるのかを明確にする。

状態数の異なるタスクにおける，センサの出力数の違いが学習効率に及ぼす影響の検証方法について述べる。検証方法の概要図を Fig.6.7 に示す。まず初めに状態数の異なる複数のタスクを用意する。次に，用意した各タスクに対してセンサの出力数の違いが学習効率に及ぼす影響の調査を行う。このとき各タスクにおける調査は，3.3 と同じ方法で行う。3.3 と同じ検証を異なるタスクで行うことで，各タスクにおいてセンサの出力数の違いが学習効率に及ぼす影響を示すことができる。

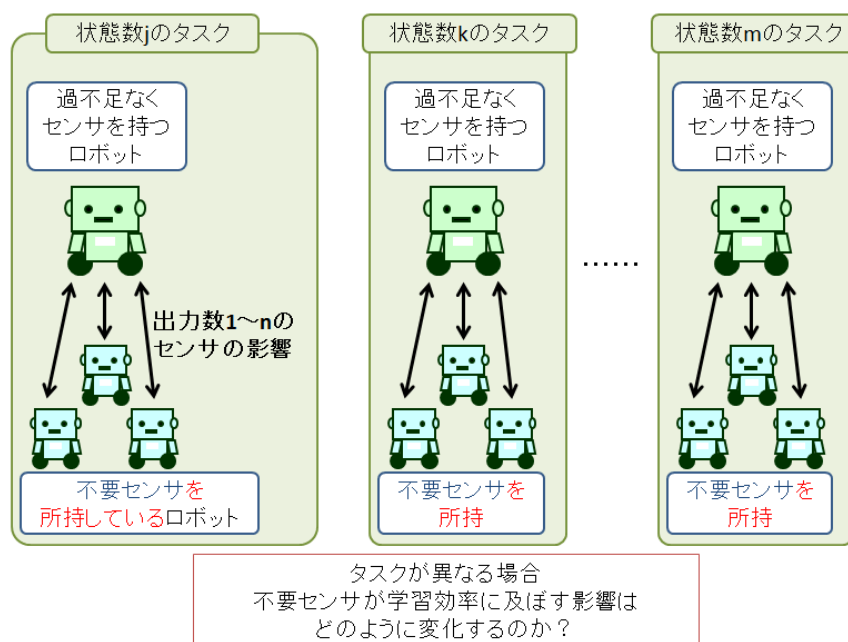


Fig.6.7 不要センサの出力数の違いが学習効率に及ぼす影響の検証方法概要図

各タスクでの実験が終了したら，全てのタスクに対して以下の二点を確認する。

- ・センサの出力数にかかわらず，学習収束後の獲得報酬は変化しない
- ・センサの出力数が大きくなると，出力数の増加に比例して学習速度が遅くなる

タスクが異なる場合でもこれらの影響が変化しなければ，5章までの実験結果はタスクが異なる場合にも同じように適用することができる。一方，タスクの変化によってこれらの影響が変化する場合には，5章までの実験結果はタスクが変わると適用できないこと

になる。この場合は、タスクの違いに対してこれらの影響の変化に規則性があるかどうかを確認する。規則性やある種の傾向が見られる場合にはそれを示し、タスクの状態数に対してセンサが学習に及ぼす影響の変化を示す。タスクの違いに対してこれらの影響の変化に規則性などが見られない場合には、今後の研究課題としてタスクの違いに対するセンサの学習効率の変化について調査の必要性を示す。

6.4. ロボットの所持センサの違いに関する検証方法

6.4.1. 具体的な調査対象

ロボットの所持センサが異なる場合に、センサの出力数の違いが学習効率に及ぼす影響の調査対象について述べる。ロボットの所持するセンサの違いについて、5章までの実験において調査が不足しているのは以下の四つの場合である。

- ・ロボットのセンサが不足している場合に、さらにセンサが不足したときの影響
- ・ロボットが不要センサを所持している場合に、さらに不要センサを搭載したときの影響
- ・ロボットのセンサが不足している場合に、ロボットに不要センサを搭載したときの影響
- ・ロボットが不要センサを所持している場合に、ロボットのセンサが不足したときの影響

ただしこのうち、ロボットが不要センサを所持している場合に、さらに不要センサを搭載したときの影響については調査の必要がない。5章までの実験結果に加え、センサの縮退の考え方を利用することで、複数の不要センサによる学習への影響を示すことができる。

センサの縮退を利用すると、状態数が同じとなるセンサの組が学習効率に及ぼす影響は全て同じであると言えることができる。例えば、出力数 6 の不要センサが学習効率に及ぼす影響と、出力数 2 の不要センサと出力数 3 の不要センサの組が学習に及ぼす影響は同じとなる。

ここで、出力数 n の不要センサを持つロボットに対して出力数 m の不要センサが学習効率に及ぼす影響を考える。この二つのセンサが学習効率に及ぼす影響は、出力数 $n \times m$ のセンサが学習効率に及ぼす影響と同じとなる。そのため、このロボットに対して出力数 m の不要センサが学習効率に及ぼす影響は、出力数 $n \times m$ のセンサの学習効率への影

響から出力数 n のセンサの学習効率への影響を引いたものであるとすることができる。不要センサはロボットの学習収束後の獲得報酬に影響を与えない。またロボット学習にかかる時間（ステップ数）については、不要センサの出力数に比例して増加する。過不足なくセンサを所持するロボットに対して出力数 2 のセンサを搭載したときの学習時間の増加量を k とする。このとき出力数 n のセンサであれば $n \times k$ 、出力数 $n \times m$ のセンサであれば $n \times m \times k$ だけ学習時間が増加する。このことから、出力数 n の不要センサを持つロボットに対して出力数 m の不要センサを搭載したとき、学習時間は m 倍になる。

このように、複数の不要センサによる学習効率への影響については 5 章の実験結果を用いることで算出することができるため、調査を行う必要はない。よって、調査の対象となるのは以下の三つとなる。

- ・ロボットのセンサが不足している場合に、さらにセンサが不足したときの影響
- ・ロボットのセンサが不足している場合に、ロボットに不要センサを搭載したときの影響
- ・ロボットが不要センサを所持している場合に、ロボットのセンサが不足したときの影響

6.4.2. 検証方法

ロボットの所持センサが異なる場合に、センサの出力数の違いが学習効率に及ぼす影響の検証方法について述べる。まず、一つのタスクを用意する。このとき、出力数 $1, 2, \dots, n$ のセンサを各一つずつ用いることで過不足なく状態を認識可能なタスクを用意する。次に、このタスクに対してロボットが持ちうる必要センサの組み合わせ全てを網羅するようにロボットを用意する。さらにこれらのロボットそれぞれに対し、不要センサを持たないロボット、出力数 $2, 3, \dots, m$ の不要センサのいずれか一つを持つロボットを用意する (Fig.6.8)。これら全てのロボットに対して共通の学習手法として強化学習を適用し、学習を行わせる。そして、各ロボットの学習効率を算出する。各ロボットの学習効率の差を比較することで、ロボットの所持センサが異なる場合にセンサの出力数の違いが学習効率に及ぼす影響を示す。

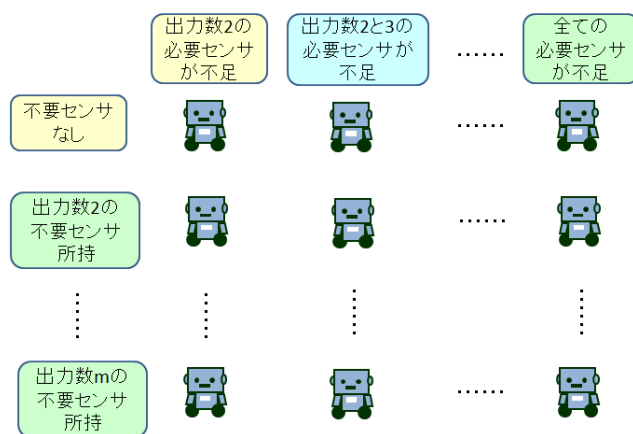


Fig.6.8 用意するロボット

調査対象ごとの、センサが学習に及ぼす影響の算出方法を述べる。算出方法を Fig.6.9 に示す。まず初めに、ロボットのセンサが不足している場合に必要センサが学習に及ぼす影響の算出方法について述べる。このときの調査対象は、任意の出力数 i のセンサが不足しているロボットに対して、出力数 $2,3,\dots,n$ の必要センサが学習に及ぼす影響である。そのため、出力数 i のセンサのみが不足しているロボットと、出力数 i のセンサに加え出力数 $2,3,\dots,n$ のうちいずれか一つのセンサが不足しているロボットとの学習効率の差を算出する。これにより、任意の出力数 i のセンサが不足しているロボットに対して、出力数 $2,3,\dots,n$ の必要センサが学習に及ぼす影響を示すことができる。

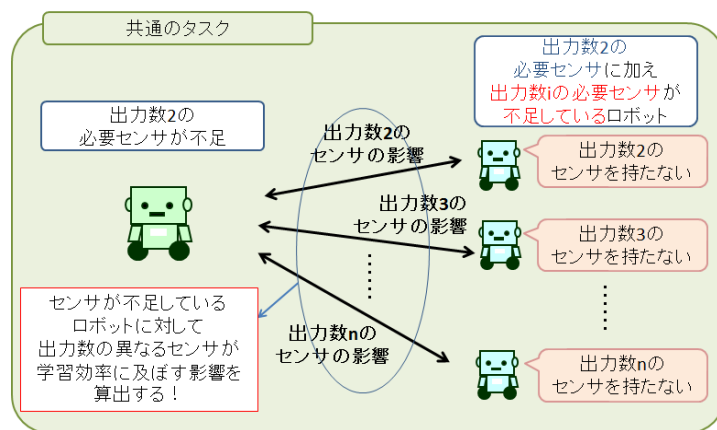


Fig.6.9 センサが不足している場合に必要センサが学習に及ぼす影響

次に、ロボットが過剰にセンサを所持している場合に必要センサが学習に及ぼす影響の算出方法について述べる。算出方法を Fig.6.9 に示す。このときの調査対象は、任意の出力数 i の不要センサを持つロボットに対して、出力数 $2,3,\dots,n$ の必要センサが学習に及

ぼす影響である。そのため、必要センサを全て持ち、出力数 i の不要センサを持つロボットと、出力数 i の不要センサを持ち、出力数 $2,3,\dots,n$ のうちいずれか一つのセンサが不足しているロボットとの学習効率の差を算出する。これにより、任意の出力数 i の不要センサを持つロボットに対して、出力数 $2,3,\dots,n$ の必要センサが学習に及ぼす影響を示すことができる。

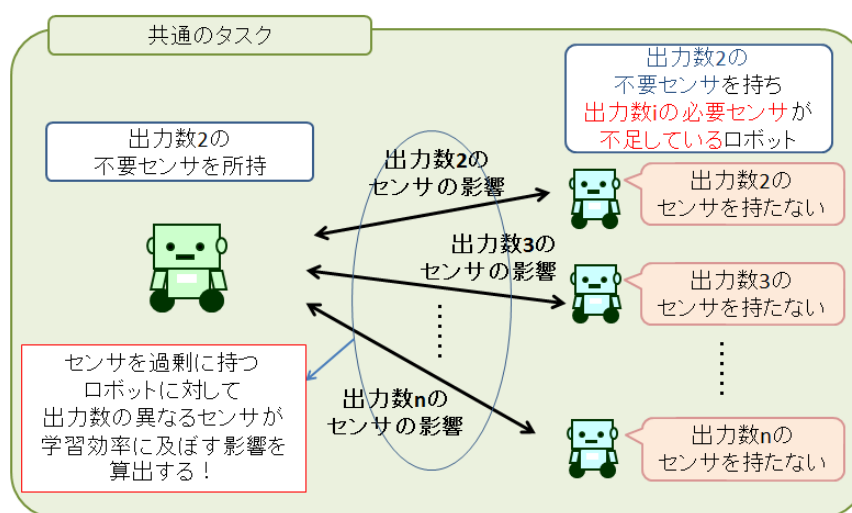


Fig.6.10 ロボットが過剰にセンサを所持している場合に必要センサが学習に及ぼす影響

最後に、ロボットのセンサが不足している場合に不要センサが学習に及ぼす影響の算出方法について述べる。算出方法を Fig.6.10 に示す。このときの調査対象は、任意の出力数 i のセンサが不足しているロボットに対して、出力数 $2,3,\dots,n$ の不要センサが学習に及ぼす影響である。そのため、出力数 i のセンサのみが不足しているロボットと、出力数 i のセンサが不足しており、出力数 $2,3,\dots,n$ のうちいずれか一つの不要センサを持つロボットとの学習効率の差を算出する。これにより、任意の出力数 i のセンサが不足しているロボットに対して、出力数 $2,3,\dots,n$ の不要センサが学習に及ぼす影響を示すことができる。

