Kobe University Repository : Kernel

PDF issue: 2025-07-16

多目的行動調停に基づく知能ロボットの行動獲得に関する研究

能島，裕介

(Degree)
博士（工学）
(Date of Degree)
2004-03-31
(Date of Publication)
2013-03-12
(Resource Type)
doctoral thesis
(Report Number)
甲3158
(URL)
https://hdl.handle.net/20.500.14094/D1003158

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。

博 士 論 文

多目的行動調停に基づく知能ロボットの行動獲得に関する研究

平成16年2月

神戸大学大学院自然科学研究科

能島裕介

目 次

第 1 章　緒論 **1**
1.1　研究背景 1
1.2　研究目的 4
1.3　本論文の構成 5

第 2 章　知能化技術とロボットシステム **7**
2.1　緒言 7
2.2　環境の定義 7
2.3　知能化技術 11
2.3.1　人工知能 12
2.3.2　ファジィ理論 14
2.3.3　ニューラルネットワーク 19
2.3.4　強化学習 23
2.3.5　進化的計算 26
2.3.6　遺伝的機械学習 31
2.4　ロボットの制御構造と学習機構 33
2.4.1　古典的人工知能とロボティクス 33
2.4.2　行動に基づくロボティクス 34
2.4.3　進化的ロボティクス 37
2.4.4　認知ロボティクス 38
2.4.5　構造化知能に基づくロボットシステム 39
2.5　結言 40

第 3 章　未知環境における移動ロボットの基本行動獲得 **43**
3.1　緒言 43
3.2　移動ロボットの基本行動 44
3.3　移動ロボットのための多目的行動調停 47
3.4　進化的計算とデルタルールによる基本行動の構造最適化と学習 51
3.5　計算機シミュレーション 54

3.5.1 基本行動獲得における動作の滑らかさ 55
3.5.2 複数未知環境における基本行動獲得の適応速度 59
3.5.3 未学習環境における実行可能性 64
3.6 結言 70

第 4 章 動的環境下における移動ロボットの局所エピソードに基づく環境適応 71
4.1 緒言 71
4.2 多目的行動調停に基づく移動ロボットの構造化学習 74
4.3 静的環境下での多目的行動調停則の獲得 76
4.4 多目的行動調停則の局所エピソード学習 77
4.5 計算機シミュレーション 78
4.5.1 静的環境下における行動調停則の獲得 79
4.5.2 動的環境下における局所エピソード学習 82
4.6 実機による実験 87
4.7 結言 90

第 5 章 実環境下におけるパートナーロボットのための多目的行動調停 91
5.1 緒言 91
5.2 パートナーロボットの構成 92
5.3 対話型進化的計算に基づく軌道生成 94
5.4 ニューラルネットワークを用いた基本行動の学習 98
5.5 多目的行動調停に基づく行動制御と学習 99
5.6 計算機シミュレーション 100
5.6.1 軌道生成と基本行動獲得 100
5.6.2 使用者との距離に基づく多目的行動調停 103
5.7 実機による実験 106
5.7.1 手渡し動作の軌道生成 106
5.7.2 多目的行動調停の適用 111
5.8 結言 116

第 6 章 結論 117
6.1 本研究のまとめ 117
6.2 今後の課題 119

謝辞 121

参考文献 **123**

第1章　緒論

1.1　研究背景

近年，知能ロボットの開発が盛んに行われ，生産現場から極限作業環境や生活環境など様々な環境でロボットが適用されている．極限作業環境や生活環境において，知能ロボットが考慮すべき目標（自己維持，探索，追従，採取など）は複数存在し，同時にそれらを満たす多目的な動作が知能ロボットに要求される．この多目的な動作は，個々の目的に対応した行動出力により構成されるが，耐故障性の観点から，一連の動作出力のギャップを緩和する滑らかさが必要とされる．また，人間との共生に対しても，多目的且つ滑らかな動作は，使用者に恐怖を与えずに円滑なコミュニケーションを行う上で必要である．さらに，極限作業環境や生活環境は，事前に想定できない未知な環境であり，その環境に適した行動を獲得する必要がある．そして，その獲得した行動は，新しい環境やタスクによる新しい状況に迅速に対応するために，再利用できなければならない．

ロボットの動作（制御構造）に関して，生産現場における古典的なロボットは，環境の完全なモデル化を必要とするが，正確にモデル化できる環境下で，事前に最適な動作計画が行える．一方，極限作業環境や生活環境において，ロボットは，事前に知ることができる情報が限られており，完全な環境モデルが得られず，観測できる情報が局所的であるため，逐次意思決定を行わなければならない．環境モデルとは，作業環境に関する記述であり，作業対象物の形状や寸法，位置，姿勢などの幾何学情報，材質や重量などの物理情報などを含み，ロボットが作業を計画したり，環境の認識を行うために必要な情報の集合である．この問題に対し，以降のロボット研究に大きな影響を与えた制御構造として，ある特定のセンサ情報に対して，実行できる行動を一対一で対応させるサブサンプションアーキテクチャが提案された．複数の目的に対して，それぞれ基本行動を設計し重層化する．そして，センサ情報のみから行動を反射的（排他的）に選択する．そのため，環境全体の事前知識を必要としない．この手法において，ロボットは，環境に置かれて初めて行動が選択され動き続けることができ，実行可能性の高い制御構造を持つ．しかし，個別の行動が排他的に用いられるため，行動の切り替わり時における制御出力信号の時系列にギャップが生じることがある．この制

御出力のギャップは，一連の動作を不安定にするだけでなく，故障の原因ともなりかねない．生活環境や極限作業環境などでは，このギャップを緩和できる滑らかな動作が必要とされている．

