

修士論文

自律ロボットのための自我を含む
多重意識エージェントの相互学習

平成22年度

三重大学大学院 工学研究科 機械工学専攻

森 敦 紀

平成 22 年度 修士論文

自律ロボットのための自我を含む 多重意識エージェントの相互学習

専攻 三重大学大学院 工学研究科 機械工学専攻
研究室 メカトロニクス研究室

平成 21 年度入学 409M151

氏名 森 敦 紀

目次

1	緒言	1
2	提案手法	3
2.1	意識と自我の定義	3
2.2	意識と自我のモデル化	4
2.3	アルゴリズム	6
2.4	他手法との比較	7
3	実験	8
3.1	2つのエージェント間での学習	9
3.1.1	画像処理エージェントにとって認識が容易な状況での学習	10
3.1.2	画像処理エージェントにとって認識が困難な状況における学習	12
3.1.3	実験結果	13
3.2	複数のエージェント間の学習	15
3.2.1	電池交換	15
3.2.2	エージェント構成	16
3.2.3	電池交換タスク中の学習	17
3.2.4	実験結果	21
4	結言	24

目次	ii
参考文献	25
謝辭	26

目 次

2.1	Proposed model	4
2.2	Agent competition	5
2.3	Learning model	6
2.4	Proposal algorithm	6
3.1	Autonomous mobile robot MieC	8
3.2	Agents for experiment1	9
3.3	Experiment 1-1	10
3.4	Experiment 1-2	11
3.5	Experiment 1-3	11
3.6	Experiment 1-4	11
3.7	Experiment 1-5	12
3.8	Experiment 1-6	12
3.9	Experiment 1-7	13
3.10	Experiment 1-8	13
3.11	Environment of experiment1	14
3.12	Result of agent ' s position in experiment1	14
3.13	Result of a agent ' s state in experiment1	15
3.14	Battery box	16

3.15 Battery exchange mechanism	16
3.16 Agents for experiment2	17
3.17 experiment2-1	18
3.18 experiment2-2	18
3.19 experiment2-3	19
3.20 experiment2-4	19
3.21 experiment2-5	19
3.22 experiment2-6	20
3.23 experiment2-7	20
3.24 experiment2-8	20
3.25 experiment2-9	21
3.26 experiment2-10	21
3.27 Environment of experiment	22
3.28 Result of agent ' s position in experiment2	22
3.29 Result of a agent ' s state in experiment2	23

第1章

緒言

本論文では、知能ロボット制御における、人の自我を含む意識のモデルに基づくマルチエージェント学習システムを提案する。人の自我を含む意識の役割をモデル化することで、学習能力の高いロボットの実現を試みる。

掃除ロボットや警備ロボットなど、自律ロボットの活躍は広がっている。そのため、ロボットのさらなる知能化が求められている。ここでは、人の知能的な学習過程に着目し、人の高度な学習機能を模倣するにあたり、個々の意識や自我について考える。まず、意識と自我について考察する。現在、意識や自我の仕組みは未だに解明されてない。その上、分野ごとに定義が異なる。例えば、意識と無意識の各分野における定義は以下のものである。

1. 心理学：体の中に「意識」は複数存在する。「無意識」は自我ののっていない「意識」である。
2. 哲学：「意識」は認識する能力。「無意識」は自分を客観的に認識する能力である。
3. 医学：「意識」とは脳の働きが活性化し、五感に対する刺激を感じ取ることが可能な状態であり、「無意識」はそれらができない状態である。
4. ウパニシャッド哲学：「意識」には4つの状態がある。そのうちの1つが「無意識」で

ある.

その他にも異なる分野で異なる定義があり，工学の分野でも，それぞれの工学者が考える自我や意識の定義がある．Minsky[1] は，階層化された心のエージェントの相互作用により，人の知的な行動が実現していると述べている．また，Maeno[2] は，エピソード記憶をする働きのある自我により制御されている意識モデルを提案している．しかし，自我について一般的なモデルはなく，現在も多くは研究されていない．また，実際にこれらのモデルをロボットの学習に用いたものは少ない．