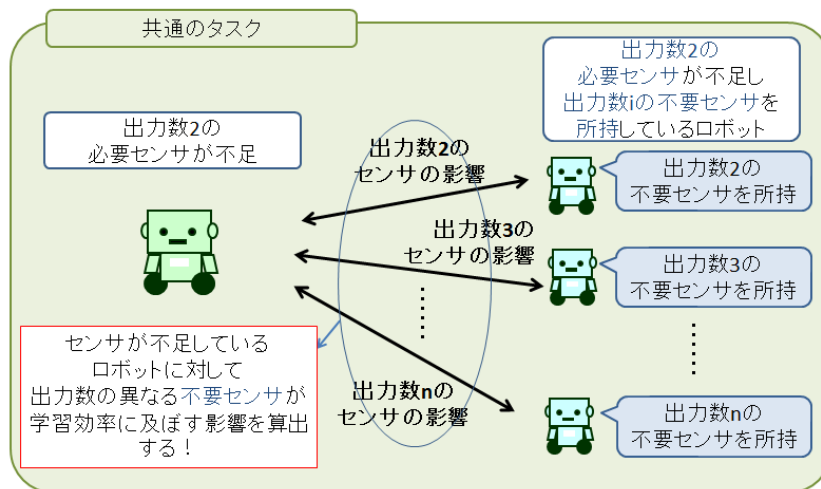


Fig.6.11 ロボットが過剰にセンサを所持している場合に不要センサが学習に及ぼす影響

結果の比較は調査対象ごとに分類して行う。まず初めに、ロボットのセンサが不足している場合に、必要センサが学習効率に及ぼす影響の比較方法について述べる。このとき注目するのは以下の二点である。

- ・センサの出力数の増加に伴い、学習収束後の獲得報酬の減少量が増加する
- ・センサの出力数が大きくなると、出力数の増加による獲得報酬の減少量が増えなくなる

ロボットのセンサに不足が生じている場合でもこれらの影響が変化しなければ、5章までの実験結果はセンサが不足している場合にも同じように適用することができる。一方、センサの不足によってこれらの影響が変化する場合には、不足しているセンサの出力数に対して、センサの出力数の違いが学習効率に及ぼす影響の変化に規則性があるかどうかを確認する。規則性やある種の傾向が見られる場合にはそれを示し、不足しているセンサの出力数に対してセンサが学習に及ぼす影響の変化を示す。

次に、ロボットが過剰にセンサを所持している場合に、必要センサが学習効率に及ぼす影響の比較方法について述べる。このとき注目するのは以下の二点である。

- ・センサの出力数の増加に伴い、学習収束後の獲得報酬の減少量が増加する
- ・センサの出力数が大きくなると、出力数の増加による獲得報酬の減少量が増えなくなる

ロボットが過剰にセンサを所持している場合でもこれらの影響が変化しなければ、5章までの実験結果は過剰にセンサを所持している場合にも同じように適用することができる。一方、不要センサによってこれらの影響が変化する場合には、不要センサの出力数

に対して、必要センサの出力数の違いが学習効率に及ぼす影響の変化に規則性があるかどうかを確認する。規則性やある種の傾向が見られる場合にはそれを示し、不要センサの出力数に対して必要センサが学習に及ぼす影響の変化を示す。

最後に、ロボットのセンサが不足している場合に、不要センサが学習効率に及ぼす影響の比較方法について述べる。このとき注目するのは以下の二点である。

- ・センサの出力数にかかわらず、学習収束後の獲得報酬は変化しない
- ・センサの出力数が大きくなると、出力数の増加に比例して学習速度が遅くなる

ロボットのセンサに不足が生じている場合でもこれらの影響が変化しなければ、5章までの実験結果はセンサが不足している場合にも同じように適用することができる。一方、センサの不足によってこれらの影響が変化する場合には、不足しているセンサの出力数に対して、不要センサの出力数の違いが学習効率に及ぼす影響の変化に規則性があるかどうかを確認する。規則性やある種の傾向が見られる場合にはそれを示し、不足しているセンサの出力数に対して不要センサが学習に及ぼす影響の変化を示す。

7. タスク及び所持センサの違いに対する実験方法

7.1. タスクの違いに対する必要センサの影響の実験方法

7.1.1. 実験の流れ

タスクの違いに対して必要センサが学習効率に及ぼす影響の実験概要について述べる。実験の概要図を Fig.7.1 に示す。まず初めに、必要センサの中に調査対象のセンサが全て含まれるタスクを用意する。具体的には、出力数 $2,3,\dots,n$ のセンサが全て必要センサに含まれるタスクとなる。次に一体のロボットを用意し、このロボットに出力数が $2,3,\dots,n$ のセンサをひとつずつ搭載する。さらに、出力数が $2,3,\dots,n$ のセンサのうちいずれか一つのみ持たないロボットを各一体ずつ用意する。これらのロボットに対して共通のタスクを与えて学習を行わせる。ただし、一つのタスクのみで実験を行うとランダムな要素によって特定のロボットに対して有利な状況が生まれる可能性がある。そこで、必要センサが同じであるタスクを E 個生成し、 E 個のタスクに対して学習を行わせる。そして各タスクでの学習結果の平均から各ロボットの学習効率を算出する。また、学習を行わずランダムに行動するロボットを用意し、他のロボットと同様に同じタスクにて行動させる。このロボットの行動結果は、学習を行うロボットの学習効率を算出する際に利用する。以上の実験の結果から各ロボットの学習効率を算出し、センサの出力数の違いに対する学習効率への影響を示す。

この実験をさらに、状態数の異なる複数のタスクに対して行う。このとき、全てのタスクについて調査対象のセンサが全て必要センサに含まれるタスクを用意する。そして、各タスクにおいてセンサが学習に及ぼす影響が変化するのか否かを調査する。また、変化する場合にはその特徴や傾向を示す。

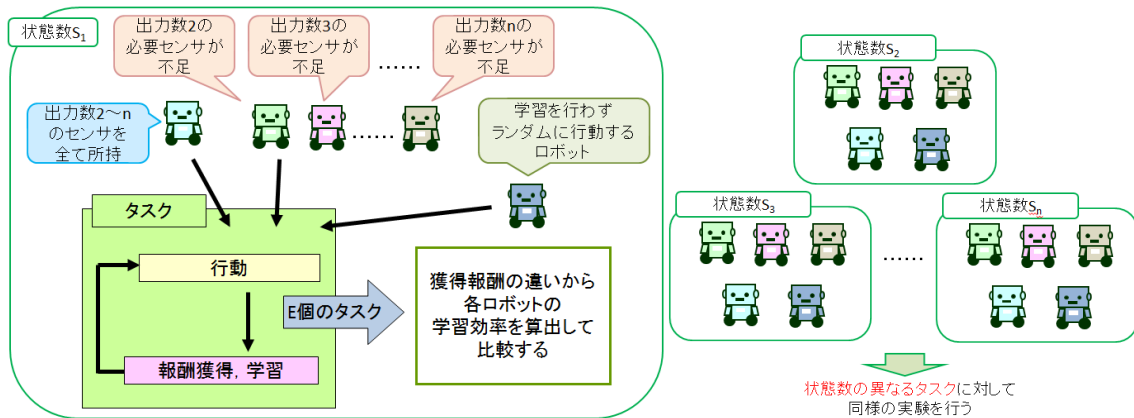


Fig.7.1 必要センサを対象にした実験の概要図

実験の際、全てのロボットは共通の学習手法として強化学習を用いる。このとき必要センサの調査と同様に、強化学習の行動学習手法として Q 学習，行動選択手法として ϵ -greedy 法を用いる。Q 学習や ϵ -greedy 法については 4.1.2 で述べたため省略する。

7.1.2. 実験に用いるタスク

ロボットが行うタスクについて述べる。ロボットが行うタスクは 4.1.3 で用いたタスクと同じであり、ルート探索のタスクである。このタスクはいくつかの状態から構成されている。このタスクは出力数 2,3,...,n の各センサを必要センサとして含む。また、これらのセンサに加えて任意の出力数のセンサを必要センサとすることで、タスクの状態数を調整する。ロボットが持つセンサが不足している場合、ロボットが認識できる状態数が減少することになる。タスクの各状態はそれぞれ一つの報酬と二つの遷移先を持つ。ロボットは各状態において行動を選択することで、選択した行動に対応した遷移先の状態へ確定的に遷移する。そして、遷移先の状態に応じて報酬を獲得する。ロボットの目的は、このタスクにおいて報酬の獲得量が最大になるルートの探索である。

次に、タスクの報酬設定について述べる。報酬設定についても 4.1.3 で述べたものと同様の設定を用いる。状態 i に与えられる報酬 Rwd_i を決定する式を式 (4.2) に示す。

$$Rwd_i = e^{\beta \cdot D_i} \quad (4.2)$$

実験ではこのタスクをランダムに E 個生成する。E 個のタスクにおける学習結果の平均を用いることでタスクによって生じる特定のロボットにのみ生じる有利，不利な条件を打ち消す。ロボットが行動を選択し、状態遷移して学習を行うまでを 1 ステップとする。各ロボットは同じ状態からタスクを開始し、一つのタスクにつき R ステップの学習

を行う。そして、各タスクでの獲得報酬の平均を用いてロボット毎に学習効率を算出する。

7.1.3.結果の比較方法

結果の比較方法について述べる。結果の比較の手順を以下に示す。

- 1.ある状態数のタスクに対して、各ロボットの学習効率を算出する
- 2.手順1で算出した各ロボットの学習効率を利用して、任意のセンサが学習効率に及ぼす影響を算出する
- 3.出力数 2,3,...,n のセンサについて手順2を行う
- 4.手順2, 3で算出した学習効率への影響が、出力数によってどのように変化するか比較を行う。
- 5.状態数の異なるタスクに対して手順1~4を行う。その後、手順4で示した学習効率への影響が、タスクの状態数によってどのように変化するか比較を行う。以下では、各手順における学習効率の算出方法や比較方法について述べていく。

手順1における各ロボットの学習効率の算出方法について述べる。手順1では4.1.4と同様にSMAの推移を用いて各ロボットの学習効率の比較を行う。tステップ目におけるSMAの式を式(4.3)に示す。 $r_{k,t}$ は、k個目のタスクにおいてロボットがtステップ目に得た報酬の値である。

$$SMA(t) = \sum_{k=1}^E \frac{r_{k,t} + r_{k,t-1} + \dots + r_{k,t-L+1}}{L} \quad (4.3)$$

手順2における、任意のセンサが学習効率に及ぼす影響の算出方法について述べる。手順2においても4.1.4と同様に、各センサのRSMAを算出することで出力数iのセンサが学習効率に及ぼす影響を示す。tステップ目において出力数iのセンサが学習効果に及ぼす影響(RSMA)を式(4.4)によって算出する。

$$RSMA_i(t) = \frac{SMA_1(t) - SMA_i(t)}{SMA_1(t) - SMA_{random}(t)} \quad (4.4)$$

これらの結果の比較をすることで、センサが学習効率に及ぼす影響として次の三点を見ることができる。

- ・センサが不足すると、学習開始時から終了時まで獲得報酬量が減少するのか
- ・センサの出力数の増加に伴い、学習収束後の獲得報酬の減少量が増加するか

・センサの出力数が大きくなると、出力数の増加による獲得報酬の減少量が増えなくなるか

この三点の影響が生じることを、状態数の異なるタスク毎に確認する。状態数が異なる場合にもこれらの影響が現れる場合、状態数にかかわらず 5 章までの実験結果をロボットに適用することができる。一方、タスクの状態数の違いによってこれら三点とは異なる影響が現れた場合には、影響の変化傾向について考察を行う。

7.1.4 実験パラメータ

実験パラメータを Table 7.1 に示す。今回は 5 章までの実験と同じく出力数が 2,3,4,5 のセンサを調査対象とした。実験に用いたタスクの状態数については 4, 5 章での実験で用いた 120 を基点とし、倍の 240, 3 倍の 360 のタスクを用いて実験を行った。

Table 7.1 必要センサに対する実験の実験パラメータ

n (調査対象のセンサの出力数の最大値)	5	β	0.65
タスクの状態数	120, 240, 360	α	0.7
R(学習ステップ数)	100000	γ	0.6
E(実験を行うタスクの数)	300	ε	0.1
Q の初期値	500	L	500

7.2.タスクの違いに対する不要センサの影響の実験方法

7.2.1.実験の流れ

タスクの違いに対して不要センサが学習効率に及ぼす影響の実験概要について述べる。実験の概要図を Fig.7.2 に示す。まず初めに、ロボットに学習を行わせるためのタスクを一つ用意する。次に一体のロボットを用意し、このロボットに過不足なくセンサを搭載する。さらに、タスクに必要なセンサに加えて出力数が 2,3,...,n の不要センサのうちいずれか一つを持つロボットを各一体ずつ用意する。これらのロボットに対して共通のタスクを与えて学習を行わせる。ただし、一つのタスクのみで実験を行うとランダムな要素によって特定のロボットに対して有利な状況が生まれる可能性がある。そこで、必要センサが同じであるタスクを E 個生成し、E 個のタスクに対して学習を行わせる。そして各タスクでの学習結果の平均から各ロボットの学習効率を算出する。また、学習を行わずランダムに行動するロボットを用意し、他のロボットと同様に同じタスクにて行動さ

せる．このロボットの行動結果は，学習を行うロボットの学習効率を算出する際に利用する．以上の実験の結果から各ロボットの学習効率を算出し，センサの出力数の違いに対する学習効率への影響を示す．

この実験をさらに，状態数の異なる複数のタスクに対して行う．このとき，全てのタスクについて調査対象のセンサが全て必要センサに含まれるタスクを用意する．そして，各タスクにおいてセンサが学習に及ぼす影響が変化するか否かを調査する．また，変化する場合にはその特徴や傾向を示す．