ロボットの学習に関して，観測情報に適した行動を獲得する様々な学習型ロボットが研究されてきた．従来の学習型ロボットは，サブサンプションアーキテクチャのような事前設計された反射行動を用いるのではなく，得られる報酬やタスクの達成度合いを高めるために自ら行動を獲得する．まず，目的を達成するための制御器をファジィコントローラやニューロコントローラ，クラシファイアシステムなどで構成する．そして，ロボットは実際の動作を通して，制御器の入出力関係を学習する．一般に，制御器の入力には観測情報が用いられ，出力には動作出力が算出される．制御器の学習方法は多岐に渡り，与えられたタスクとロボットが観測できる情報に依存する．例えば，報酬関数が与えられるような環境下において，強化学習に基づく制御器の学習手法がある．目的に対する価値を報酬関数で表し，ロボットは，未来に得られる価値を最大化することで最適な行動を探索する．ロボットは，事前に環境を知る必要がなく，行動が状態の価値として環境状態に意味付けされる．強化学習は，未知な環境において有効な学習手法の一つであるが，現在の状態から次の状態を決定できる性質であるマルコフ性が仮定されなければならず，生活環境のような他の行動主体が存在する環境下では適用が難しい．他にも，環境全体のモデルを必要としない手法に，進化的計算に基づく進化的ロボットがある．タスク終了時に実際得られた性能（例えば作業時間，精度，移動距離等）を評価することで学習が行えるため，マルコフ性が仮定されている必要がない．目的ごとの評価項目を用意し重み付けした評価関数を用いることで，多目的な行動を獲得することも可能である．しかし，獲得した行動は，単目的へと分割不可能であり，再利用しにくく，新たなタスク，新たな環境下で，新たに学習し直す必要がある．この再学習の問題は，強化学習を用いても起こりうる．

従来のロボットの制御構造や学習機構では，環境モデルが事前に得られない未知環境下で動き回ることができたり，特定の環境でタスクを遂行するための多目的な動作が獲得できるようになった．しかし，生活環境では，新しい作業環境や新たなタスクがロボットに与えられることが想定される．このとき，ロボットは，行動の再利用や再学習を行う必要がある．これら新しい作業環境やタスクに迅速に対応するためには，新たに行動を学習し直すよりも，これまでに獲得した行動を再利用できる方が良い．行動の再利用に関して，戦略（strategy）と戦術（tactics）の関係を例にとりあげる．戦略は，目標を達成するための手段の選択や配分のことを意味し，戦術は，具体的な行動がとれるように決めた方法（手段）のことを意味する．この意味で，ある状況下において，どのように戦術を用いるかが戦略を基にトップダウン的に決定され，個々の

戦術の統合により，ボトムアップ的に実際の振る舞いを決定する．本稿では，トップダウンを，マクロな秩序や挙動がミクロな振る舞いや相互作用に影響し変化をもたらすこととし，ボトムアップを，ミクロな相互作用を通してマクロな秩序や挙動が生成されることとする．一方，新たな目標が立てられた時や異なる状況に遭遇した時，一般に，戦略や戦術の一部を変更することで目標に対応する．また戦略や戦術は一旦有効であることが示された場合，同じような状況下で再び用いるために，破棄せず蓄積されるものである．そこで，個々の戦術を基本行動とみなし，戦略を行動調停則と考えると，複数の基本行動を調停する階層型分散システムが考えられる．行動調停則により基本行動へトップダウン的に役割（秩序）が与えられ，個々の基本行動が学習され，個々の基本行動の選択或は融合の結果，ボトムアップ的に実際の動作（振る舞い）が生成される．上記の戦略と戦術の観点から，階層型分散システムによるロボットは，過去に獲得した基本行動や行動調停則を明示的に保持できる構造を持ち，新たな目標や新たな状況に対して，基本行動や行動調停則の一部を変更することで適応するべきである．

周辺研究に関して，与えられた問題を部分問題に分割し，部分問題の解の統合により最終的な解を求め，問題を解決しようとする分割統合戦略（divide-and-conquer strategy）として，様々な階層型分散システムが，周辺分野でも盛んに研究されている．例えば，混合エキスパートやマルチモデル制御などは，分割されたサブモジュールの出力をゲーティングやスイッチングすることで，一時刻の状態に対して，特定の機能を個別に用いる手法であるが，一連の出力の滑らかさを考慮していない．また，多重順逆モデルやアンサンブル学習などは，サブモジュールの出力を重み付け平均により制御出力や学習結果として算出する手法であるが，分割されたサブモジュール間の相互作用によりボトムアップ的な振る舞いを主に扱い，トップダウン的な役割付けを行っていない．さらに，これらの階層型分散システムは，新たな環境や問題に対して，サブモジュールを追加する或は，事前に冗長なサブモジュールを用意しておく必要がある．このように，周辺研究においても，動作出力の滑らかさや，サブモジュールの再利用性に関してあまり議論されていないのが現状である．

以上の研究の背景から，現在，解決されていない知能ロボットの問題点として，行動切り替えによるロボットの制御出力信号の時系列に滑らかさがない点と，新たな環境やタスクに対する基本行動の再利用性が考慮されておらず最初から学習し直す必要がある点があげられる．これらの問題を解決するために，以下，本研究の目的について説明する．

1.2 研究目的

本研究の目的は，時系列観測情報を考慮した多目的行動調停に基づくロボットシステムを提案し，環境条件に合わせた基本行動の獲得や行動調停則の学習を通して，研究背景で取り上げた従来のロボットシステムにおける問題点を解決することである．

多目的行動調停は，時系列観測情報を考慮した行動調停則により，基本行動の出力を重み付け平均し，多目的な動作を生成する手法である．基本行動は，おかれた環境で与えられたタスクを達成するために必要となる副目標に対して，それぞれファジィコントローラで構築される．行動調停則は，直面する状況における基本行動の使用度合いを表す行動重みを逐次更新するために，簡単なプロダクションルールやファジィルールにより構成される．この行動重みの更新により，基本行動の使用度合いは，トップダウン的に調節される．そして，ロボットは基本行動の出力の重み付け平均により，ボトムアップ的に多目的な動作を生成する．ボトムアップ的な動作により，ロボットは異なる観測情報を得て，次の行動重みを更新し，トップダウン的に基本行動の使用度合いを決定する．これを繰り返すことで，行動調停則と基本行動との間に，互いに限定し合う相互依存の入れ子構造が生じる．この行動調停則と基本行動の入れ子構造により，どのような行動が獲得できるか，どのように行動を再利用できるかを議論する．

最初に，行動切り替え型ロボットの制御出力信号の時系列に滑らかさがないという問題に対して，多目的行動調停の動作生成により問題解決を試みる．重み付け平均による動作は，各基本行動の出力間のギャップを緩和し，制御出力信号の時系列の断続を押さえることができると考えられる．ただし，基本行動自体が別々に設計されていると，重み付け平均された動作が滑らかになる保証はなく，また，タスクを達成できる保証もない．そのため，他の基本行動や行動調停則に依存した行動として，基本行動が獲得される必要がある．そこで，タスクの達成度合いを評価する事で基本行動の獲得を行う．この基本行動の獲得を通して，生成される動作の滑らかさについて検証する．