本論文では，ドイツの心理学者 Wundt の考えを取り上げる．彼は人の意識を内観で分析し，自我は意識の主体であると考えた．そこで，自我を含む意識をマルチエージェントシステムと考え定義し，機械学習に適用することを試みる．本研究は，それぞれの意識をエージェントとして扱い，自我ののる意識とその他の意識との相互作用により，自律ロボットが学習するシステムを提案する．機械学習は大きく分けて，教師あり学習，教師なし学習，強化学習に分類される．マルチエージェント強化学習の従来研究としては，以下のようなものがある．Wiering[3], Osada[4] は，エージェントを階層的に配置し各階層にサブゴールを設置することで学習効率を向上させている．Ono[5], Fujita[6] は，Q-learning の状態空間を自分と他の 1 体のエージェントから構成させる部分状態を用い，学習時に範囲を限定することで効率の良い学習をする．本研究では教示情報としてシステム外部からだけでなく，内部（システム内の他のエージェント）からの情報も教示として用いる設計とする．そのため，従来手法と比べ，より多くの教示情報を用いるため，効率のよい学習が期待できる．また，環境の変化に応じて学習エージェントと教示エージェントが役割を換えて学習するようにするため，従来手法と比べより柔軟な学習が期待できる．

本論文では，提案手法が実際のロボットに効果的であることを実験により確認する．

第2章

提案手法

2.1 意識と自我の定義

本論文では自我を含む意識の役割を以下のように定義し，モデル化する．そのモデルは，主観的な内観により考え，客観的な解剖学に反しないものとする．

1. 意識は一つではなく，体の中に複数存在する．
2. 人の動作は多くの意識の協力と競争で引き起こされる．
3. 自我は一つだけ存在する．
4. 自我は最も優先度の高い意識に存在する．
5. 意識が対立すると，自我が乗った意識は優先的に資源を使用し，他の意識は停止する．
6. 学習する場合，自我の乗った意識が学習する．
7. 他の意識の認識結果が，他の意識の学習に用いられる．

2.2 意識と自我のモデル化

本論文では，上記の意識の役割をマルチエージェントシステムとして扱い (**Fig. 2.1**) 以下のようにモデル化した．図中の円はすべてエージェントを表す．

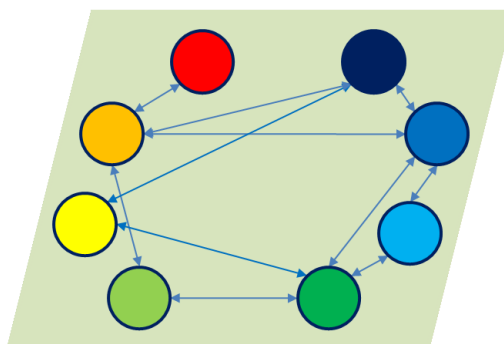
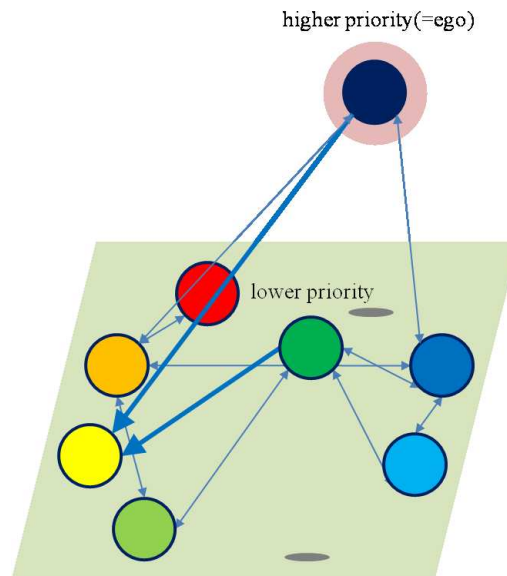


Fig. 2.1 Proposed model

1. 人のそれぞれの意識に対応するデータプロセッサかセンサそれぞれをエージェントとする．
2. エージェントはそれぞれ優先度と起動条件をもつ．
3. 環境が起動条件を満たせばエージェントは処理を実行する．
4. 優先度は設定される条件に合えば合うほど強くなる．
5. エージェントは他のエージェントから指令と優先度を取得する．
6. エージェントは複数のエージェントから競合する指令を受け取ると，優先度の高い指令のみを用いる． (**Fig. 2.2** の場合指令を受け取ったエージェントは，優先度の高いエージェントからの指令のみ実行し，優先度が低いエージェントからの指令は無視する．)

**Fig. 2.2** Agent competition

7. 学習する場合，優先度の高いエージェントが学習する．
8. 他のエージェントの認識結果が，他のエージェントの学習に用いられる．

エージェントの優先度は環境に応じて変化するため，学習エージェントと教示エージェントが入れ替わりながら学習する． (**Fig. 2.3**)