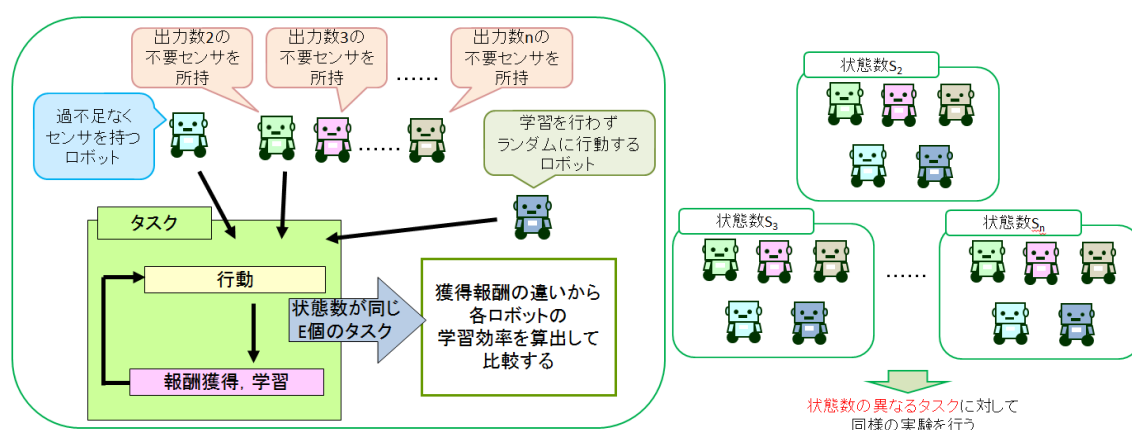


Fig.7.2 必要センサを対象にした実験の概要図

実験の際，全てのロボットは共通の学習手法として強化学習を用いる．このとき必要センサの調査と同様に，強化学習の行動学習手法として Q 学習，行動選択手法として ϵ -greedy 法を用いる．Q 学習や ϵ -greedy 法については 4.1.2 で述べたため省略する．

7.2.2. 実験に用いるタスク

ロボットが行うタスクについて述べる．ロボットが行うタスクは 4.1.3 で用いたタスクと同じであり，ルート探索のタスクである．このタスクはいくつかの状態から構成されている．ロボットがセンサを過剰に所持している場合，ロボットの認識できる状態数が増加する．タスクの各状態はそれぞれ一つの報酬と二つの遷移先を持つ．ロボットは各状態において行動を選択することで，選択した行動に対応した遷移先の状態へ確定的に遷移する．そして，遷移先の状態に応じて報酬を獲得する．ロボットの目的は，このタスクにおいて報酬の獲得量が最大になるルートの探索である．

次に，タスクの報酬設定について述べる．報酬設定についても 4.1.3 で述べたものと同様の設定を用いる．状態 i に与えられる報酬 Rwd_i を決定する式を式 (4.2) に示す．

$$Rwd_i = e^{\beta \cdot D_{ii}} \quad (4.2)$$

実験ではこのタスクをランダムに E 個生成する。E 個のタスクにおける学習結果の平均を用いることでタスクによって生じる特定のロボットにのみ生じる有利、不利な条件を打ち消す。ロボットが行動を選択し、状態遷移して学習を行うまでを 1 ステップとする。各ロボットは同じ状態からタスクを開始し、一つのタスクにつき R ステップの学習を行う。そして、各タスクでの獲得報酬の平均を用いてロボット毎に学習効率を算出する。

7.2.3.結果の比較方法

不要センサについては特に学習速度に注目して行う。ただし、学習収束後の獲得報酬が変化しないことも確認する。結果の比較の手順を以下に示す。

- 1.ある状態数のタスクに対して、各ロボットの学習効率を算出する
- 2.各ロボットの学習収束後の獲得報酬を比較し、学習収束後の獲得報酬に変化が無いことを確認する
- 3.手順 1 で算出した各ロボットの学習効率を利用して、任意のセンサが学習速度に及ぼす影響を算出する
- 4.出力数 2,3,...,n のセンサについて手順 2 を行う
- 5.手順 2, 3 で算出した学習速度への影響が、出力数によってどのように変化するか比較を行う。
- 6.状態数の異なるタスクに対して手順 1~5 を行う。その後、手順 2, 5 で示した学習効率への影響が、タスクの状態数によってどのように変化するか比較を行う。以下では、各手順における学習効率の算出方法や比較方法について述べていく。

手順 1 における各ロボットの学習効率の算出方法について述べる。必要センサでの実験と同様に、時刻 t において、ロボットが直近の L ステップで得た報酬の平均 (SMA) を算出し、SMA の推移を用いて各ロボットの学習効率の比較を行う。t ステップ目における SMA の式を式 (4.4) に示す。r_{k,t} は、k 個目のタスクにおいてエージェントが t ステップ目に得た報酬の値である。

$$SMA(t) = \sum_{k=1}^E \frac{r_{k,t} + r_{k,t-1} + \dots + r_{k,t-L+1}}{L} \quad (4.4)$$

不要センサの実験では、各ロボットの学習収束後の SMA に注目する。学習収束後の各

ロボットの SMA に違いが無ければ、センサの出力数にかかわらず不要センサは学習収束後の獲得報酬に影響を与えないといえる。

学習速度を比較する手法として、4.2.3 と同様に時定数を利用する。この実験では、エージェントが獲得した SMA の最低値を 0%，ロボットが獲得した SMA の最大値（最終的に取る値）を 100%としたときに、ロボットの SMA が 63%に到達するまでのステップ数を時定数とする。各ロボットについて時定数を算出する。その上で、センサを過不足なく持つロボットの時定数を 1としたときに出力数*i*の不要センサを持つロボットの時定数が何%増加したのか、その増加割合を比較する。これにより、センサの出力数の違いによる学習速度への影響の違いを示す。

7.2.4.実験パラメータ

実験パラメータを Table 7.2 に示す。今回は出力数が 2,3,4,5,6 の不要センサを調査対象とした。実験に用いたタスクの状態数については 4, 5 章での実験で用いた 120 を基点とし、1/2 の 60, 1/3 の 40 のタスクを用いて実験を行った。また、今回は不要センサとして、出力する可能性のある値をランダムにひとつ出力するものを用いた。

Table 7.2 必要センサに対する実験の実験パラメータ

n (調査対象のセンサの出力数の最大値)	5	β	0.65
タスクの状態数	120, 60, 40	α	0.7
R(学習ステップ数)	100000	γ	0.6
E(実験を行うタスクの数)	300	ε	0.1
Qの初期値	500	L	500

7.3.ロボットの所持センサの違いに関する実験方法

7.3.1.実験の流れ

ロボットの所持センサが異なる場合に対する、出力数の異なるセンサによる学習効率の変化についての実験方法について述べる。用意するロボットを Fig.7.3, 実験の概要図を Fig.7.4 に示す。まず初めに、必要センサの中に調査対象のセンサが全て含まれるタスクを用意する。具体的には、出力数 2,3,...,n のセンサが全て必要センサに含まれるタスクとなる。次に、このタスクに対してロボットが持ちうる必要センサの組み合わせ全てを網羅するようにロボットを用意する。さらにこれらのロボットそれぞれに対し、不要

センサを持たないロボット，出力数 $2,3,\dots,m$ の不要センサのいずれか一つを持つロボットを用意する．これら全てのロボットに対して共通の学習手法として強化学習を適用し，学習を行わせる．ただし，一つのタスクのみで実験を行うとランダムな要素によって特定のロボットに対して有利な状況が生まれる可能性がある．そこで，必要センサが同じであるタスクを E 個生成し， E 個のタスクに対して学習を行わせる．そして各タスクでの学習結果の平均から各ロボットの学習効率を算出する．また，学習を行わずランダムに行動するロボットを用意し，他のロボットと同様に同じタスクにて行動させる．このロボットの行動結果は，学習を行うロボットの学習効率を算出する際に利用する．以上の実験の結果から各ロボットの学習効率を算出し，センサの出力数の違いに対する学習効率への影響を示す．

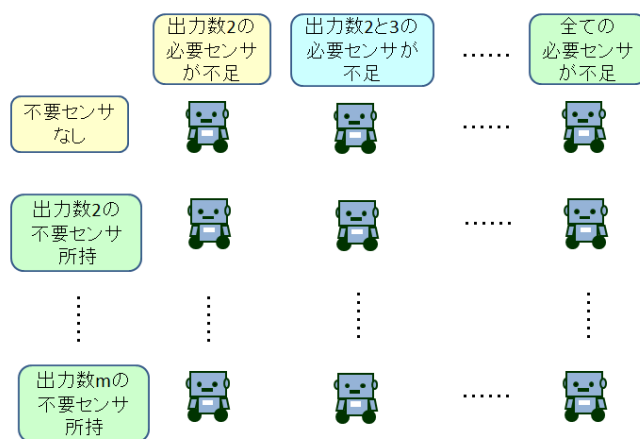


Fig.7.3 用意するロボット

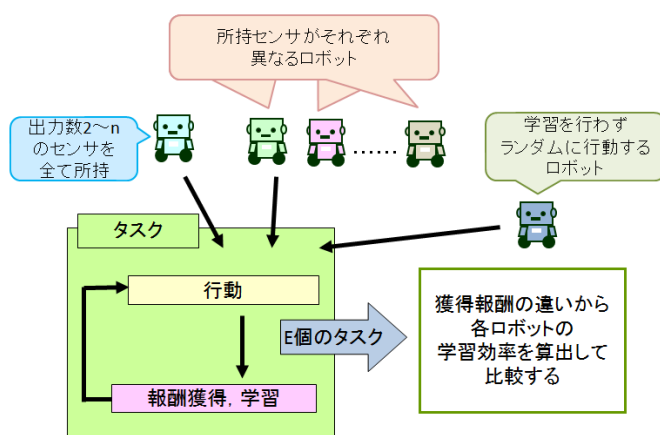


Fig.7.4 ロボットの所持センサの違いに対する実験の概要図

実験の際，全てのロボットは共通の学習手法として強化学習を用いる．このとき必要センサの調査と同様に，強化学習の行動学習手法として Q 学習，行動選択手法として ϵ

-greedy 法を用いる。Q 学習や ϵ -greedy 法については 4.1.2 で述べたため省略する。

7.3.2. 実験に用いるタスク

ロボットが行うタスクについて述べる。ロボットが行うタスクはルート探索のタスクである。このタスクはいくつかの状態から構成されており、ロボットは出力数 $2, 3, \dots, n$ の必要センサを一つずつ、合計 $n-1$ 個のセンサを用いることですべての状態を過不足なく認識することが出来る。そのため、タスクの状態数は $n!$ 個となる。ロボットの持つセンサが不足している場合、ロボットが認識できる状態数が減少する。また、ロボットがセンサを過剰に所持している場合、ロボットの認識できる状態数が増加する。タスクの各状態はそれぞれ一つの報酬と二つの遷移先を持つ。ロボットは各状態において行動を選択することで、選択した行動に対応した遷移先の状態へ確定的に遷移する。そして、遷移先の状態に応じて報酬を獲得する。ロボットの目的は、このタスクにおいて報酬の獲得量が最大になるルートの探索である。

7.3.3. 結果の比較方法

結果の比較方法について述べる。結果の比較方法は、調査の対象によって異なる。まず初めに、ロボットのセンサが不足している場合に必要センサが学習効率に及ぼす影響について、結果の比較の手順を以下に示す。

1. 各ロボットの学習効率を算出する
2. 手順 1 で算出した各ロボットの学習効率を利用して、任意のセンサが学習効率に及ぼす影響を算出する
3. 出力数 $2, 3, \dots, n$ のセンサについて手順 2 を行う
4. 手順 2, 3 で算出した学習効率への影響が、出力数によってどのように変化するか比較を行う。

手順 1 における各ロボットの学習効率の算出方法について述べる。手順 1 では 4.1.4 と同様に SMA の推移を用いて各ロボットの学習効率の比較を行う。t ステップ目における SMA の式を式 (4.3) に示す。 $r_{k,t}$ は、k 個目のタスクにおいてロボットが t ステップ目に得た報酬の値である。

$$SMA(t) = \sum_{k=1}^E \frac{r_{k,t} + r_{k,t-1} + \dots + r_{k,t-L+1}}{L} \quad (4.3)$$

手順 2 における、任意のセンサが学習効率に及ぼす影響の算出方法について述べる。手順 2 においても 4.1.4 と同様に、各センサが学習効果に及ぼす影響を算出することで出力数 i のセンサが学習効率に及ぼす影響を示す。ただし今回は、センサが不足しているロボットに対して出力数 i のセンサが学習効率に及ぼす影響が比較の対象となる。そのため学習効果に及ぼす影響を算出する際にも、センサが不足しているロボットと出力値 i のセンサを持たないロボットの SMA （獲得報酬）の差を利用する。 t ステップ目において、出力数 j のセンサを持たないロボットに対して出力数 i のセンサが学習効果に及ぼす影響 ($MSMA_{j,i}$) を式 (7.1) によって算出する。

$$MSMA_{j,i}(t) = \frac{SMA_j(t) - SMA_i(t)}{SMA_j(t) - SMA_{random}(t)} \quad (7.1)$$

SMA_j は必要センサのうち出力数 j のセンサのみが不足しているロボットの SMA である。 SMA_i は出力値 i のセンサを持たないエージェントの SMA である。 SMA_{random} は学習を行わずランダムに行動するエージェントの SMA である。 $RSMA$ が正の値であれば、ロボットから出力数 i のセンサを取り除いたときの学習効果は減少することを示す。逆に $RSMA$ が負の値であれば、学習効果が上昇することを示す。また、 $RSMA$ の絶対値が大きいほど出力数 i のセンサが学習効果に及ぼす影響は大きいことを示す。

これらの結果の比較をすることで、センサが学習効率に及ぼす影響として次の三点を見ることができる。

- ・センサが不足すると、学習開始時から終了時まで獲得報酬量が減少するのか
- ・センサの出力数の増加に伴い、学習収束後の獲得報酬の減少量が増加するか
- ・センサの出力数が大きくなると、出力数の増加による獲得報酬の減少量が増えなくなるか

ロボットのセンサが不足している場合に、この三点の影響が生じるか否かの確認を行う。ロボットのセンサが不足している場合にもこれらの影響が現れる場合、ロボットの所持センサが不足している場合であっても 5 章までの実験結果をロボットに適用することができる。一方、センサの不足によってこれら三点とは異なる影響が現れた場合には、センサの不足による学習効率への影響の変化の傾向について考察を行う。

次に、ロボットが過剰にセンサを持つ場合に必要センサが学習効率に及ぼす影響について、結果の比較の手順を以下に示す。

- 1.各ロボットの学習効率を算出する
- 2.手順 1 で算出した各ロボットの学習効率を利用して、任意のセンサが学習効率に及ぼ

す影響を算出する

3.出力数 2,3,...,n のセンサについて手順 2 を行う

4.手順 2, 3 で算出した学習効率への影響が, 出力数によってどのように変化するか比較を行う.