次に，新たな環境に対して最初から行動を学習し直す必要がある問題に対して，複数の静的未知環境下での基本行動の学習による汎化と，行動調停則の適応学習により問題解決を試みる．多目的行動調停は，時系列観測情報を考慮することで，遭遇した状況に対して，どの基本行動がどれだけ動作に寄与すればよいかを決定でき，各基本行動の役割を明確にすることができる．つまり，与えられた役割に対して基本行動が意味付けられることで，おかれた環境に対して基本行動が特化する従来の行動学習手法よりも，未学習の未知環境に適応できると考えられる．これを，複数の静的未知環境で獲得した基本行動を用いて，未学習環境で性能評価を行うことにより検証する．また，質的に異なる新たな環境に対しても，最初から基本行動や行動調停則を学習し直

すのではなく，新たな役割（基本行動の用い方）を与えれるように行動調停則を学習することで，基本行動を再利用することができると考えられる．新たな環境として，移動障害物が存在する動的環境を対象とし，移動障害物との遭遇による異なる経験に対して行動調停則を学習する手法を検討し，基本行動の再利用性について検証する．

最後に応用として，人間が存在する環境における人型ロボットの基本行動獲得と，その基本行動を用いた多目的行動調停により，多目的且つ滑らかな動作によるヒューマンフレンドリなロボットの構築について検討する．

1.3 本論文の構成

第二章では，ロボットの適用される環境から問題のクラスを定義し，解決方法としての知能化技術について，環境を考慮に入れた説明を行う．

第三章では，ロボットが動作している間，障害物が移動しない静的環境下でのナビゲーション問題を取り上げる．この章で，本論文の軸となる多目的行動調停に関して説明を行う．主に基本行動の役割をトップダウン的に決定する行動調停と，基本行動の出力の融合によるボトムアップ的な動作について議論する．また，この行動調停と基本行動との間の入れ子構造について説明し，行動調停則が制約となって獲得される基本行動の特徴について検証する．

第四章では，より実環境に近づけるための適用環境の拡張として，移動障害物が存在する環境を取り上げる．静的環境下において，シナリオを評価することで基本行動や行動調停則を獲得することができるが，移動障害物との遭遇を考慮されておらず，ロボットは移動障害物と衝突する時がある．この動的環境では，静的環境で用いた行動調停則を移動障害物による新たな状況に適応させる必要がある．そこで，移動障害物との衝突の原因が最近の一連の動作にあると仮定し，その過去の動作に関与した行動調停則を局所的に改善する手法を提案する．この局所的な改善により，移動ロボットとの遭遇による新しい状況に対して，適した回避動作が生成できるようになることを検証する．

第五章では，パートナーロボットの構築を目指して，基本行動の学習と多目的行動調停の導入を試みる．時々刻々と変化する人間の評価を含む人間–ロボット環境での基本行動の獲得手法として，人間の評価モデルを同定しながら軌道探索を行う対話型進化的計算手法を提案する．さらに，獲得した基本行動を他の基本行動と多目的行動調停により融合することで，主行動の滑らかな遷移が実現できることを示す．

最後に，第六章において，三章から五章にかけて検証してきた構造的な学習手法についてまとめ，本研究を通して明らかになった事と残る課題をまとめる．

第2章　知能化技術とロボットシステム

2.1　緒言

本章では，問題環境の観点から，ロボットの知能化と制御構造について説明する．解くべき問題環境に対して，必要な意思決定機構と学習機構は異なる．これら意思決定機構と学習機構の内，本研究に関連する方法論について簡単に説明を行い，問題環境と合わせて議論を行う．まず，人工知能（artificial intelligence; AI）では，古典的人工知能（old-fashioned AI）から近年の身体性（embodiment）に関する議論について説明する．次に，我々が日常的に行う多様な問題解決や行為を再認し，柔軟な情報処理機能を実現しようという知能化アプローチであるソフトコンピューティング（soft computing）において，本研究で用いるファジィ理論（fuzzy theory），ニューラルネットワーク（neural network），強化学習（reinforcement learning; RL），進化的計算（evolutionary computation; EC）について説明する．そして，人工知能やソフトコンピューティング，認知科学などに基づいて構築される代表的なロボットアーキテクチャを紹介する．

2.2　環境の定義

ロボットを設計するには，最初に問題のクラスを定義する必要がある．これは，チェスのようなゲームと我々の日常環境とでは，ロボットが観測できるセンサ情報や動作の制限などが異なるためである．本研究では，問題のクラスを環境によって場合分けする．

Russell ら [1] は，環境を次の観点から分類している．

- アクセス可能 $\leftrightarrow$ アクセス不可能
- 決定的 $\leftrightarrow$ 非決定的
- エピソード的 $\leftrightarrow$ 非エピソード的
- 静的 $\leftrightarrow$ 動的

- 離散的 ↔ 連続的

＜アクセス可能 ↔ アクセス不可能＞：ロボットのセンサが環境の状態を知覚でき，動作選択に必要な情報が全て得られるのであれば，環境はアクセス可能という．アクセス可能であれば，センサ情報はマルコフ性が保証され，過去の情報を内部状態として保持する必要がない．（マルコフ性については，強化学習の項で説明する．）基本的に，実機でのセンシングには，観測範囲の制限や，ノイズや位置誤差などの外乱があり，正確に環境を観測することはできない場合が多い．＜決定的 ↔ 非決定的＞：現在のセンサ情報と，そのときのロボットの動作によって，次のセンサ情報が決まる場合を，環境は決定的という．また，マルコフ性が保証されるときにおいて，次の状態が特定の遷移確率で決まる環境をマルコフ決定過程という．アクセス可能で且つ決定的な環境の例として，チェスや将棋などのボードゲームがあげられる．盤面を見るだけで現在の状態を知ることができ，意思決定することができる．アクセス不可能な環境では，観測情報が不確かであり，環境が決定的であっても，見かけ非決定的になる．基本的に未知な環境では，どのようなセンサ情報が得られるかが事前に知ることはできず，環境は非決定的であるとみなされる．＜エピソード的 ↔ 非エピソード的＞：ロボットの知覚と動作の時系列が経験的に決まっており，現在の動作がエピソードだけに依存する環境をエピソード的という．この場合，一連の動作が固定されているシステムにのみ用いることができ，工場での組み立てや塗装ロボットがこれに該当する．この時，一連の動作が決まっているため，ほとんど意思決定を行う必要がない．＜静的 ↔ 動的＞：ロボットが意思決定を行っている間に環境が変化するとき，環境は動的であるといい，どれだけ意思決定に時間をかけても，環境が変化しないとき，環境は静的であるという．ただし，意思決定の間に環境が変化しないが，実際に取った動作によって環境が変化する場合，環境は準動的であるという．動的な環境下では，環境の変化に対応するために，有限時間での意思決定が必要となる．単一ロボットの問題では，静的環境を扱っているものが多いが，複数ロボット（マルチエージェント）の問題では，他のロボットの意思決定が非同期非通信を前提とするため，動的環境として扱われているものが多い．また，人間の生活環境における問題は，静的な環境と動的な環境の複合的な環境になっている．ただし，近年のロボット研究は，どちらか一方を対象としたものが多い．＜離散的 ↔ 連続的＞：知覚と動作を有限数で扱えるとき，環境は離散的という．ロボットの知覚や動作は，実環境に近づくほど連続値として取り扱う必要がある．