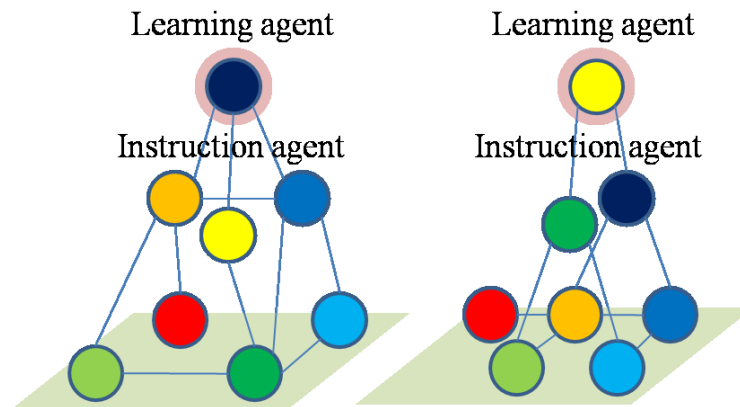


Fig. 2.3 Learning model

2.3 アルゴリズム

データプロセッサ型エージェントのアルゴリズムを Fig. 2.4 に示す.

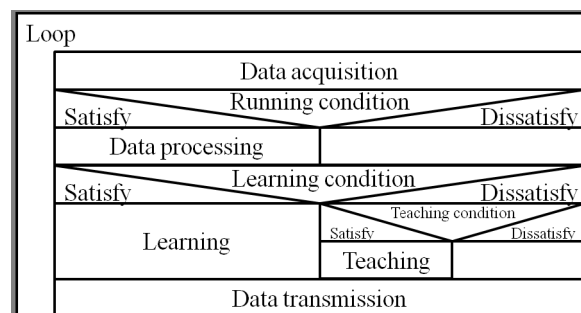


Fig. 2.4 Proposal algorism

まずエージェントは他のエージェントから指令やセンサデータを取得する。得られたデー

タを処理し，自身の起動条件から優先度を計算する．今回はロボットが上手く動作するように優先度の変化量は人が恣意的に設計したが，将来的には優先度の変化量も自動的に学習させる．学習状態に移行する条件を満たしていれば学習する．教示状態に移行する条件を満たしていれば教示する．最後に他のエージェントにデータを送信する．センサ型エージェントの処理はセンサからのデータを他のエージェントに渡すのみである．

2.4 他手法との比較

提案モデルと同じ人を模倣したモデルとしてニューラルネットワークモデルがある．その中でもスパイクニューラルネットワークモデルは時間遅れまでをも表現できるため，理論的には任意の非線形制御関数を表現できる．しかし，欠点として学習をしようとするとき，複数のパラメータがお互いに干渉し合い学習が困難になる．提案モデルは学習するエージェントが常に 1 つに限定されるため学習が安定的である．

本手法は，強化学習の一種であるが，従来手法とは異なり，教示情報がシステムの外のみでなく，システム内部からのものもあるので学習が効率よく進む．また，学習すべきエージェントは，環境の変化により入れ替わり学習が進むので，柔軟である．

第3章

実験

本論文では，提案方法が効率的であることを実験的に確かめる．実験には当研究室が設計開発した実際のロボット MieC(Movable Intelligent Easy Computer) を使用する．(Fig. 3.1)

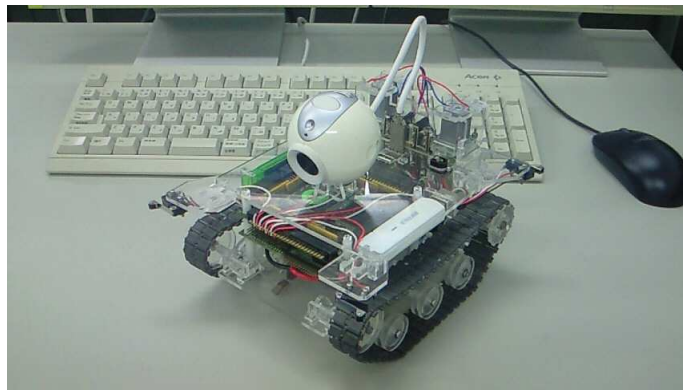


Fig. 3.1 Autonomous mobile robot MieC

MieC は長さ 21cm，幅 16.5cm，高さ 16.5cm であり，電池ボックス(単 3 電池 6 本) で起動し，左右 2 つのモータにより駆動するクローラーで移動する．OS として Linux を採用し，周辺機器は，視覚センサとしての WebCam(自由度 0) や無線 LAN，USB メモリを用いてい

る。また、本実験に使う機器として、前後左右 4 箇所にタッチセンサを配置した。これにより MieC は、WebCam による視覚センサとタッチセンサによる触覚センサ、そしてモータに付くエンコーダから環境情報を得る。リアルタイム制御が必要なモータ関係は FPGA ボードで制御し、画像処理などのエージェントの処理は CPU ボード上で行う。実験環境では、MieC は垂直な面に囲まれた水平面上に存在し、垂直面と水平面の間に境界が存在する。