手順 1 における各ロボットの学習効率の算出方法について述べる. 手順 1 では 4.1.4 と同様に SMA の推移を用いて各ロボットの学習効率の比較を行う. t ステップ目における SMA の式を式 (4.3) に示す. $r_{k,t}$ は, k 個目のタスクにおいてロボットが t ステップ目に得た報酬の値である.

$$SMA(t) = \sum_{k=1}^E \frac{r_{k,t} + r_{k,t-1} + \dots + r_{k,t-L+1}}{L} \quad (4.3)$$

手順 2 における, 任意のセンサが学習効率に及ぼす影響の算出方法について述べる. 手順 2 においても 4.1.4 と同様に, 各センサの学習効果に及ぼす影響を算出することで出力数 i のセンサが学習効率に及ぼす影響を示す. ただし今回は, 不要センサを持つロボットに対して出力数 i のセンサが学習効率に及ぼす影響が比較の対象となる. そのため学習効果に及ぼす影響を算出する際にも, 不要センサを持つロボットと出力値 i のセンサを持たないロボットの SMA (獲得報酬) の差を利用する. t ステップ目において, 出力数 j のセンサを持たないロボットに対して出力数 i のセンサが学習効果に及ぼす影響 ($DSMA_{k,i}$) を式 (7.2) によって算出する.

$$DSMA_{k,i}(t) = \frac{SMA_k(t) - SMA_i(t)}{SMA_k(t) - SMA_{random}(t)} \quad (7.2)$$

SMA_k は出力数 k の不要センサを持つロボットの SMA である. SMA_i は出力値 i のセンサを持たないエージェントの SMA である. SMA_{random} は学習を行わずランダムに行動するエージェントの SMA である. $DSMA$ が正の値であれば, ロボットから出力数 i のセンサを取り除いたときの学習効果は減少することを示す. 逆に $DSMA$ が負の値であれば, 学習効果が上昇することを示す. また, $DSMA$ の絶対値が大きいくほど出力数 i のセンサが学習効果に及ぼす影響は大きいことを示す.

これらの結果の比較をすることで, センサが学習効率に及ぼす影響として次の三点を見ることができる.

- ・センサが不足すると, 学習開始時から終了時まで獲得報酬量が減少するのか
- ・センサの出力数の増加に伴い, 学習収束後の獲得報酬の減少量が増加するか
- ・センサの出力数が大きくなると, 出力数の増加による獲得報酬の減少量が増えなく

なるか

ロボットが不要センサを持つ場合に、この三点の影響が生じるか否かの確認を行う。ロボットが不要センサを持つ場合にもこれらの影響が現れる場合、ロボットが不要センサを所持していても 5 章までの実験結果をロボットに適用することができる。一方、不要センサによってこれら三点とは異なる影響が現れた場合には、不要センサによる学習効率への影響の変化の傾向について考察を行う。

最後に、ロボットのセンサが不足している場合に必要センサが学習効率に及ぼす影響について、結果の比較の手順を以下に示す。

- 1.ある状態数のタスクに対して、各ロボットの学習効率を算出する
- 2.各ロボットの学習収束後の獲得報酬を比較し、学習収束後の獲得報酬に変化が無いことを確認する
- 3.手順 1 で算出した各ロボットの学習効率を利用して、任意のセンサが学習速度に及ぼす影響を算出する
- 4.出力数 2,3,...,n のセンサについて手順 2 を行う
- 5.手順 2, 3 で算出した学習速度への影響が、出力数によってどのように変化するのかわかりやすく比較を行う。

・各エージェントの学習効率 (SMA) の算出方法について

手順 1 における各ロボットの学習効率の算出方法について述べる。必要センサでの実験と同様に、時刻 t において、ロボットが直近の L ステップで得た報酬の平均 (SMA) を算出し、SMA の推移を用いて各ロボットの学習効率の比較を行う。 t ステップ目における SMA の式を式 (4.4) に示す。 $r_{k,t}$ は、 k 個目のタスクにおいてエージェントが t ステップ目に得た報酬の値である。

$$SMA (t) = \sum_{k=1}^E \frac{r_{k,t} + r_{k,t-1} + \dots + r_{k,t-L+1}}{L} \quad (4.4)$$

不要センサの実験では、各ロボットの学習収束後の SMA に注目する。学習収束後の各ロボットの SMA に違いが無ければ、センサの出力数にかかわらず不要センサは学習収束後の獲得報酬に影響を与えないといえる。

学習速度を比較する手法として、4.2.3 と同様に時定数を利用する。この実験では、エージェントが獲得した SMA の最低値を 0%、ロボットが獲得した SMA の最大値 (最終的に取る値) を 100%としたときに、ロボットの SMA が 63%に到達するまでのステップ

数を時定数とする。各ロボットについて時定数を算出する。この実験における時定数の比較対象は、出力数 j の必要センサが不足しており不要センサを持たないロボットと、出力数 j の必要センサが不足しており、さらに出力数 i の不要センサを持つロボットである。出力数 j の必要センサが不足しており不要センサを持たないロボットの時定数を 1 としたときに、出力数 j の必要センサが不足しており、さらに出力数 i の不要センサを持つロボットの時定数が何%増加したのか、その増加割合を比較する。これにより、センサの出力数の違いによる学習速度への影響の違いを示す。

これらの結果の比較をすることで、センサが学習効率に及ぼす影響として次の二点を見ることができる。

- ・センサの出力数にかかわらず、学習収束後の獲得報酬は変化しない
- ・センサの出力数が大きくなると、出力数の増加に比例して学習速度が遅くなる

ロボットのセンサが不足している場合に、この二点の影響が生じるか否かの確認を行う。ロボットのセンサが不足している場合にもこれらの影響が現れる場合、ロボットの所持センサが不足している場合であっても 5 章までの実験結果をロボットに適用することができる。一方、センサの不足によってこれら二点とは異なる影響が現れた場合には、センサの不足による学習効率への影響の変化の傾向について考察を行う。

7.3.4. 実験パラメータ

実験パラメータを Table 7.3 に示す。今回は出力数が 2,3,4,5 の必要センサと、出力数が 2,3,4,5,6 の不要センサを調査対象とした。そのため、タスクは出力数 2,3,4,5 のセンサを一つずつ用いることで過不足なく認識可能なタスクを用いた。また、今回は不要センサとして、出力する可能性のある値をランダムにひとつ出力するものを用いた。

Table 7.3 必要センサに対する実験の実験パラメータ

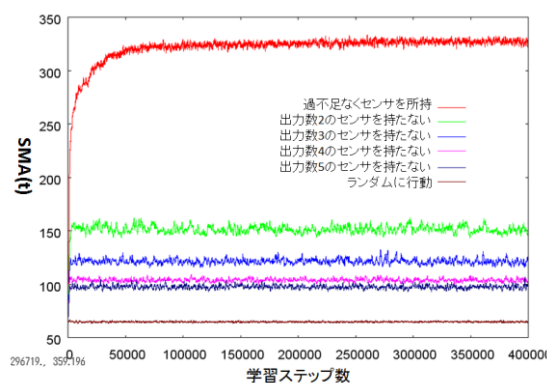
n (必要センサの出力数の最大値)	5	β	0.65
M (不要センサの出力数の最大値)	6	α	0.7
タスクの状態数	120	γ	0.6
R(学習ステップ数)	100000	ε	0.1
E(実験を行うタスクの数)	300	L	500
Qの初期値	500		

8. タスク及び所持センサの違いに対する実験結果

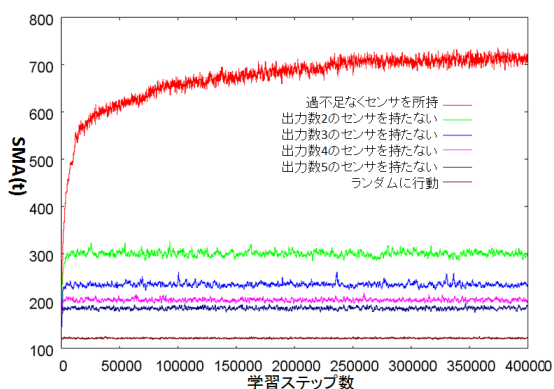
8.1. タスクの違いに対する必要センサの影響の実験

8.1.1. 実験結果

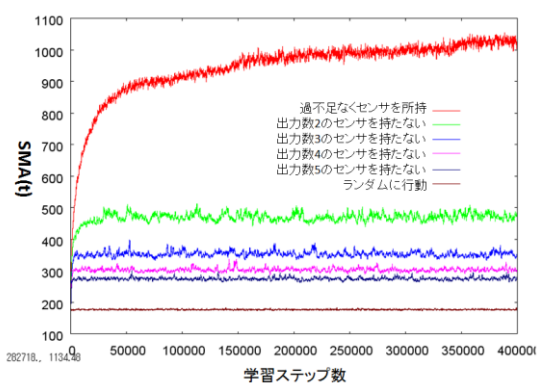
タスクの状態数が異なる場合に、必要センサが学習効率に及ぼす影響の実験結果を示す。実験結果を Fig.8.1～Fig.8.4 に示す。Fig.8.1 は、状態数 120, 240, 360 のタスクそれぞれにおいて、各ロボットが獲得した SMA の推移である。Fig.8.1 は、40 万ステップまでに全てのロボットが学習収束していることを確認するためのグラフである。全てのロボットについて、学習終了前数万ステップに渡って同程度の SMA を獲得している。そのため全てのロボットが学習収束しているといえる。



(a) 状態数 120 のタスク



(b) 状態数 240 のタスク



(c) 状態数 360 のタスク

Fig.8.1 各タスクにおける各ロボットの SMA の推移

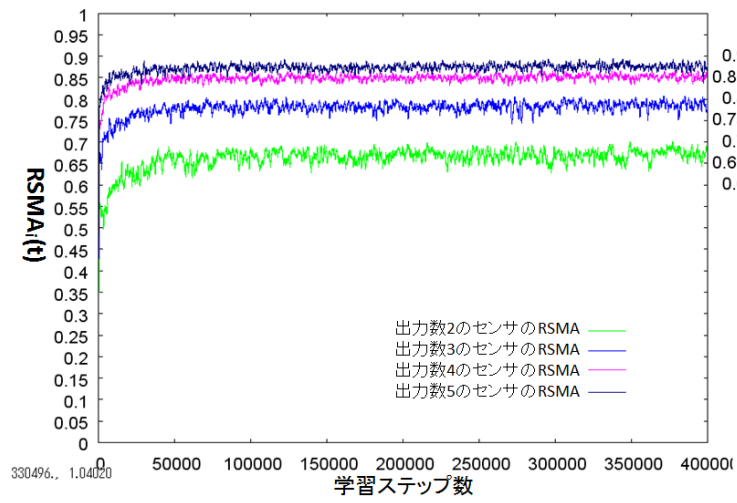


Fig.8.2 状態数 120 のタスクにおける，出力数が異なるセンサの RSMA の推移

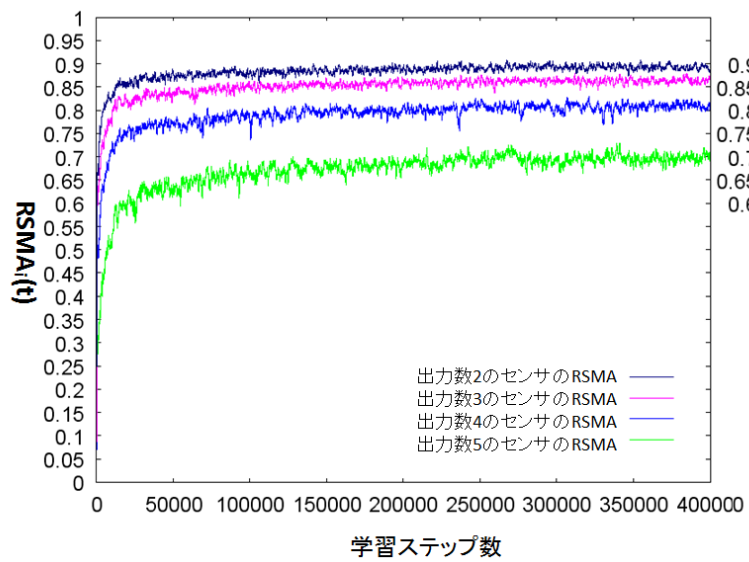


Fig.8.3 状態数 240 のタスクにおける，出力数が異なるセンサの RSMA の推移

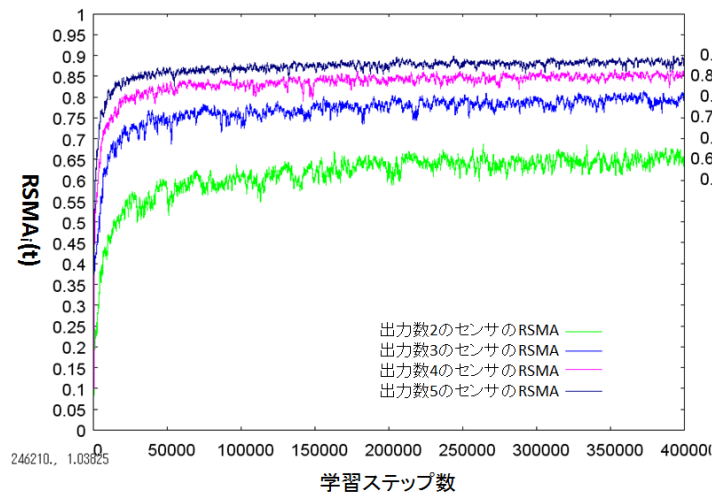


Fig.8.4 状態数 360 のタスクにおける，出力数が異なるセンサの RSMA の推移

Fig.8.2～Fig.8.4 は，各タスクにおいて出力数が異なるセンサの RSMA の推移である．これらのグラフは出力数 i のセンサによって変化した獲得報酬の割合の推移を表している．RSMA が大きいほど獲得報酬量の差が大きく，正の値であれば獲得報酬が減少したことになる．まず初めに，Fig.8.2～Fig.8.4 において，いずれのタスクのいずれのセンサに関しても学習初期から最後まで正の値となっている．すなわち，タスクの状態数に関わらず，センサが不足すると学習全体を通して獲得報酬が減少することがわかる．