上記五つの特徴から，本研究で議論するロボットの適用環境の定義を行う．図 2.1 に示すとおり，静的既知環境，静的未知環境，動的環境，人間–ロボット環境の四つに分類する．

（1）静的既知環境は，ロボットが環境の大域的な状態を観測できるアクセス可能な

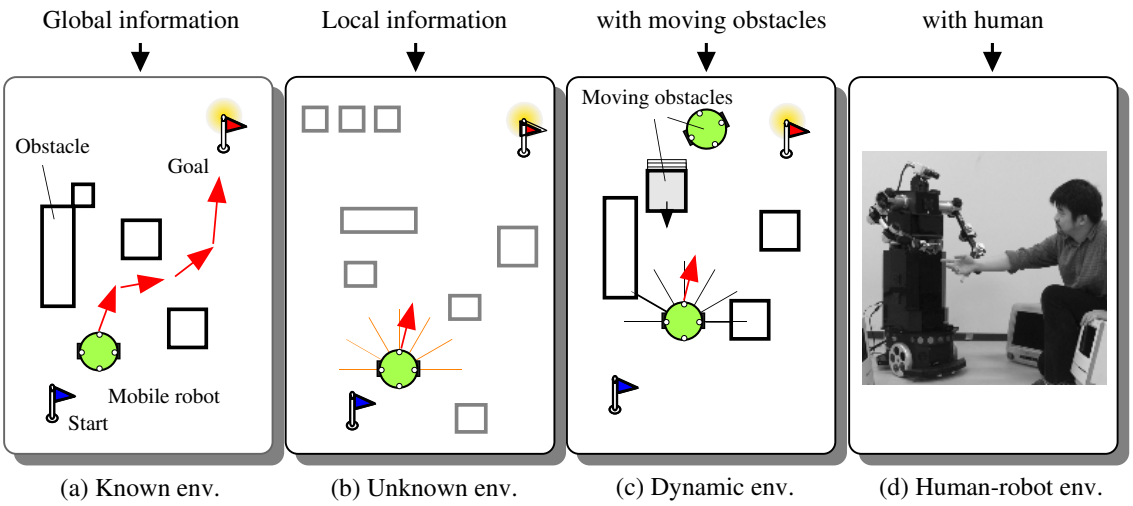


Fig 2.1 Environmental conditions.

環境とする．これは，環境モデルを知っていることと同義である．この環境内では，ロボットと固定配置された障害物のみ存在するものとする．このような環境は，古くから研究されており，ナビゲーション問題を，障害物の配置情報からグラフ探索問題として解いたり，人工ポテンシャル場を仮定することで，事前に最適な経路を探索することが可能である．つまり，事前に最適な動作列を決定でき，エピソード的であるとも言える．

(2) 静的未知環境は，ロボットが障害物の配置やそれらの密度などを事前に知らない環境である．固定配置した障害物に対して，意思決定の時間的な制限はなく，ロボットの動作にのみ観測情報が変化する静的環境である．また，観測できる範囲はロボットの近傍であり，完全な情報がアクセス不可能な環境とみなす．このことから，ロボットは，実際に環境の中におかれ，動作した結果を評価することで学習を行う．ここで，未知環境はさらに二つに細分することができる．一つは，設計者が知っている環境で，ロボットを設計する場合と，もう一つは，設計者も知らない環境であり，必然的にロボットが学習しなければならない場合とがある．ただし，環境が静的なことから，定量的な評価，例えば，タスク実行時間や達成度などからロボットの行動を評価することができる．

(3) 動的環境は，環境の中に移動障害物が存在するものとする．移動障害物の存在により，ロボットは有限時間で意思決定を行い動作する必要がある．ただし，移動障害物のダイナミクスを事前にロボットが知る場合は，それを考慮にいれて事前に意思決定が行えるため，動的環境とはいえない．その意味で，移動障害物の未来の動きがわからない場合，環境は非決定的であり，決められたエピソードを規定できず，局所的な

範囲のセンサ情報や，過去の時系列センサ情報，動作出力情報のみ用いることができる．静的未知環境では，一定時間の動作結果を評価し比較することで学習が行えるが，動的環境では，一定時間の中で，いつ，どこで，どのように移動障害物と遭遇するかが特定できず，一定時間の評価を学習器の比較には用いにくい．そのため，移動障害物に遭遇したときとそうでないときとを切り分けて学習する方が良いと考えられる．

(4) 人間–ロボット環境は，ロボットと人間が存在する環境であり，パートナーロボットとしてロボットにタスクが与えられる．そのため，実際に使用する人間の評価を考慮する必要がある．人間の評価は，ロボットの動作や動作の時系列だけでなく，人間個人の経験によって決まる．使用者が異なれば，評価も異なり，事前に評価関数を設計することはできない．また，人間–ロボット環境では，観測情報が局所的であるだけでなく，人間とロボットが互いに正確な情報を送り合うことができず，不確定な要素が大きい．これらをまとめると，環境は，アクセス不可能かつ，非決定的，非エピソード的で，動的且つ連続的であり，意思決定に用いることができる情報が時々刻々と変化し，問題のクラスとしては最も難しいと考えることができる．人間–ロボット環境全体を一度に解くことは不可能であり，様々な観点から問題を分割し，意思決定および学習を行う方が得策であると考えられる．ただし，単に分割するのではなく，全てにおいて整合性がとれる構造的な学習を検討する必要がある．

本論文では，3 章で，移動ロボットのナビゲーション問題に関して静的未知環境を，4 章では，移動障害物を含む移動ロボットのナビゲーション問題として動的環境を，5 章では，パートナーロボットの構築に関して，人間–ロボット環境を，それぞれ対象とする．