今回は、2 つのエージェント間の学習 (3-1) と複数のエージェント間の学習 (3-2) で実験する。

3.1 2 つのエージェント間での学習

ロボットのための教示と学習のためのエージェントとして、画像処理エージェントと体感エージェントを取り上げる。エージェント全体の構成を Fig. 3.2 に示す。

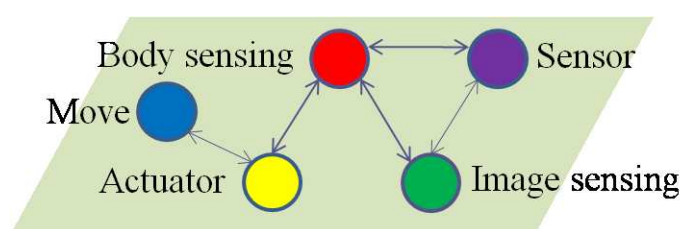


Fig. 3.2 Agents for experiment1

画像処理エージェントと体感エージェントは境界からのロボットの距離をそれぞれ探す。画像処理エージェントは適当な閾値によって画像上の境界を探索し、画像座標からロボット座標に変換することで境界の 3 次元姿勢を概算する。体感エージェントはもしロボットが面に触れたら、垂直面の下にある境界を探索する。画像処理エージェントとは異なり、体感エージェントは適当なパラメータによるデッドレコニングによって境界とロボットとの距離を概算する。互いのエージェントの境界の姿勢の認識は正確ではない場合、優先度は高く

なる。

上記 2 つのエージェントは能動的に移動するエージェントではないため、実験するに当たり環境を変化させるために移動エージェントを追加した。移動エージェントの優先度は一定であり、学習も教示もしない人間が操作したようにロボットを動かすエージェントである。センサーエージェントはカメラセンサ、タッチセンサ、エンコーダからデータをそれぞれ取得するエージェントが存在するが **Fig. 3.2** ではまとめて表現している。アクチュエータエージェントは他のエージェントから受け取ったモータへの指令値を FPGA ボードへ入力するエージェントである。

3.1.1 画像処理エージェントにとって認識が容易な状況での学習

ある環境では、画像処理エージェントが物体の境界を認識することができる (**Fig. 3.4**). 言い換えると画像処理エージェントは普通歩ディセンシングエージェントより優先度は低くなる。初期の状態では、移動エージェントが画像処理エージェントと体感エージェントより高い優先度をもつ。体感エージェントは境界の姿勢を知らない。ロボットは移動エージェントによって垂直面に近づくと接触する (**Fig. 3.5**). そしてその時、体感エージェントは他のエージェントより高い優先度を得る。移動エージェントは中断され、体感エージェントが起動し、画像処理エージェントからの情報を用いて適切なパラメータを学習する (**Fig. 3.6**). 具体的には、体感エージェントはロボットの画像処理エージェントによって見つけた境界の姿勢から自己位置を概算するデッドレコニングパラメータを調節する。

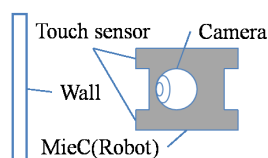


Fig. 3.3 Experiment 1-1

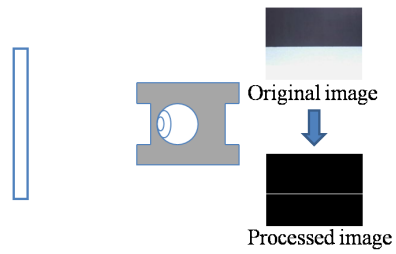


Fig. 3.4 Experiment 1-2

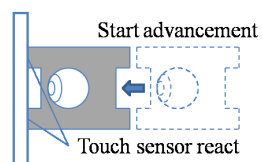


Fig. 3.5 Experiment 1-3

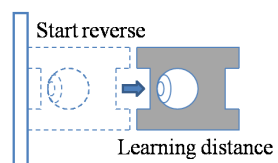


Fig. 3.6 Experiment 1-4

3.1.2 画像処理エージェントにとって認識が困難な状況における学習

他のある環境では，画像処理エージェントが認識できない (**Fig. 3.7**). 言い換えると，画像処理エージェントは普通体感エージェントより高い優先度を得る．ロボットは移動エージェントによって再び垂直面に接触するだろう (**Fig. 3.8**). そしてその時，画像処理エージェントは他のエージェントより高い優先度を得る．移動エージェントが中断され画像処理エージェントが起動しすでに調節されたデッドレコニングパラメータをもつ体感エージェントからの情報を用いて適切な閾値を学習する．具体的には，画像処理エージェントは体感エージェントによって概算された境界の姿勢から境界が探索できるように閾値を調整する (**Fig. 3.9, Fig. 3.10**).