次に，Fig.8.2～Fig.8.4 において各センサの学習収束後（40 万ステップ付近）の RSMA の大きさに注目する．どのタスクについても 5 章での結果と同様に，出力数が大きいセンサほど RSMA が大きく，学習収束後の学習効果に及ぼす影響が大きいことがわかる．またどのタスクについても，出力数 2 と 3 のセンサの RSMA の差に対して，出力数が 4 と 5 のセンサの RSMA の差が小さくなっていることがわかる．このことから，タスクの状態数にかかわらず，センサの出力数が増えると学習への影響は次第に増加しなくなっていくと言える．

さらに，Fig.8.2～Fig.8.4 において，出力数 2 のセンサの RSMA に注目する．どのタスクに関しても，出力数 2 のセンサの RSMA はおよそ 0.7 程度で収束していることがわかる．出力数が異なるセンサについても同様であり，タスクの状態数にかかわらず，出力数 3 のセンサは約 0.8，出力数 4 のセンサは約 0.86，出力数 5 のセンサは約 0.89 程度で学習が収束していることがわかる．

このように，タスクの状態数が変化した場合でも，センサが学習に及ぼす影響は全く変わらないことがわかる．このことから 5 章で示したセンサが学習に及ぼす影響は，ロ

ロボットが行うタスクの状態数に関わらずロボットに対して適用することができるといえる。

8.1.2. 考察

出力数の異なるセンサが学習効率に及ぼす影響の考察として、5章ではセンサの不足に対してロボットが認識可能な状態数の変化に注目した。ここでも、各タスクにおいて、各センサの不足による状態数の変化量に注目する。出力数 2,3,4,5 のセンサが不足したときに減少する状態数は、タスクの状態数に関わらず以下の通りになる。

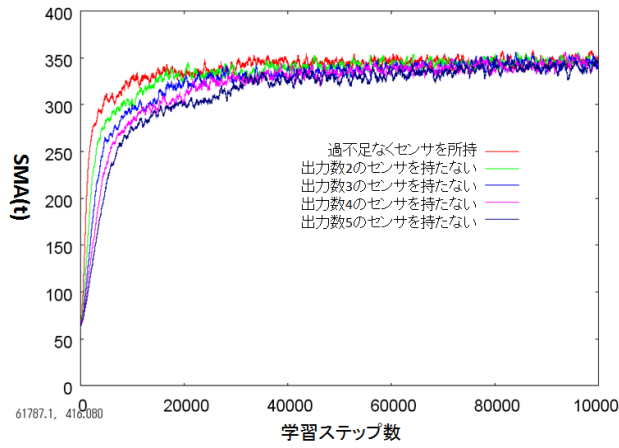
- ・出力数 2 のセンサが失われたとき：認識できる状態数は 50%減少
- ・出力数 3 のセンサが失われたとき：認識できる状態数は約 67%減少
- ・出力数 4 のセンサが失われたとき：認識できる状態数は約 75%減少
- ・出力数 5 のセンサが失われたとき：認識できる状態数は 80%減少

タスクの状態数が異なる場合でもセンサの出力数が同じであれば、センサが失われたときの状態数の変化割合は同じである。そのためタスクの状態数が異なる場合でも、センサが学習に及ぼす影響の大きさは変わらないのだと考えられる。

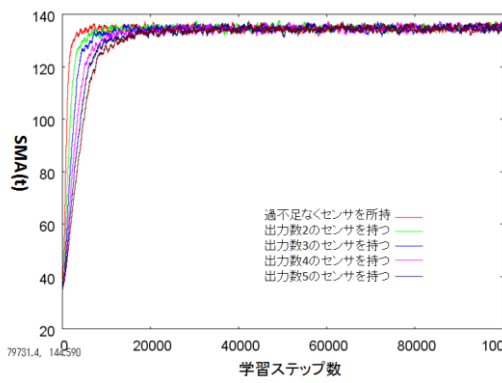
8.2.タスクの違いに対する不要センサの影響の実験

8.2.1.実験結果

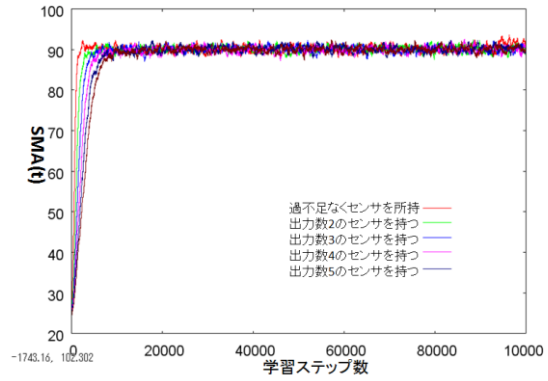
タスクの状態数が異なる場合に、不要センサが学習効率に及ぼす影響の実験結果を示す。実験結果を Fig.8.5, Fig.8.6, Table 8.1~8.3 に示す。Fig.8.6 は各タスクにおける各ロボットの SMA の推移のグラフである。Fig.5.3 の 80000 ステップ以降の SMA に注目すると、状態数 120 のタスクではいずれのロボットも SMA が 320 から 330 付近で安定して推移していることがわかる。同様に、状態数 60 のタスクではいずれのロボットも SMA が 130 付近、状態数 40 のタスクではいずれのロボットも SMA が 90 付近で推移していることがわかる。このことからタスクの状態数やセンサの出力数に関わらず、不要センサは学習収束後の獲得報酬に影響を及ぼさないことがわかる。



(a) 状態数 120 のタスク



(b) 状態数 60 のタスク



(c) 状態数 40 のタスク

Fig.8.5 各タスクにおける各ロボットの SMA の推移

Table 8.1 状態数 120 のタスクにおける各ロボットの時定数

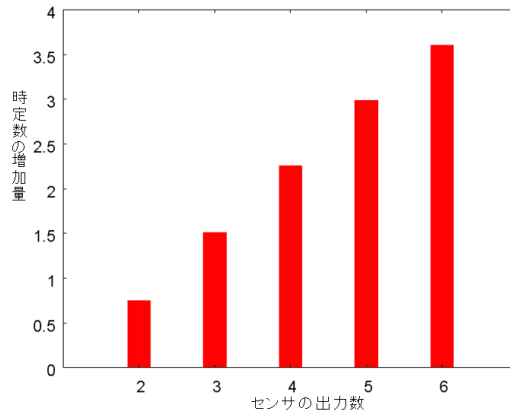
	過不足なし	出力数2の不要センサを所持	出力数3	出力数4	出力数5	出力数6
時定数 (ステップ数)	1750	3074	4403	5718	6994	8072
時定数の割合	1.0	1.752566	2.510262	3.259977	3.987457	4.602052
時定数の増加量	0	0.752566	1.510262	2.259977	2.987457	3.602052

Table 8.2 状態数 60 のタスクにおける各ロボットの時定数

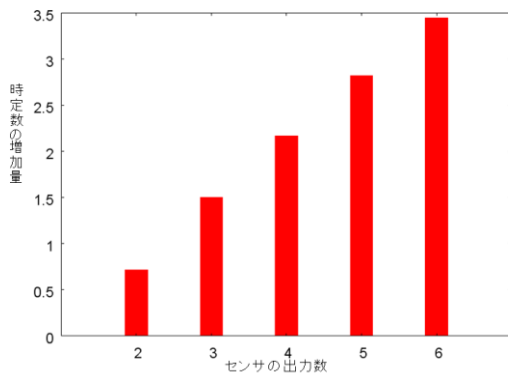
	過不足 なし	出力数 2 の 不要センサ を所持	出力数 3	出力数 4	出力数 5	出力数 6
時定数 (ステップ数)	1074	1842	2684	3404	4102	4778
時定数の割合	1.0	1.715084	2.499069	3.169460	3.819367	4.448790
時定数の増加量	0.0	0.715084	1.499069	2.169460	2.819367	3.448790

Table 8.3 状態数 40 のタスクにおける各ロボットの時定数

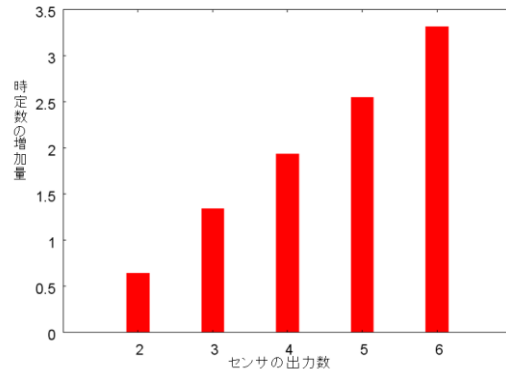
	過不足 なし	出力数 2 の 不要センサ を所持	出力数 3	出力数 4	出力数 5	出力数 6
時定数 (ステップ数)	848	1394	1984	2491	3011	3660
時定数の割合	1.0	1.643868	2.339623	2.937500	3.550708	4.316038
時定数の増加量	0.0	0.643868	1.339623	1.937500	2.550708	3.316038



(a) 状態数 120 のタスク



(b) 状態数 60 のタスク



(c) 状態数 40 のタスク

Fig.8.6 各タスクにおける，センサの出力数の違いによる時定数の増加量

Fig.8.5 の結果を元に各ロボットの時定数を算出し，タスク毎に表にまとめたのが Table 8.1～8.3 である．1 行目は各ロボットの時定数を示している．2 行目は，過不足のないロボットの時定数を 1 としたときの各ロボットの時定数の割合を示している．3 行目は，過不足のないロボットの時定数を 1 としたときの，各ロボットの時定数の増加量を示している．また，出力数の違いによる時定数の増加量を棒グラフで示したのが Fig.8.6 である．

Table 8.1～8.3 や Fig.8.5 における各ロボットの時定数の増加量に注目する．状態数 120 のタスクでは，センサの出力数に応じて時定数が約 75% ずつ増加している．また，状態数 60 のタスクではセンサの出力数に応じて時定数が約 71% ずつ，状態数 40 のタスクではセンサの出力数に応じて時定数が約 65% ずつ増加していることがわかる．このようにタスクの状態数にかかわらず，センサの出力数に対して時定数の増加量はほぼ比例の関係であることがわかる．増加量は出力数によって多少ばらつきが見られる．しかし今回用いているタスクはルート探索である上にルートの長さも不定なので，SMA を取ってい

てもばらつきが生じやすいタスクである。また、ロボットが用いている行動選択手法の ϵ -greedy 法も、ロボットの学習結果にかかわらず 10% の確率でランダムに移動するため、結果にばらつきを生む原因となっている。こうした要因を考えると、Table 8.1~8.3 に見られる時定数の増加量のばらつきは誤差の範囲内だと考えられる。以上のことから不要センサをエージェントに搭載した場合、タスクの状態数にかかわらず、センサの出力数に比例して学習速度が遅くなるといえる。

ただし、出力数が 1 増加したときの時定数の増加量はタスクによって異なり、状態数の少ないタスクほど、センサの出力数に対する時定数の増加量は小さくなっている。このことから、状態数の小さいタスクほど不要センサによる学習速度への影響は小さくなっていくことがわかる。

8.2.2. 実験結果の考察

ロボットが不要センサを所持すると、ロボットの認識する状態の数が増加する。ロボットは 1 ステップにつき一つの状態しか学習を行うことができない。そのため、増加した状態数に応じて学習にかかるステップ数が増加したのだと考えられる。ロボットが出力数 i の不要センサを所持すると、タスクの状態数にかかわらずロボットの状態数が i 倍に増える。つまり、不要センサによる状態数の増加量はセンサの出力数に比例する。そのためタスクの状態数にかかわらず、ロボットの学習にかかる時間がセンサの出力数に比例して増加したのだと考えられる。