2.3 知能化技術

本研究では，2.2 による環境の定義に対応づけて，学習手法を検討する．一般に，学習手法は大別すると下記のようになる．

- 機械的な記憶による学習（rote learning）：知識や技術を直接ロボットに埋め込む学習であり，問題を熟知した設計者によるアルゴリズム開発やヒューリスティックな設計がこれに該当し，古典的な人工知能で成功をあげたプロダクションシステムなどがある．また，ノイズを含むような曖昧な観測情報に対しては，ファジィルールが適用され，感覚的なヒューリスティックを設計に反映させやすい．基本的にロボットが，ある環境におかれ，実際に動作することで学習を進めていくことが理想的である．しかし，未知環境においては，学習の指向性が全く与えられない場合，膨大な学習時間がかかることがある．機械的な記憶により知識を与えることによって，学習の速度や収束などが改善されることもある．

- 教示による学習（learning from instruction and advice taking）：指示やアドバイスから得た情報をロボットが元々持っていた知識や技術と統合する学習手法である．ある観測情報に対して，どのような制御出力を出力すべきかを学習する．代表的な学習手法として，ニューラルネットワークの誤差伝搬学習などがあげられる．静的既知な環境においては，教師値を用意できる場合が多く，求める性能が実現しやすい．逆に，静的未知環境や動的環境では，一時刻における理想的な状態が定義しにくく，適用は困難である．

- 事例からの学習（learning from example and by practice）：正例・負例あるいは実際の経験を通して，ロボットの知識や技術を摘出（extraction）・洗練（refinement）する手法である．ロボットの取った動作に対して，報酬と罰から制御器を学習する強化学習などがこれに該当する．スカラー値で表される報酬–罰のみから学習が行え，適切な教師値が与えられないような静的未知環境において，よく適用されている．ただし，報酬の構造を事前に与える必要があることと，報酬の構造が局所的にしか定義されないことから，大規模な問題（複数タスクが与えられる問題や，報酬が得られるまで時間が長くかかる問題）には適していない．

- 類推による学習（learning by analogy）：ロボットの持っている知識や技術を変形して，いままでに解いた問題と似ている問題を解決する．ファジィルールによるクラスタリングがこれに該当する．

- 発見的学習（learning by discovery）：観察や経験から新しい知識や制御器を構築する．一定時間の動作の後に評価値が計算できる場合に適用される．代表的な学

習手法として，進化的計算や遺伝的機械学習などがあげられる．例えば，遺伝的機械学習の中でルール集合を個体としたピッツアプローチという手法があるが，評価値に依存して遺伝的操作を行うことで，知識表現あるいは制御器を構成するルール集合が獲得できる．明示的な教師値や報酬の構造を考える必要がなく，必要な性能を指標とする評価関数さえ設定すれば学習ができる．この学習手法は，未知な環境に適用できるが，評価値に個別の動作の結果が反映されないため，動的環境では適用しにくい．

これらを踏まえて，人工知能，ファジィ理論，ニューラルネットワーク，強化学習，進化的計算，遺伝的機械学習について，基礎的な説明を以下に行う．

2.3.1 人工知能

初期のロボット研究に多大な影響を与えた人工知能について説明する．人工知能という概念が最初に提唱されたのは1956年であり，一時期，人工知能ですべての問題が解決できるとまで言われていた．当時の人工知能研究は，知識ベース（knowledge base）と推論エンジン（inference engine）を用いたプロダクションシステム（production system）に代表されるアルゴリズム中心の論理学に基づくアプローチであった．そのとき提案されたエキスパートシステム（expert system）は，専門的な知識を構造的に表現することが可能となり，産業分野に色々と応用された．簡単にプロダクションシステムについて説明すると，知識ベースはルール形式（基本的に IF-THEN ルール）に記述された知識をデータベースとして保持する．推論エンジンはワーキングメモリ（working memory）が持つモデルの状態にあわせて，一致するルールを知識ベースから選択する．そして，そのルールに従って，ワーキングメモリの状態は書き換えることにより目標に近付けたり，検証したりするシステムである．

知識表現としては，Minsky によるフレーム理論や，カーネギー・メロン大学の音声理解システムで開発された黒板モデル，心理学の分野で用いられた意味ネットワーク

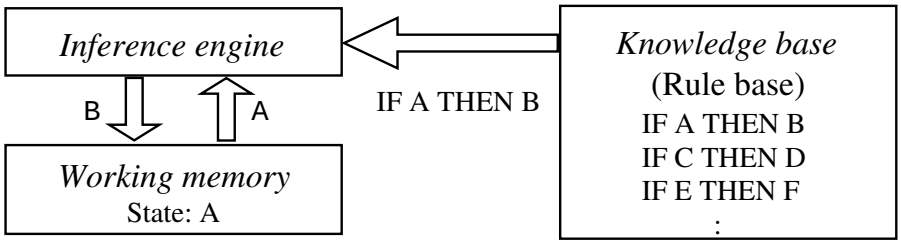


Fig 2.2 Concept of production system.

などが提唱された．

Russell は,「人工知能の研究は，知的存在を理解しようとするもので，我々自身をよりよく知ることであり，さらに，理解するだけでなく，それらを構築し，構築された知的存在が有効であることを示すもの」であると言っている [1]．実際には，人間のように考えれるかどうかを，AI の計算機モデルと認知心理学における実験とを比較することにより議論したり，計算機プログラムが人間のように行動できるかどうかをチューリングテストを用いて議論している．また，知的存在として人間の枠にとらわれず，合理的に考えたり行動できるかを，論理主義的アプローチにより議論されている．

人工知能は，知識の表現，推論規則の学習，問題解決，目標指向の動作など記号処理システムを基礎とする．記号処理システムでは，知識は記号（symbol）間の関係のネットワーク構造として表現される．そのネットワーク構造は種々の概念（concept）を表し，意思決定のもととなる．したがって，人工知能は記号主義あるいは表象主義と呼ばれる．人工知能の多くの研究は，システムをこの記号あるいは表象といった抽象的なモノで記述することにより知能を発現させようとしてきた．また，人工知能は，主としてコンピュータ上で議論されてきたため，知的主体としての身体を無視，或いは軽視してきた．つまり,「知能」が，環境と切り離された，閉じた，抽象的な，身体の動作を伴わない，静的な表象操作として捉えられてきた．環境と切り離された記号処理的なもので，人工知能が人間より優れた例として，チェスの世界チャンピョン Kasparov に，IBM のスーパーコンピュータ，ディープ・ブルーが勝利したことを容易に思い出せる．プログラムは単なる全探索だと言われたりもするが，目的を与え合理的な手を決定するという観点から人工知能であるとみなすことができる．ここで，対象とする問題（チェス）とコンピュータの関係が明確になる．チェスは離散化された問題空間で定義（構成）されており，高性能のコンピュータを用いることによって有限時間内に合理的な行動（手）を行うことができる．このように，限られた空間，構造化された問題において，人工知能は有効性を示すことができる．しかし，実際の環境は問題空間が構造化されておらず，従来の人工知能アプローチでは解ける問題も限定されている．現実世界は時間的空間的に連続であり，記号による記述が不可能であるためである．では，連続的な問題空間に対応できるように，記号による離散的な記述を避け，莫大な量の知識と処理能力を持つようにすれば良いのか？しかし，いずれは記述しきれない，或は処理しきれないという無限後退が起こりうる．