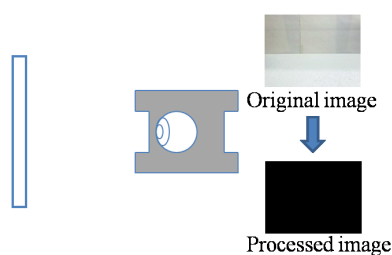


Fig. 3.7 Experiment 1-5

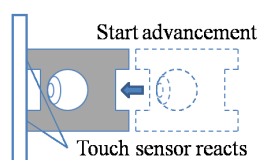


Fig. 3.8 Experiment 1-6

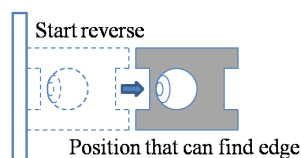


Fig. 3.9 Experiment 1-7

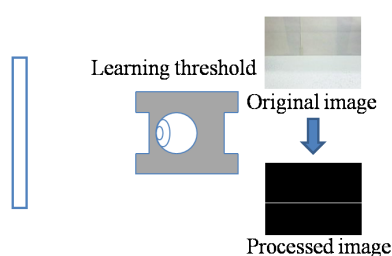


Fig. 3.10 Experiment 1-8

3.1.3 実験結果

Fig. 3.11 のような環境で実験を行った。Fig. 3.11 の左側では、画像処理エージェントは垂直面と水平面の色が大きく異なるため容易に境界が認識できる。Fig. 3.11 の上側では、壁と床の色が似ており、境界が認識し難い。ロボットは上記の 2 つの環境で連続して学習した。

上記のような実験を行ったところ環境に応じて 2 つのエージェントは協調的に学習でき、提案手法が効果的であることを確認した。Fig. 3.12 はロボットの 2 次元座標系における位置と向きを示している。Fig. 3.13 は体感エージェントと画像処理エージェントの学習・教示状態を示している。Fig. 3.12 と Fig. 3.13 から以下のことが確認できる。初めに、体感

エージェントが学習エージェントとして，教示エージェントである画像処理エージェントからの教示によって学習した．その後，画像処理エージェントが学習エージェントとして，教示エージェントである体感エージェントからの教示で学習した．