5 章では得られなかった結果として、状態数の小さいタスクほどセンサの出力数に対する学習速度への影響が小さいという結果が得られた。これは、今回用いたタスクの設定が原因であると推測できる。今回のタスクは、ある状態 i から再び状態 i へ戻るための最短経路 D_{ii} を元に報酬が決定される。状態数が小さい場合には D_{ii} の最大値も小さくなるため、少ない探索でも高い報酬が得られるルートを探しやすくなる。一方状態数が大きい場合には D_{ii} の最大値も大きくなるため、高い報酬が得られるルートを探し出すことが困難になる。そのため不要センサによって状態数が増加したときに、高い報酬が得られるルートを探し出すことがより困難になり、状態数が大きいタスクほど不要センサの影響が大きくなったのだと考えられる。

ただし、今回の実験では検定などによる正確な検証は行っていない。そのためタスクの状態数の変化に対する獲得報酬量の変化は、偶然起きた事象である可能性もある。この点の真偽については今後の調査課題とする。

8.3. ロボットの所持センサの違いに関する実験結果

8.3.1. センサが不足しているロボットに対する必要センサの影響の実験結果

ロボットのセンサが不足している場合に、必要センサが学習効率に及ぼす影響の実験結果を示す。実験結果を Fig.8.7, Fig.8.8 に示す。Fig.8.7 は、出力数 2 のセンサが不足しているロボットに関して、各ロボットが獲得した SMA の推移である。Fig.8.7 は、20 万ステップまでに全てのロボットが学習収束していることを確認するためのグラフである。全てのロボットについて、学習終了前数万ステップに渡って同程度の SMA を獲得している。そのため全てのロボットが学習収束しているといえる。

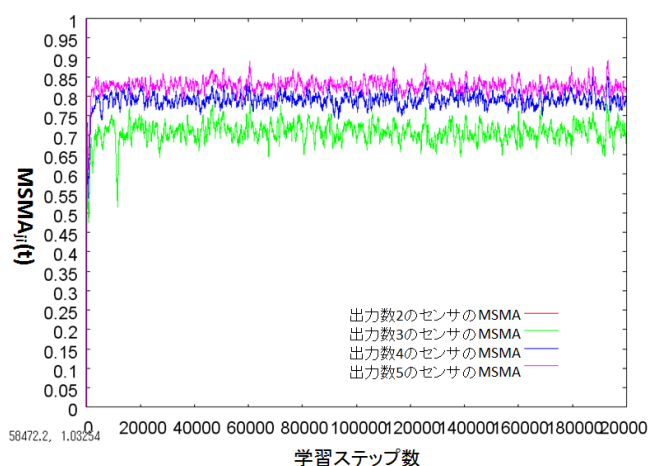


Fig.8.8 出力数 2 のセンサが不足しているロボットに対する
出力数の異なるセンサの MSMA の推移

Fig.8.8 は、出力数 2 のセンサが不足しているロボットに対する、出力数の異なるセンサの MSMA の推移である。このグラフは出力数 i のセンサによって変化した獲得報酬の割合の推移を表している。MSMA が大きいほど獲得報酬量の差が大きく、正の値であれば獲得報酬が減少したことになる。このグラフではいずれのセンサに関しても学習初期から最後まで正の値となっている。すなわち、ロボットのセンサが不足している場合でも、センサが不足すると学習全体を通して獲得報酬が減少することがわかる。

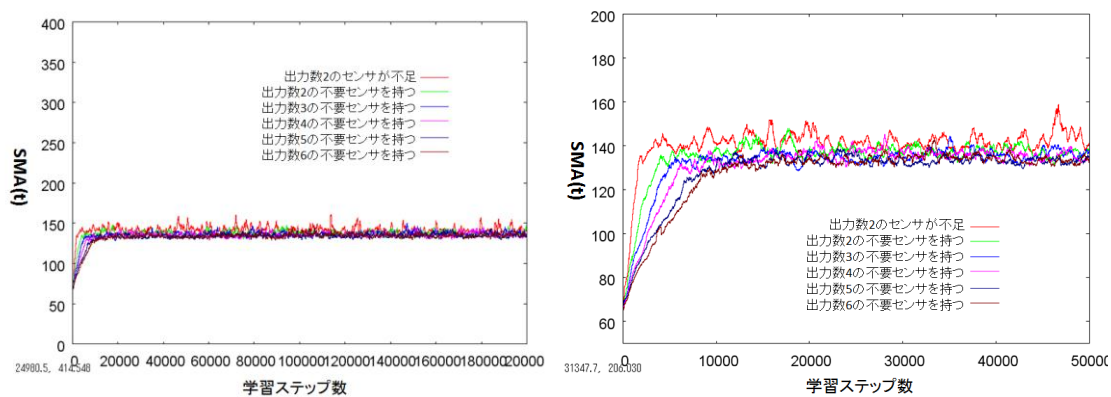
次に、各センサの学習収束後（2 万ステップ以降）の MSMA の大きさに注目する。すると 5 章での結果と同様に、出力数が大きいセンサほど MSMA が大きく、学習収束後の学習効果に及ぼす影響が大きいことがわかる。また、出力数 3 と 4 のセンサの MSMA の差に対して、出力数が 4 と 5 のセンサの MSMA の差が小さくなっていることがわかる。

このことから、ロボットのセンサが不足している場合でも、センサの出力数が増えると学習への影響は次第に増加しなくなっていくと言える。

このように、ロボットのセンサが不足している場合でも、センサが学習に及ぼす影響は全く変わらないことがわかる。このことから 5 章で示したセンサが学習に及ぼす影響は、センサが不足している場合ロボットに対しても適用することができるといえる。

8.3.2. センサが不足しているロボットに対する不要センサの影響の実験結果

ロボットのセンサが不足している場合に、不要センサが学習効率に及ぼす影響の実験結果を示す。実験結果を Fig.8.9, Fig.8.10, Table 8.4 に示す。Fig.8.9 は出力数 2 のセンサが不足しているロボットと、出力数 2 のセンサが不足しており、さらに不要センサを持つロボットの SMA の推移のグラフである。Fig.8.9 (a) は 20 万ステップまで、Fig.8.9 (b) は 5 万ステップまでの結果を示している。Fig.8.9 (b) の 20000 ステップ以降の SMA に注目すると、いずれのロボットも SMA が 130 付近で安定して推移していることがわかる。このことからセンサが不足している場合にも、不要センサは学習収束後の獲得報酬に影響を及ぼさないことがわかる。



(a) 200000 ステップまでの結果

(b) 50000 ステップまでの結果

Fig.8.9 出力数 2 のセンサが不足している各ロボットの SMA の推移

Table 8.4 出力数 2 のセンサが不足している各ロボットの時定数

	出力数 2 の センサが不足	出力数 2 の 不要センサ を所持	出力数 3	出力数 4	出力数 5	出力数 6
時定数 (ステップ数)	1549	2547	3650	4508	5576	6905
時定数の割合	1.0	1.644287	2.356359	2.910265	3.664300	4.457715
時定数の増加量	0.0	0.644287	1.356359	1.910265	2.664300	3.457715

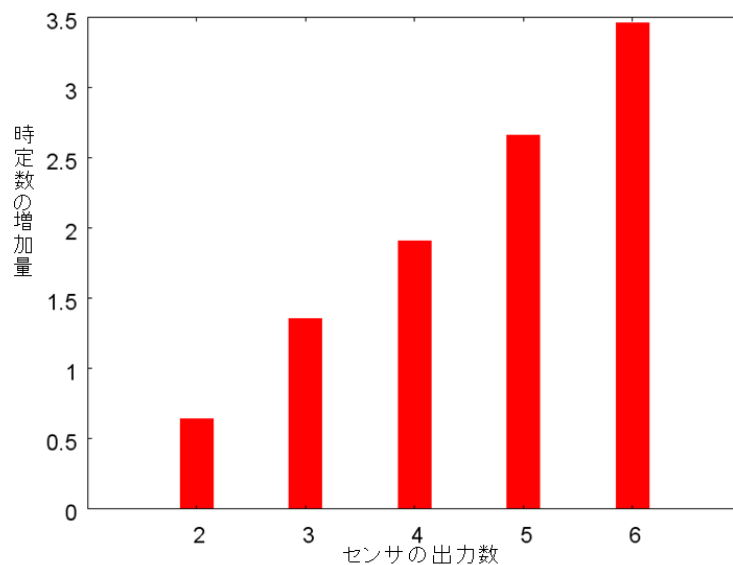


Fig.8.10 出力数 2 のセンサが不足しているロボットに対する
センサの出力数の違いによる時定数の増加量

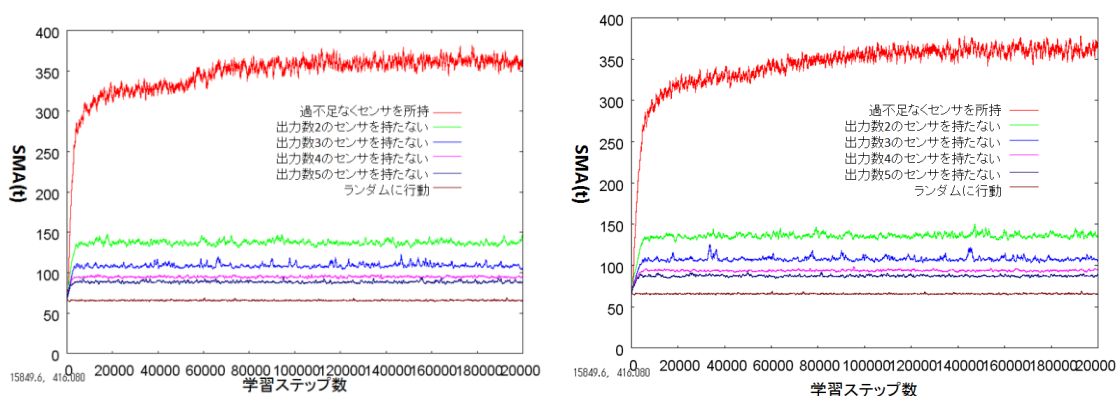
Fig.8.9 の結果を元に各ロボットの時定数を算出し、タスク毎に表にまとめたのが Table 8.4 である。1 行目は各ロボットの時定数を示している。2 行目は、不要センサを持たないロボットの時定数を 1 としたときの各ロボットの時定数の割合を示している。3 行目は、不要センサを持たないロボットの時定数を 1 としたときの、各ロボットの時定数の増加量を示している。また、出力数の違いによる時定数の増加量を棒グラフで示したのが Fig.8.10 である。

Table 8.4 や Fig.8.10 における各ロボットの時定数の増加量に注目する。すると、不要センサを持つロボットは、不要センサの出力数に応じて時定数が約 70% ずつ増加している。時定数の増加量は出力数によってばらつきがあり、センサを過不足なく持つロボットの結果にくらべて増加量のばらつきは大きくなっている。しかしこれは、センサが不足す

ることにより学習が不安定になっていることが原因だと考えられる。ロボットのセンサが不足するとロボットから見た環境が動的空間となるため、安定して報酬を得られることが難しくなる。そのためセンサを過不足なく所持するロボットに比べてランダムな要素による結果のばらつきが大きくなったのだと考えられる。こうした要因を考えると、Table 8.4に見られる時定数の増加量のばらつきは誤差の範囲内だと考えられる。以上のことから、センサが不足しているロボットに対して不要センサをエージェントに搭載した場合でも、センサの出力数に比例して学習速度が遅くなるといえる。

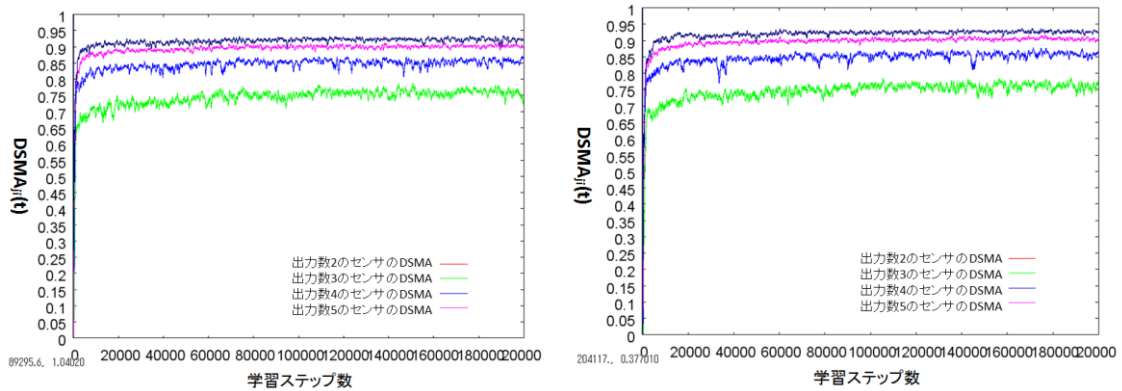
8.3.3. センサを過剰に所持するロボットに関する実験結果

ロボットが不要センサを所持している場合に、必要センサが学習効率に及ぼす影響の実験結果を示す。実験結果を Fig.8.11, Fig.8.12 に示す。Fig.8.11 は、不要センサを持つロボットと、不要センサを持ち、さらにセンサが不足しているロボットが獲得した SMA の推移である。Fig.8.11 (a) は全てのロボットが出力数 2 の不要センサを所持しており、Fig.8.11 (b) は全てのロボットが出力数 3 の不要センサを所持している。Fig.8.11 を見ると、全てのロボットについて、学習終了前数万ステップに渡って SMA の値が同程度の値で推移している。そのため全てのロボットが学習収束しているといえる。