知的存在であろう私達人間は，限られた記憶と限られた情報処理能力，限られた行動で，うまく生活している．限られた能力だからこそ，知的な行為を起こすことを可能にしているのではないか．そこで，人工知能の分野や認知科学の分野で，主体の能力が環境との関わりの中で制限される身体性というものを考慮し，知的主体である環

境の中に存在する身体に関してもう一度深く考えてみようというアプローチが盛んに議論されるようになってきている．

2.3.2 ファジィ理論

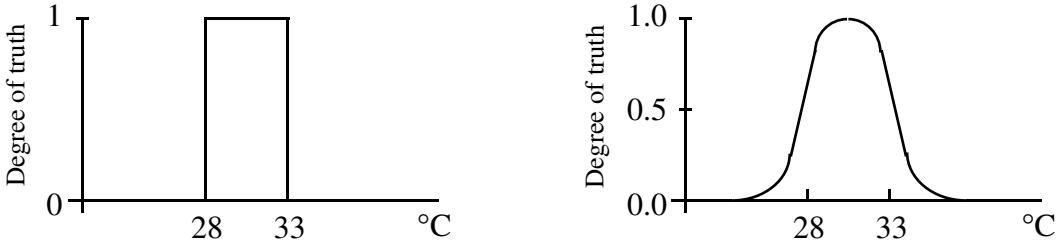
ファジィ理論は，1965 年に Zadeh が提案したファジィ集合論（fuzzy set）から始まり，集合論の拡張としてファジィ演算（fuzzy operation），確率論の拡張としてファジィ測度（fuzzy measure），多値論理の拡張としてのファジィ論理（fuzzy logic）とファジィ推論（fuzzy inference）の基礎理論から構成される．従来の科学において，あるシステムのモデルを構築する場合には，モデルを記述する精密さが重要とされていた．しかし Zadeh は，複雑なシステムを構築する場合に，モデルの正しさと精密さとは両立しないことを示唆し，ファジィ理論がもつ曖昧さがモデルを正しく表現できる場合があると主張した．この節では，特にファジィ集合から，ファジィ推論，その応用的方法論のファジィ制御（fuzzy control）を中心に説明する．

複雑なシステムは，実世界において無数に存在するが，人間が介在するシステムや，人間の知識や思考を模擬したシステム，さらに自然言語を用いたシステムなどは，人間の主観や知識，思考，自然言語などを用いるため，モデルに曖昧さが必要となる．ファジィ集合が提案されたのは，そのようなシステムを確率論で表現するのは不適であるということからである．従来の集合において，集合への要素の含まれ方が属するか否かの 2 値（1, 0）をとるクリスプ集合（crisp set）が用いられてきた．ファジィ集合は，属するか否かが 1 ∼ 0 の中間的な状態もとれるように拡張したものである（図 2.3）．例えば，気温が「暑い」という自然言語で表す温度は何度かとい っ たときに，クリスプ集合にて 28 ℃から 33 ℃までが「暑い」で，それ以外は暑くないという記述になり，非現実的である．そこでフ ァ ジ ィ 集合を用いて「暑い」範囲に曖昧さを持たせ，どの程度「暑い」のかを度合い（$fitness$）により示す．

ファジィ集合 A は，全体集合 X において，

$$\mu_A : X \to [0, 1] \tag{2.1}$$

と表す．ここで，μ は，メンバーシップ関数（membership function）といい，$\mu_A(x)$ は，要素 x がファジィ集合 A に属する度合い（適合度）を表す．ファジィ集合も，通常（クリスプ集合）の演算で定義されている包含関係，共通集合，和集合，補集合を，次の



(a) Crisp set　　(b) Fuzzy set

Fig 2.3 An example of crisp set and fuzzy set (hot).

ように定義することができる．

$$A \subset B \Leftrightarrow \mu_A(X) \leq \mu_B(X) \tag{2.2}$$

$$A \cap B \Leftrightarrow \mu_{A \cap B}(X) = \mu_A(X) \wedge \mu_B(X) \tag{2.3}$$

$$A \cup B \Leftrightarrow \mu_{A \cup B}(X) = \mu_A(X) \vee \mu_B(X) \tag{2.4}$$

$$\bar{A} \Leftrightarrow \mu_{\bar{A}}(X) = 1 - \mu_A(X) \tag{2.5}$$

A，B はファジィ集合，$\wedge$，$\vee$ はそれぞれ min 演算，max 演算を表す．

次に，ファジィ推論とファジィ制御について説明する．ファジィ推論は，ファジィ集合で構成されるファジィ命題（fuzzy proposition）からファジィ真理値（fuzzy truth value）を求めることにより，近似的な推論（approximate conclusion）を行うものである．ファジィ命題とは,「x は A」いう命題があった場合，述語 A がファジィ集合により記述されるものをいう．たとえば,「身長は高い」とか,「水は冷たい」．ファジィ命題は，述語修飾演算子を用いて記述することもできる.（例，ファジィ量限定子：「水はほとんどない」，ファジィ質限定子：「記録はほぼ正しい」）基本的にファジィ命題は IF-THEN ルールで表現される．

$$If\ x\ is\ A\ then\ y\ is\ B \tag{2.6}$$

ファジィ真理値においても,「やや真」のように言語的にその度合いを示す言語真理値や,「やや 0.8」のようなファジィ数真理値,「0.8」のような数値真理値で表すことができる．ファジィ制御では，複数の IF–THEN ルールを制御規則として用いて，ファジィ推論により操作量を決定する．ファジィ制御に対する推論法として，Mamdani のミニマックス重心法（min–max–gravity method）や代数積–加算–重心法（product–sum–gravity method），簡略化ファジィ推論法（simplified fuzzy reasoning method）などが用いら