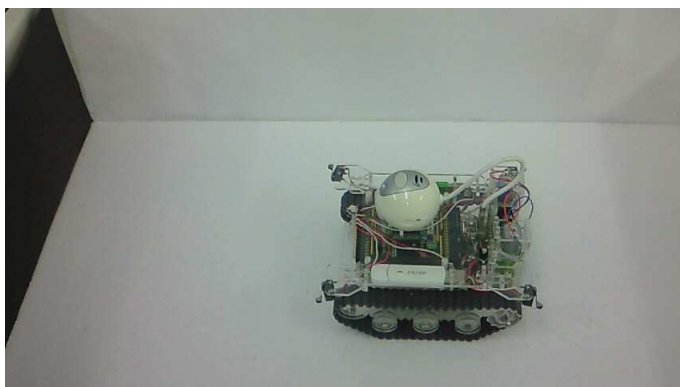


Fig. 3.11 Environment of experiment1

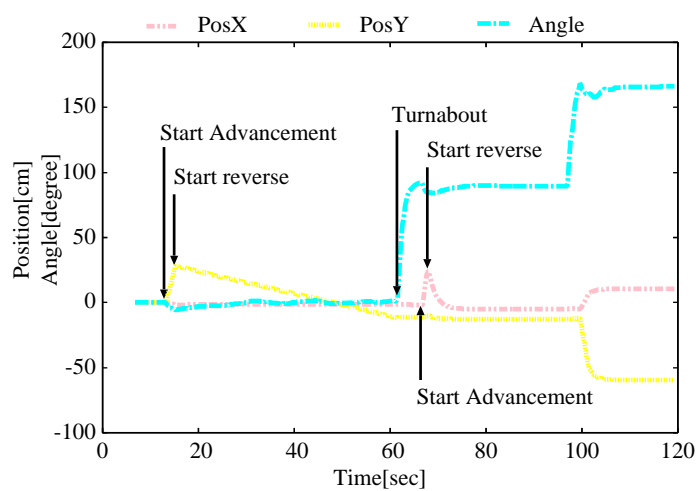


Fig. 3.12 Result of agent 's position in experiment1

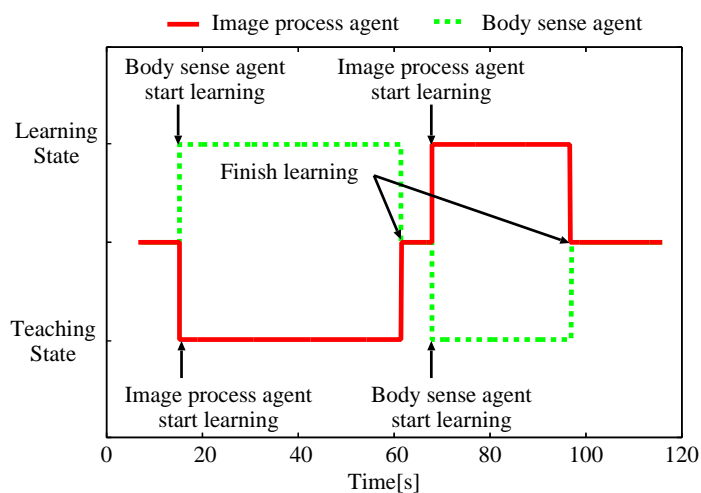


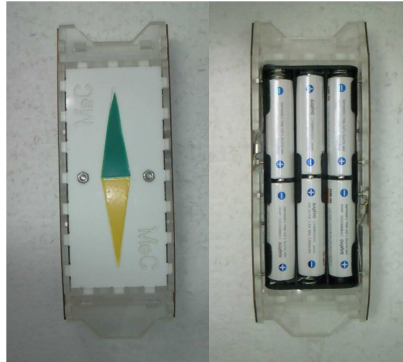
Fig. 3.13 Result of a agent ' s state in experiment1

3.2 複数のエージェント間の学習

3-1 ではロボットは何もタスクを実行せず，人に操作されるままに移動しながら，学習した．今回は電池交換というタスクを行いつつ複数のエージェントが学習することを実験的に確かめた．

3.2.1 電池交換

MieC は電池ボックス (Fig. 3.14) を自ら新しい電池ボックスに交換できる機構をもつ (Fig. 3.15)．電池は直列に電池ボックス両側面の電極に接続されている．電池ボックスは MieC 下部の金属レールに支えられ，MieC 側の電極に抑えつけられているだけである．そのため，電池ボックスが MieC の正面にあれば電池ボックスを壁などに押しつけることで新しい電池ボックスと交換できる．

**Fig. 3.14** Battery box**Fig. 3.15** Battery exchange mechanism

3.2.2 エージェント構成

エージェント全体の構成を **Fig. 3.16** に示す。画像処理エージェントからの情報を基に、接近、後退、交換エージェントが優先度を変化させながら電池交換する。画像処理エージェントは電池ボックスのマーカーの画像中の位置と角度を検出する。接近エージェントは電池

ボックスが見えた時にそれが交換できる位置になれば、優先度が上昇し、正面に回り込みながら接近する。後退エージェントは電池ボックスにそれ以上接近も交換もできなければ、優先度が上昇し、接近をやり直すために後退する。初期状態では後退する距離はランダムに与えられており、後退距離が多すぎて電池ボックスを見失うと優先度が上昇する。交換エージェントは電池ボックスが交換できる位置にあれば、優先度が上昇し、前進して交換する。初期状態では前進する距離はランダムに与えられており、前進距離が足りずに交換に失敗すると優先度が上昇する。体感エージェントは初期状態では 3-1 で壁との距離を学習した状態で実験する。

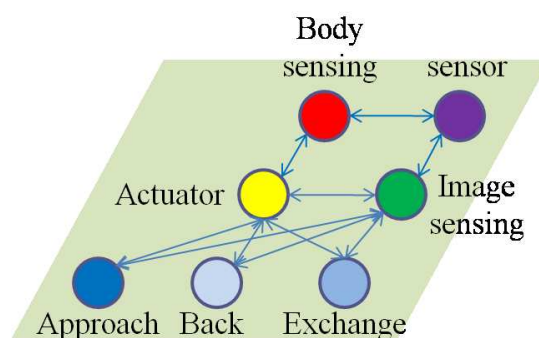


Fig. 3.16 Agents for experiment2

3.2.3 電池交換タスク中の学習

接近エージェントにより電池ボックスに接近する (Fig. 3.17). 接近も交換もできないので後退エージェントにより後退する (Fig. 3.18). 後退量が多く電池ボックスを見失うと後退エージェントの優先度が増加する (Fig. 3.19). 後退エージェントが画像処理エージェントからの教示情報を使い学習する (Fig. 3.20). 学習が完了すると、再び電池ボックスに接近する (Fig. 3.20). 交換不可能なので、適切な距離だけ後退する (Fig. 3.21). 再び接近