(a) 出力数 2 の不要センサを持つロボット (b) 出力数 3 の不要センサを持つロボット

Fig.8.11 不要センサを持つ各ロボットの SMA の推移



(a) 出力数 2 の不要センサを持つロボット (b) 出力数 3 の不要センサを持つロボット
 Fig.8.11 不要センサを持つロボットに対する各センサの DSMA の推移

Fig.8.11 は、不要センサを所持しているロボットに対する、出力数の異なるセンサの DSMA の推移である。Fig.8.11 (a) は出力数 2 の不要センサを持つロボット、Fig.8.11 (b) は出力数 3 の不要センサを持つロボットに対する各センサの DSMA の推移である。これらのグラフは出力数 i のセンサによって変化した獲得報酬の割合の推移を表している。DSMA が大きいほど獲得報酬量の差が大きく、正の値であれば獲得報酬が減少したことになる。これらグラフではいずれのセンサに関しても学習初期から最後まで正の値となっている。すなわち、ロボットが不要センサを所持している場合でも、センサが不足すると学習全体を通して獲得報酬が減少することがわかる。

次に、各センサの学習収束後（20 万ステップ付近）の DSMA の大きさに注目する。すると Fig.8.11 (a) , (b) どちらも 5 章での結果と同様に、出力数が大きいセンサほど DSMA が大きく、学習収束後の学習効果に及ぼす影響が大きいことがわかる。また、出力数 2 と 3 のセンサの DSMA の差に対して、出力数が 4 と 5 のセンサの DSMA の差が小さくなっていることがわかる。このことから、ロボットのセンサが不足している場合でも、センサの出力数が増えると学習への影響は次第に増加しなくなっていくと言える。

このように、ロボットが不要センサを所持している場合でも、センサが学習に及ぼす影響は全く変わらないことがわかる。このことから 5 章で示したセンサが学習に及ぼす影響は、不要センサを所持しているロボットに対しても適用することができるといえる。

8.3.4.実験結果の考察

ロボットのセンサが異なる場合にも、必要センサや不要センサが学習効率に及ぼす影

響は全く変わらないことがわかった。これらの結果が生じたのは、センサによる状態数の変化割合が同じであることが原因であると考えられる。ロボットにセンサを搭載した場合や除去した場合、ロボットが認識できる状態数の変化割合はセンサの出力数にのみ依存する。例えば出力数 2 の必要センサをロボットから除去した場合、ロボットが認識できる状態数は必ず 50%減少する。出力数 2 の不要センサを搭載すれば、ロボットが認識できる状態数は必ず 2 倍になる。また、必要センサも不要センサも、ロボットの認識可能な状態数の変化割合の違いが原因となって、学習効率に様々な影響を及ぼしていた。こうしたことから、ロボットの認識可能な状態数にかかわらず、センサが学習効率に及ぼす影響はセンサの出力数、およびセンサがタスクに対して必要センサであるか不要センサであるかという要素のみによって決まったのだと言える。

8.3.5. 調査結果のまとめ

本章は、5 章までに行った実験を解決するための実験を行った。5 章までの実験では、実験結果を適用可能な範囲が非常に限られているという問題があった。そこで 6 章では、5 章の実験結果の適用範囲を広げるために必要な要素を考察した。その結果、タスクが異なる場合と、ロボットの所持センサが異なる場合にセンサが学習効率に及ぼす影響の変化を調査する必要があることを示した。そして 6 章にて、タスクや所持センサが異なる場合にセンサが学習効率に及ぼす影響の検証方法を述べ、この検証方法に従って調査実験を行った。

実験の結果、ほとんどの場合においてタスクの状態数や所持センサにかかわらず、センサが学習効率に及ぼす影響はセンサの出力数、およびセンサがタスクに対して必要センサであるか不要センサであるかという要素のみから決まることを示した。ただし唯一の例外として、不確定な結果ではあるが、タスクの状態数が異なる場合に不要センサが学習速度に及ぼす影響に変化が見られた。

ロボットに不要センサを搭載した場合、搭載した不要センサの出力数に比例してロボットの学習速度が減少する。この学習速度の減少量は、ロボットが行うタスクの状態数に応じて増減する可能性がある。ロボットの行うタスクの状態数が多い場合には不要センサによる学習速度の減少量も大きくなる。一方タスクの状態数が少ない場合には、不要センサによる学習速度の減少量も小さくなる。ただしこの影響は確実に生じる影響であるとは現時点では言えないため、今後の調査課題として挙げるに止める。

以上の結果から、出力数の異なるセンサが学習効率に及ぼす影響をまとめると以下のようになる。

- ・必要センサが学習効率に及ぼす影響

- ・必要センサが不足すると，学習開始時から終了時まで獲得報酬量が減少する
- ・センサの出力数の増加に伴い，学習収束後の獲得報酬の減少量が増加する
- ・センサの出力数が大きくなると，出力数の増加による獲得報酬の減少量が増えなくなる

- ・不要センサが学習効率に及ぼす影響
 - ・センサの出力数にかかわらず，学習収束後の獲得報酬は変化しない
 - ・センサの出力数が大きくなると，出力数の増加に比例して学習速度が遅くなる
 - ・ロボットが行うタスクの状態数が増えると，不要センサによる学習速度の減少量も大きくなる可能性がある

4章及び5章では調査結果を元に，ロボットに搭載するセンサの指針について述べた．ロボットのタスクの状態数や，ロボットの所持するセンサの過不足が異なる場合にも，センサが学習に及ぼす影響には基本的に変化がない．そのため，4章及び5章で示したセンサ搭載の指針についても，ロボットのタスクや所持センサの違いを気にせずに適用することが可能であるといえる．

9. まとめ

本論文ではロボットの学習におけるセンサの重要性を示すことを目的として，出力数の異なるセンサが学習効率に及ぼす影響の調査を行った．調査にはモデル化したセンサを利用し，シミュレーションによる調査実験を行った．また，調査の際にはセンサを必要センサと不要センサに分類し，それぞれのセンサについて出力数の違いと学習効率への影響について調査を行った．実験では任意の出力数 i のセンサを持つロボットと持たないロボットについて学習効率の違いを比較した．これにより出力数 i のセンサが学習効率に及ぼす影響を算出した．そして，出力数 i が変化したときにセンサが学習効率に及ぼす影響がどのように変化するのかを比較した．これにより，出力数の異なるセンサが学習効率に及ぼす影響を示した．

ここまでの調査を行ったのが1章から5章である．6章以降では5章までの調査の問題点として調査結果の適用範囲が限定的であることを挙げ，5章の実験結果の適用範囲を広げるために必要な要素を考察した．その結果，タスクが異なる場合と，ロボットの所持センサが異なる場合にセンサが学習効率に及ぼす影響の変化を調査する必要があること

を示した。そして 6 章にて、タスクや所持センサが異なる場合にセンサが学習効率に及ぼす影響の検証方法を述べ、この検証方法に従って調査実験を行った。

これらの調査の結果、センサの出力数に応じてロボットの学習効率に影響を及ぼす影響は変化することが判明した。センサが学習に及ぼす影響はロボットの所持センサに関係なく、センサの出力数のみから決定できることを示した。また、タスクの状態数に関しては不要センサが学習効率に及ぼす影響にのみ関係があることを示した。出力数の異なるセンサに関して、必要センサが学習効率に及ぼす影響と不要センサが学習効率に及ぼす影響をまとめると以下のようになった。

- ・必要センサが学習効率に及ぼす影響
 - ・必要センサが不足すると、学習開始時から終了時まで獲得報酬量が減少する
 - ・センサの出力数の増加に伴い、学習収束後の獲得報酬の減少量が増加する
 - ・センサの出力数が大きくなると、出力数の増加による獲得報酬の減少量が増えなくなる

- ・不要センサが学習効率に及ぼす影響
 - ・センサの出力数にかかわらず、学習収束後の獲得報酬は変化しない
 - ・センサの出力数が大きくなると、出力数の増加に比例して学習速度が遅くなる
 - ・ロボットが行うタスクの状態数が増えると、不要センサによる学習速度の減少量も大きくなる可能性がある

これらの調査結果を元に、センサの出力数を参考にしてロボットにセンサを搭載するための指針についても考察を行った。センサの出力数が重要となるのは、ロボットのタスクが複数の場合やタスクが複雑である場合に、タスクに対して必ずしも過不足なくセンサを搭載できない場合である。必要センサも不要センサも、センサの出力数に応じて学習への影響が増加する。ただし必要センサの場合には、センサの出力数が増えていくと、学習への影響はある程度の大きさ以上には増えないという特徴がある。一方不要センサについてはセンサの出力数に単純に比例してロボットの学習速度が遅くなるという特徴がある。こうした特徴を踏まえて搭載するセンサ、除去するセンサを選択することでロボットに対してより適切にセンサを搭載することができる。

このように、タスクに対して過不足なくセンサを搭載するのが難しい場合であっても、センサの出力数に応じて搭載するセンサを選択することで学習効率の低減を防ぐことができることを示した。これによりロボットの学習においては、学習手法だけでなくセンサが学習効率に及ぼす影響を考慮することが非常に重要であることを示すことができた。

最後に、本論文での調査結果を踏まえて今後の研究課題について述べる。本論文ではセンサの能力として、ロボットの学習において最も重要であるセンサの出力数に注目して調査を行った。しかしセンサの能力を決める要素は他にもいくつか存在する。その中でも最も重要なのが、センサの分解能とサンプリング周波数である。

センサの分解能は、センサが認識可能な値を何段階に分割するかを示すセンサの能力である。分解能が大きいセンサはより細かく状態を認識することができるが、タスクに対して必要以上に状態の切り分けを行ってしまう可能性が出てくる。一方分解能が小さいセンサはより粗く状態を認識するため不要な情報は少なくなるが、必要な状態の切り分けも行えなくなってしまう恐れがある。これまではセンサを必要センサと不要センサに分類していたが、分解能を考慮すると、一つのセンサの中にタスクに必要な状態と不要な状態が含まれるという状況が起こりうる。こうした分解能が学習に及ぼす影響も調査を行う必要がある。

センサのサンプリング周波数は、センサが状態を認識する周期を決めるセンサの能力である。サンプリング周波数が荒いセンサでは、ある時間内で状態を認識できる回数が減少する。状態を認識してから次に状態を認識するまでにロボットの状態遷移が複数回起こった場合、ロボット遷移する途中の様子を認識できなくなる。このような場合に学習にどのような影響が生じるのか調査を行う必要がある。

センサの能力として特に重要なのはこの二点である。しかしロボットの学習においてセンサに注目した背景には、ロボットの身体に注目したという背景もある。ロボットは身体を持つ学習者であるため、学習における状態認識はセンサを通して行い、学習結果を元に決定した行動はアクチュエータを利用して行われる。このときアクチュエータの性能が悪いと、ロボットが決定した行動を正しく行えない可能性がある。このとき、ロボットは学習をうまく行えないことが予想される。そのためセンサだけではなく、ロボットのアクチュエータについても学習に対して適切に設計されている必要がある。

以上、センサの分解能、センサのサンプリング周波数、そしてロボットのアクチュエータが学習に及ぼす影響の調査を今後の研究課題とする。これらの調査を行っていくことで、学習手法以外の要因からロボットの学習効率を改善できるようにしたい。また、ロボットの学習をより効率的に行う方法を模索していきたい。

謝辞

本論文を結ぶにあたり，日ごろから様々な面で有益なご指導・ご助言を頂きました主指導教員の倉重健太郎先生に深く感謝の意を表します．また，中間発表にてご指導・ご助言を下さいました本学情報工学科の・・・先生，・・・，にこの場をお借りして厚くお礼を申し上げます．また，研究活動全般において有意義なディスカッションや意見を頂いた認知ロボティクス研究室の木島康隆さん，宮崎愛央君，北山直樹君，梅津祐介君，渋谷和さん，杉本大志君，高泉昇太郎君，沼田利伸君，三浦丈典君に心より感謝致します．

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] 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

- [4] 谷口, 榎木: “報酬設計を通じた社会的相互作用による行為概念群の構築: シェマ理論に基づいた累増的強化学習”, 知能と情報, vol. 18, no. 4, pp. 629--640, 2006.

- [5] 中須賀: “人工知能は宇宙開発・宇宙利用に貢献できるか?”, 人工知能学会誌, vol. 21, no. 1, pp. 2--13, 2006.

- [6] 三宅陽一郎, “デジタルゲームにおける人工知能技術の応用”, 人工知能学会誌, Vol23, No1, pp44-55, 2008

- [7] 中須賀: “人工知能は宇宙開発・宇宙利用に貢献できるか?”, 人工知能学会誌, vol. 21, no. 1, pp. 2-13, 2006.

- [8] 浅田: “強化学習の実ロボットへの応用とその課題”, 人工知能学会誌, vol. 12, no. 6, pp. 831--836, 1997.

- [9] 細田: “形態が学習にもたらすもの, 学習が形態にもたらすもの”, 日本ロボット学会誌, vol. 22, no. 2, pp. 186--189, 2004.

- [10] 浅田: “認知発達ロボティクスにおける学習”, 人工知能学会誌, vol. 18, no. 5, pp. 550--554, 2003.