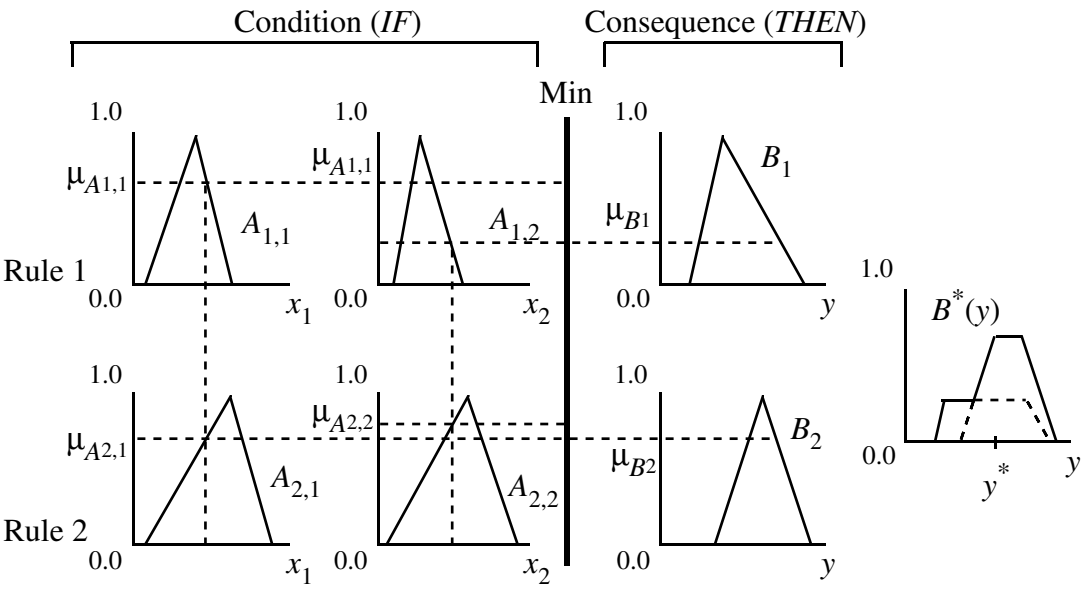


Fig 2.4 Inference procedure of the min-max-gravity method.

れる．次の一般化した IF-THEN ルールをもとに上記推論法を説明する．

$$\textbf{If}\ x_1\ is\ A_{i,1}\ and\ x_2\ is\ A_{i,2}\ and, ...,\ and\ x_n\ is\ A_{i,n}$$

$$\textbf{then}\ y_1\ is\ B_{i,1}\ and\ y_2\ is\ B_{i,2}\ and, ...,\ and\ y_o\ is\ B_{i,o} \tag{2.7}$$

ここで，$A_{i,j}$ と $B_{i,j}$ は，それぞれルール i の j 番目の入力と出力に関するメンバーシップ関数であり，n と o はそれぞれ入出力数である．入力 x_1 に対するファジィ集合への適合度を μ_i とし，適合度に対する出力を y とする．簡略化のためルールは 2 入力 1 出力とし，ルール総数を 2 つとする．ミニマックス重心法について，まず，各ルールの適合度 μ_i を求める．

$$\mu_i = \mu_{A_{i,1}}(x_1) \times \mu_{A_{i,2}}(x_2) \tag{2.8}$$

各ルールの後件部推論結果は次式となる．

$$\mu_{B_i^*}(y) = \mu_i \times \mu_{B_i}(y) \tag{2.9}$$

よって全体の推論結果は，

$$\mu_{B^*}(y) = \mu_{B_1^*}(y) \vee \mu_{B_2^*}(y) \tag{2.10}$$

となる．そして，出力を後件部推論結果の重心をとることにより求める．

$$y^* = \frac{\int B^*(y)\, y dy}{\int B^*(y)\, dy} \tag{2.11}$$

上記の操作を図 2.4 に示す．

次に，代数積–加算–重心法について説明する．代数積–加算–重心法は，ミニマックス重心法の min 演算部分を代数積演算に，max 演算部分を加算演算にしてものである．それにより推論結果の線形補間が行われ，非線形性が弱められる．各ルールの適合度 μ_i を求める．

$$\mu_i = \mu_{A_{i,1}}(x_1) \times \mu_{A_{i,2}}(x_2) \tag{2.12}$$

各ルールの後件部推論結果は次式となる．

$$\mu_{B_i^*}(y) = \mu_i \times \mu_{B_i}(y) \tag{2.13}$$

よって全体の推論結果は，

$$\mu_{B^*}(y) = \mu_{B_1^*}(y) + \mu_{B_2^*}(y) \tag{2.14}$$

となる．(2.11) 式と同様に後件部推論結果の重心をとることにより出力 $y*$ を求める（図 2.5）．

次に，簡略化ファジィ推論について説明する．簡略化ファジィ推論は，代数積–加算–重心法のルール後件部のファジィ集合を実数値 w に換えたものであり，計算の高速化と簡略化が行われる．各ルールの適応度は (2.12) 式により求める．推論結果の出力 y は次式により求める．

$$y = \frac{\sum \mu_i w_i}{\sum \mu_i} \tag{2.15}$$

図 2.6 にて推論の手順を図示する．

ファジィ制御の特徴として，制御規則を入力変数の領域別，出力変数別，目的別に並列に記述することができ，推論により全体の調和をとる出力が得られる．また，制御規則として用いる IF–THEN ルールは言語ラベル（linguistic label）により記述することができるので，曖昧さを含む知識を表現することが容易である．また，直感的に把握しやすいため，制御規則の修正・変更も言語レベル（linguistic level）で行える．これらの特徴から，ファジィ制御は，エキスパート的，ヒューリスティックな制御に用いることができる．応用分野としては，機械の故障診断，意思決定，人工物デザイン，スケジューリング，プロセス制御，家電製品，自然言語解析，画像理解，ロボティクス，心理学的モデル等に用いられている．