し，交換可能な状態になる (**Fig. 3.22**)．交換エージェントにより，交換するために前進する (**Fig. 3.23**)．交換できたか確認するために後退する (**Fig. 3.24**)．電池交換に失敗すると，交換エージェントの優先度が上がり体感エージェントからの教示情報により学習する．再び交換するために前進する (**Fig. 3.25**)．交換の成否の確認のために後退し，交換が完了する (**Fig. 3.26**)．

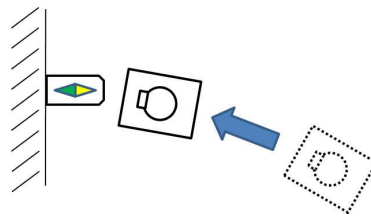


Fig. 3.17 experiment2-1

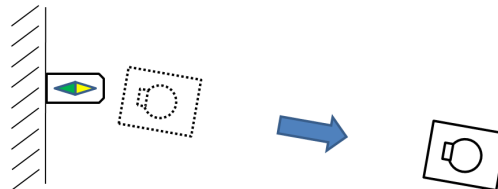


Fig. 3.18 experiment2-2

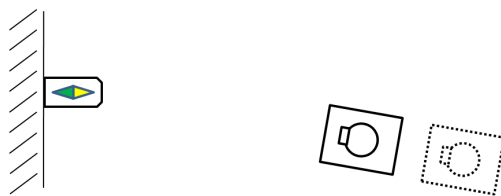


Fig. 3.19 experiment2-3

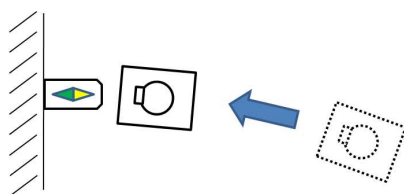


Fig. 3.20 experiment2-4

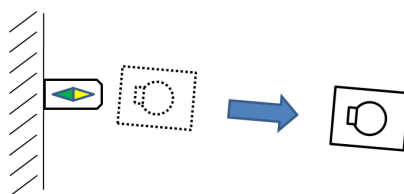


Fig. 3.21 experiment2-5

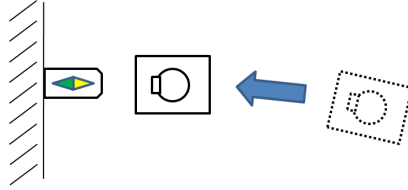


Fig. 3.22 experiment2-6

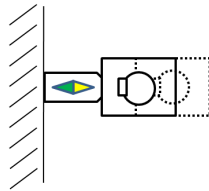


Fig. 3.23 experiment2-7

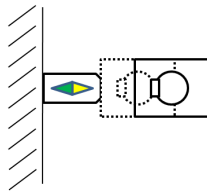


Fig. 3.24 experiment2-8

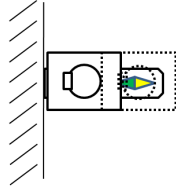


Fig. 3.25 experiment2-9

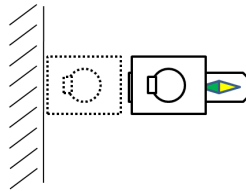
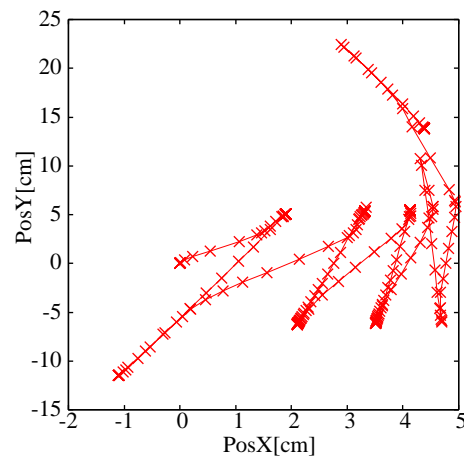


Fig. 3.26 experiment2-10

3.2.4 実験結果

Fig. 3.27 のような環境で実験を行った。Fig. 3.28 はロボットの 2 次元座標系における位置と向きを示している。Fig. 3.29 はエージェントの学習・教示状態を示している。Fig. 3.28 と Fig. 3.29 から以下のことが確認できる。電池交換のタスクを実行しつつ複数のエージェントが学習した。