- [11] 伍賀, 妻屋, 田浦: “身体性の拡張に関連する進化型自立移動ロボットの特徴解析”, 2008年度人工知能学会全国大会, pp. 4, 2008.
- [12] 浅田稔, 石黒浩, 國吉康夫, “認知ロボティクスを目指すもの”, 日本ロボット学会誌, Vol17, No1, pp2-6, 1999
- [13] 中村, 松本, 石黒: “生態ゆらぎを模倣したロボット制御”, 日本ロボット学会誌, vol. 28, no. 4, pp. 470--478, 2010.
- [14] 伊藤, 高山: “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律ヘビ型”, 知能と情報, vol. 21, no. 3, pp. 402--403, 2009.
- [15] 山村, 馬野, 瀬田: “段階的な視覚をも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, vol. 18, no. 4, pp. 561--570, 2006.
- [16] 杉浦, 川上, 片井: “移動ロボットにおけるセンサ形態と制御系の同時設計法”, 電学論C, vol. 128, no. 7, pp. 1154--1161, 2008.
- [17] Y. Onoue, K. Kurashige, “A relationship between ability of perception and learning efficiency,” Proceeding of World Automation Congress 2008, 2008.
- [18] 高橋, 浅田: “実ロボットによる行動学習のための状態空間の漸次的構成”, 日本ロボット学会誌, vol. 17, no. 1, pp. 118--124, 1999.
- [19] 釜谷, 北山, 藤村, 阿部: “連続状態空間における強化学習”, 計測自動制御学会東北支部 第229回研究集会, pp. 229-11, 2006.
- [20] Richard. S. Sutton, Andrew. G. Barto: Reinforcement Learning. pp. , The MIT Press, 1998.
- [21] 中南: “環境認識能力の変化が学習に及ぼす影響について”, 室蘭工業大学卒業論文, 2010.

研究業績

[1] 中南：“センサ数の過不足が学習に及ぼす影響について”，精密工学会 2011 年度北海道支部学術講演会，pp. 1-2, 2010.

センサ数の過不足が学習に及ぼす影響について

室蘭工業大学 ○中南義典, 室蘭工業大学 倉重健太郎, 株式会社本田技術研究所 尾上由希子

要旨

ロボットはセンサの認識情報を用いて学習を行う。そのためロボットが持つセンサが異なると学習効率も異なると考えられる。しかし学習手法の有用性を検証する際、センサ数に過不足が生じたまま検証が行われている可能性がある。このときセンサ数の過不足により学習効率に変化し、検証が適切に行われていない可能性がある。本論文ではセンサ数の過不足が学習に及ぼす影響を調査し、学習に合わせたセンサ選択の重要性を示す。

1. はじめに

ロボットが学習を行う時、自身に搭載されたセンサから得られた情報を元に自身の状態を認識して学習を行う。そのため、タスクに対してセンサ数が不足しているロボットは適切な学習を行うことが難しい。一方、ロボットに多くのセンサを搭載してセンサから得られる情報量が増えると、より多くの情報に対して学習することになり学習にかかる時間が増えてしまう。このように、ロボットの学習においてはセンサ数の過不足が学習効率に大きく影響を与えると考えられる[1, 2]。

従来研究においては、ロボットに与えるタスクに対して必要と考えられるセンサを研究者が予測し、ロボットに搭載するセンサを選択している[3]。しかし、ロボットが実際に過不足なくセンサを搭載していることを判断する明確な指標は無い。つまり、ロボットのセンサ数に過不足が生じたまま研究が行われている可能性がある。このときも、センサ数の過不足によって学習効率に変化が生じるならば、その影響が見過ごされてしまう。このような状況では、学習手法の有用性を検証する際にセンサ数の過不足により本来の学習効率が発揮されず、学習手法が本来持つ有用性が見過ごされてしまう可能性がある。

そこで本論文では、ロボットのセンサ数の過不足が学習に及ぼす影響について、強化学習[4]を対象として調査を行う。この調査により、ロボットの学習において学習に合わせたセンサを搭載することの重要性を示す。

2. 調査対象となるセンサの定義

本論文ではいかなるセンサにも共通して起こる学習への影響を調査する。そのために、センサの基本的な機能について抜き出して抽象化したセンサを定義する。この抽象化したセンサに対してシミュレーション実験を行うことで、センサが学習に及ぼす影響を調査する。

まず、センサの基本的な役割について考える。あらゆるセンサに共通する役割は、対象の状態を認識し、認識した状態を離散値に変換して出力することである。そこで本論文では、認識した状態を式(1)に示す離散値として出力するものをセンサとして定義する。

$$V = \{1, 2, \dots, V_{\max}\} \quad (1)$$

エージェントはこのセンサを一つ、または複数用いて状態を認識する。エージェントは、自身が持つ各センサから出力された値の組み合わせによって自身の状態を決定する。そのため、エージェントが認識する状態空間は自身が持つセンサの組み合わせによって決まる。

次に、エージェントが持つセンサの過不足について考える。エージェントが必要とするセンサは、エージェントに与えるタスクによって決まる。そこで、エージェントに与えられたタスクの達成に必要なセンサを必要センサと呼

び、タスクに対して必要センサが少ないエージェントはセンサ数が不足しているとする。一方、エージェントに与えるタスクが決まれば、タスク達成には必要のないセンサも同時に決まる。このセンサを不用センサと呼び、不用センサを持つエージェントは過剰にセンサを所持していると言える。本論文では強化学習を対象として調査を行うことから、必要センサと不用センサを次のように定義する。

- 必要センサ：タスクに与えられた報酬に対して従属である値を出力するセンサ
- 不用センサ：タスクに与えられた報酬に対して独立である値を出力するセンサ

3. センサの過不足が学習に及ぼす影響の検証方法

本論文では、センサ数の過不足が学習に及ぼす影響を調査する。そのために、センサ数の異なるエージェントに同じタスクを与え、同じ学習手法を用いて学習させ、各エージェントの学習効率を比較する。

初めにセンサ数の異なるエージェントを用意する。まず、必要センサを持たず、0~M個の不用センサを持つエージェントを用意する。同様に、1~N個のセンサを持ち、0~M個の不用センサを持つエージェントを用意する。このとき、センサの数以外の要素が学習に影響を及ぼさないように、センサの能力を均一化する必要がある。そのために各センサが出力する値Vの最大値 V_{\max} を共通の値にする。

次に、エージェントが用いる学習手法について述べる。学習手法の違いによって学習の影響が変化することを防ぐため、全てのエージェントは強化学習を用いて学習を行う。行動学習手法にはQ学習、行動選択手法にはsoftmax法を用いる。Q学習の式を(2)に、softmax法の式を(3)に示す。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (2)$$

ここで $Q(s_t, a_t)$ は状態 s_t における行動 a_t のQ値を示している。また、 α は学習率、 γ は割引率を表している。

$$\pi(s_t, a_t) = \frac{e^{Q(s_t, a_t)/\tau}}{\sum_{b=1}^n e^{Q(s_t, b)/\tau}} \quad (3)$$

ここで $\pi(s_t, a_t)$ は状態 s_t における行動 a_t の選択確率である。また、 τ は温度と呼ばれる正定数である。

次に、エージェントが行うタスクについて述べる。エージェントに与えるタスクはルート探索のタスクである。タスクは複数の状態を持ち、各状態は二つの遷移先を持つ。エージェントはN個の必要センサを用いることでタスク内のすべての状態を認識できる。また、エージェントは各状態において行動を選択することで、行動に対応した遷移先の状態へ確定的に遷移する。そして、遷移先の状態に応じてエージェントは報酬を獲得する。エージェントの目的は、このタスクにおいて最も高い報酬を得られ続けるルー

トを探索することである。このとき、高い報酬を得られる状態ほど辿り着くのが困難であるように報酬を設定した。各状態における報酬設定の式を(4)に示す。

$$Rwd_i = e^{\beta \cdot D_i} \quad (4)$$

ここで D_{ii} は状態 i から遷移して状態 i へ再び辿り着くまでに経由する最小の状態数である。 D_{ii} が大きい状態ほど高い報酬が設定される。また、 β は \exp の値を調整するための D_{ii} の係数である。

このタスクに対して各エージェントは R ステップの学習を行う。このときランダムな要素によって特定のエージェントに有利な状況が生じるのを防ぐため、 E 個のタスクに対して学習を行わせる。そして各タスクで得られた獲得報酬の平均から各エージェントの学習効率の違いを比較する。

次に、各エージェントの学習効率の比較方法について述べる。今回用いるタスクでは、より多くの報酬を獲得しているエージェントほど学習効率が高いと言える。そこで各エージェントが各ステップで獲得した報酬の推移を比較する。これにより各エージェントの学習速度の違いや、学習収束後の学習効率の違いについて比較する。しかし、今回用いるタスクでは高い報酬を得るために報酬の低い状態を経由する必要があるため、1ステップ毎の獲得報酬が安定しない。そこで t ステップ目における直近の L ステップで得た報酬の平均の推移を用いて比較を行う。この式を(5)に示す。

$$SMA = \frac{r_t + r_{t-1} + \dots + r_{t-L+1}}{L} \quad (5)$$

ここで r_t はエージェントが t ステップ目で得た報酬の値である。

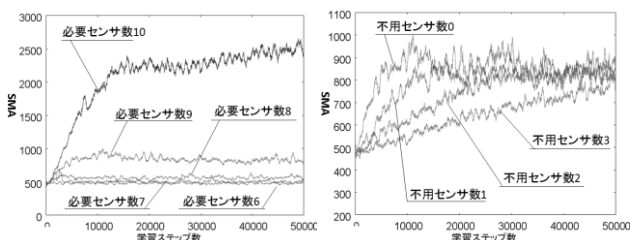
最後に、本実験で実際に用いたパラメータを表1に示す。

表1. 実験パラメータ

N(必要センサの最大数)	10	β	0.65
M(不用センサの最大数)	3	α	0.7
R(学習ステップ数)	50000	γ	0.6
E(実験を行うタスクの数)	20	τ	1000
V_{max} (センサの出力の最大値)	2	L	500

4. センサの過不足が学習に及ぼす影響の実験結果

初めに、エージェントが持つ必要センサ数が不足している場合の学習への影響を示す。不用センサ数0、必要センサ数6~10のエージェントのSMAの推移を図1(a)に示す。



(a) 必要センサ6~10、不用センサ0のエージェントのSMAの推移
(b) 必要センサ9、不用センサ0~3のエージェントのSMAの推移
図1. 各エージェントのSMAの推移の比較

図1(a)において、必要センサ数0~7のエージェントはSMAの推移の様子がほとんど同じであったので、必要センサ数5以下のエージェントのSMAについては省略した。各エージェントの学習収束後のSMAに注目すると、必要センサ数が減っていくにつれてSMAの値が小さくなっている。このことから、タスクに対して必要センサが不足している

エージェントは学習収束後の学習効率が減少すると言える。

センサ数が不足しているエージェントは、本来切り分けるべき状態を一つと同じ状態として認識してしまう。つまり不適切な状態空間に対して学習を行うことになる。そのため、適切に学習を行っているにもかかわらず学習の対象が不適切であるため、学習効率が減少するのだと考えられる。

次に、エージェントが不用センサを持つ場合の学習への影響を示す。必要センサ数9、不用センサ数0~3のエージェントのSMAの推移を図1(b)に示す。各エージェントの学習収束のタイミングに注目すると、いずれのエージェントもSMAが900程度に到達した時点で学習が収束している。また、学習収束までのステップ数は不用センサの数に応じて増加している。今回は必要センサ数が9のエージェントの比較結果しか示していない。しかし必要センサ数が同じエージェント同士で不用センサの影響を比較した場合、不用センサの影響は全て同じ傾向であった。これらのことから不用センサを持つエージェントは学習収束に必要なステップ数が増加するといえる。また、学習収束後は不用センサを持たないエージェントと同じ学習効率を發揮できるといえる。

5. まとめ

本論文ではセンサ数の過不足が学習に及ぼす影響の調査を行った。調査は抽象化したセンサを対象とし、シミュレーションによって行った。実験ではセンサ数の異なるエージェントを複数用意し、各エージェントに同じ学習手法と同じタスクを与えて学習効率の違いを比較した。

調査の結果、エージェントが所持するセンサがタスクに対して不足すると学習収束後の学習効率が減少することを示した。また、エージェントが不用センサを所持すると、学習収束に必要なステップ数が増加することを示した。

初めに述べたように、現在のロボットの学習の研究ではセンサ数に過不足が生じたまま研究が行われている可能性がある。そこで本論文では、センサ数の過不足が学習に及ぼす様々な影響を調査した。その結果、センサ数に過不足が生じている場合、学習に様々な影響を及ぼすことを示した。そのため、センサの過不足が生じたまま学習手法の有用性の検証を行った場合に、学習手法が持つ本来の有用性が発揮できず、有用性の検証が適切に行えないと言える。以上のことから、ロボットの学習においては学習に合わせたセンサを搭載することが重要であることを示した。

今回の調査ではセンサ数の違いが学習に及ぼす影響に注目した。そのため各センサの能力は均一のものとして調査を行った。そこで今後の課題として、各センサの能力が異なる場合に学習に及ぼす影響について調査を行いたい。これによりロボットの学習におけるセンサの重要性をさらに示していきたい。

参考文献

- [1] Y. Onoue, K. Kurashige, "A relationship between ability of perception and learning efficiency", Proceeding of World Automation Congress 2008, 2008
- [2] 杉浦, 川上, 片井: "移動ロボットにおけるセンサ形態と制御系の同時設計法", 電学論C, vol. 128, no. 7, pp. 1-8, 2008.
- [3] 細田: "形態が学習にもたらすもの, 学習が形態にもたらすもの", 日本ロボット学会誌, vol. 22, no. 2, pp. 186-189, 2004.
- [4] R. S. Sutton, A. G. Barto, "Reinforcement learning", The MIT Press, 1998