**Fig. 3.27** Environment of experiment**Fig. 3.28** Result of agent 's position in experiment2

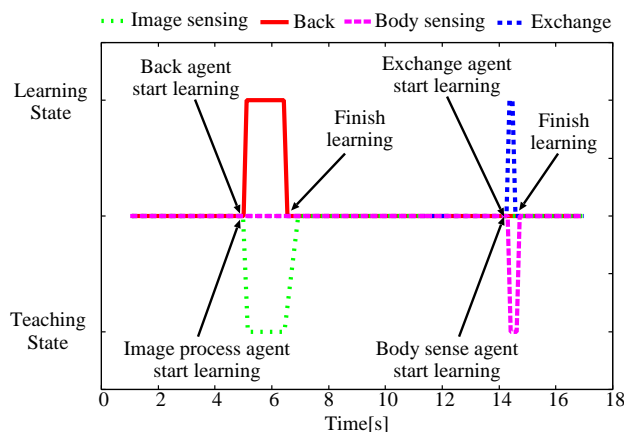


Fig. 3.29 Result of a agent ' s state in experiment2

現在は、エージェントは全て人の手によって作り込まれており、一部の環境のためのものである。また、学習エージェントはそれぞれ1つのパラメータを調整するだけである。しかし、将来全てのエージェントは提案モデルの中で自動的に生成され、より多くのパラメータを学習することで複雑な状況にも対応できるようになる。

具体的には、以下のようなアルゴリズムを提案モデルに組み込むことで可能になると考えられる。(1)いくつかのエージェントのテンプレートを用意する。(2)ロボットが環境に適応できない時、自動的にテンプレートを選択しエージェントが生成される。(3)初期状態ではエージェントのパラメータはランダムに与えられる。(4)エージェントは学習によってパラメータを調整する。(4)もし学習できなければ、他のテンプレートから新しいエージェントが作られる。

上記の手法も人間の学習手法を内観によって考えたものである。人間は遺伝的に持っている認識のための枠組みをもっており、必要に応じてそれを複製して学習することでより良い認識を得ていると考えたものである。

第4章

結言

本論文では心理学により定義される自我を含む意識を含む人間の学習の仕組みに基づく学習モデルを提案した。提案モデルを、当研究室で開発した実機ロボットに適用し、その有効性を確認した。具体的には、実験で画像処理エージェントと体感エージェントの2つのエージェントのみの場合と電池を交換する複数のエージェントが、MieCが移動することにより生じる環境の変化に合わせて学習者と教示者が役割を換えたことを確認した。

提案手法は教示情報を外部からのものだけでなく内部からのものも用いるため従来の強化学習より少ない試行回数で学習できることが期待できる。しかし、今回、学習するエージェントは環境の一部のみのためのプログラムであり、全て人の手で作り込まれている。今後の課題として、提案手法を、より多くのエージェントタイプを追加し、環境によって自動でエージェントを選択させることで、より大きく複雑な環境に適応できるように発展させたい。

参考文献

- [1] M.Minsky : “心の社会”, 産業図書, 1995
- [2] 前野隆司: “ロボットの心の作り方 受動意識仮説に基づく基本概念の提案 ”, 日本ロボット学会誌, vol. 23, no. 1, pp51-62, 2005
- [3] M.Wiering: “ HQ-learning ”, Adaptive Behavior, 6(2), pp219-246, 1997
- [4] H.Osada, S.Fujita: “ CHQ:A Multi-Agent Reinforcement Learning Scheme for Partially Observable Markov Decision Processes”, IEICE Trans. Inf. & Syst., vol.e88-D, No5, pp.1004-1011, 2005
- [5] N.Ono: “ Synthesis of Herding and Specialized Behavior by Modular Q-learning Animals”, ALIFE V Poster Presentations, 26-30, 1996
- [6] 藤田和幸, 松尾啓志: “状態空間の部分的高次元化法によるマルチエージェント強化学習”, 信学会論文誌, vol.J88-D-I, No.4, pp.864-872, 2005

謝辞

本研究を遂行するにあたり，御指導と御鞭撻を賜りました三重大学 松井 博和 助教に対して，謹んで感謝の意を表します．

本論文をまとめるにあたり，三重大学 加藤典彦 准教授から懇切なる御指導と御助言を頂き，深く感謝致します．

本研究に際して、様々なご指導を頂きました野村由司彦 理事，矢野賢一 教授に対して深謝いたします．

また，本研究を進めるにあたり，助け合い，相談しあった情報処理研究室の加藤雄一氏，柴田信雄氏，ならびにメカトロニクス研究室諸氏にも重ねて感謝致します．