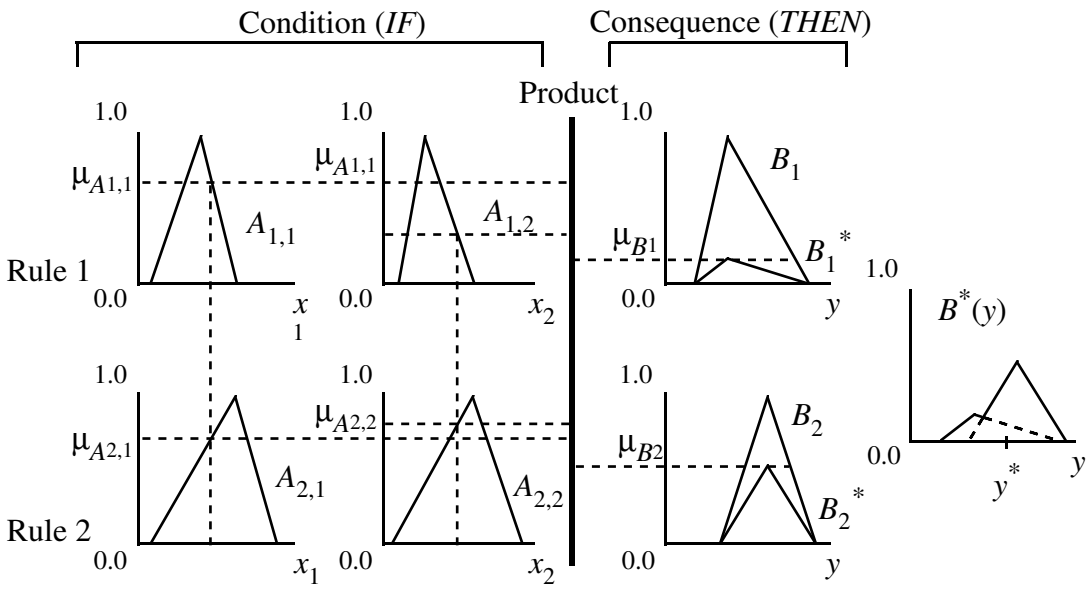


Fig 2.5 Inference procedure of the product-sum-gravity method.

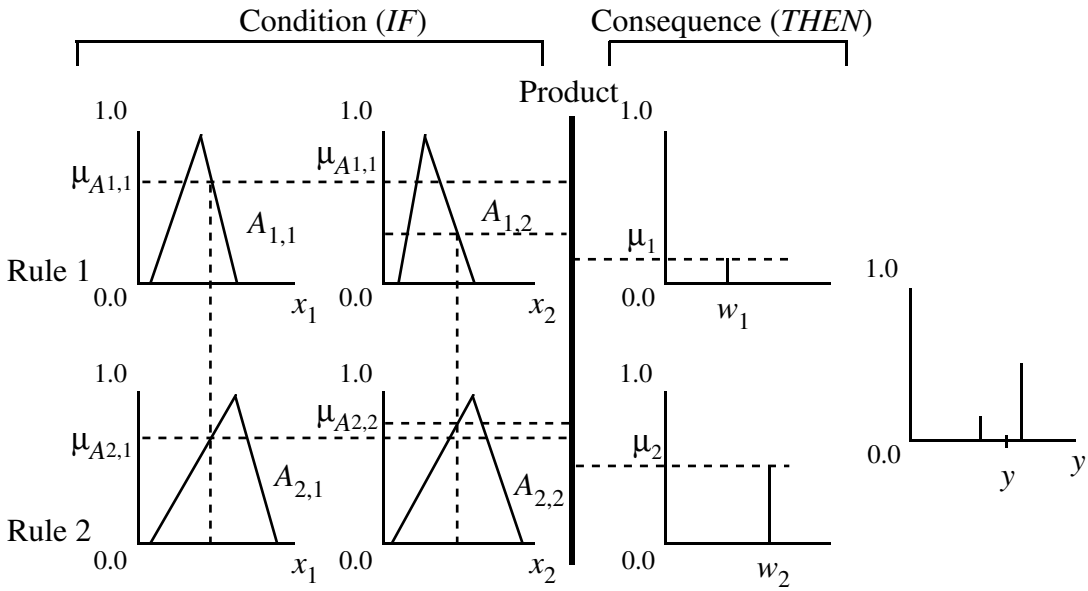


Fig 2.6 Inference procedure of the simplified fuzzy reasoning method.

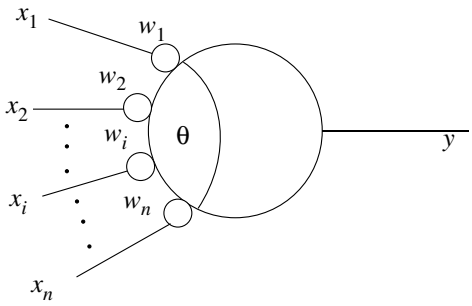
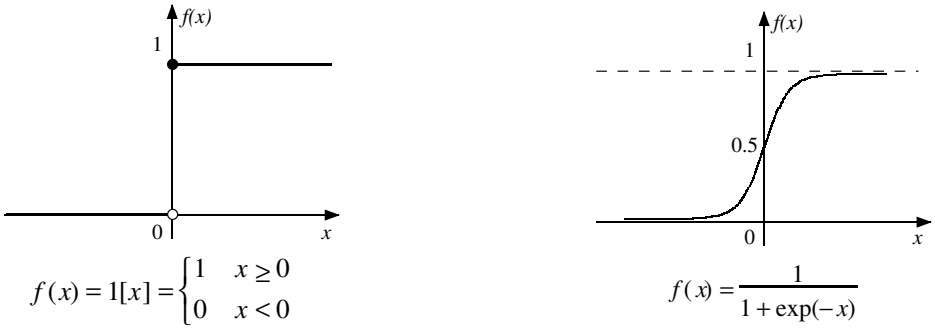


Fig 2.7 Neuron model.



(a) Step function

(b) Sigmoid function

Fig 2.8 Step function and sigmoid function.

2.3.3 ニューラルネットワーク

ニューラルネットワーク（neural network; NN）は，生物の神経系の特徴的な機能に着目して，そのモデル化を行ったものである [9]．1943 年に，McCulloch-Pitts が，生物系内のニューロンの動作原理に基づいて図 2.7 と ((2.16)) 式で示されるニューロンのモデルを提案した．

$$y = 1\left[\sum_i w_i x_i - \theta\right] \tag{2.16}$$

x_i，y はニューロンへの入力とニューロンからの出力，w_i はシナプス結合強度を表している．また，$1[x]$ は図 2.8(a) に示すように，$x \geq 0$ のときは 1，$x < 0$ のときは 0 となる単位ステップ関数である（1 は興奮状態，0 は静止状態を示す）．((2.16)) 式は，ニューロンに伝わった信号 x_i が重み付けされて加算され，それがある閾値 θ を超えるとニューロン i が興奮することを意味している．また，((2.16)) 式における $1[x]$ という単位ステップ関数の代わりに ((2.17)) 式で表されるシグモイド関数 (sigmoid function) やラジアル基底関数 (radial-based function) が使用されることもある．シグモイド関数や，ラジアル基底関数を用いた場合，連続値を取り扱うことができ，ロボットの距離センサや光

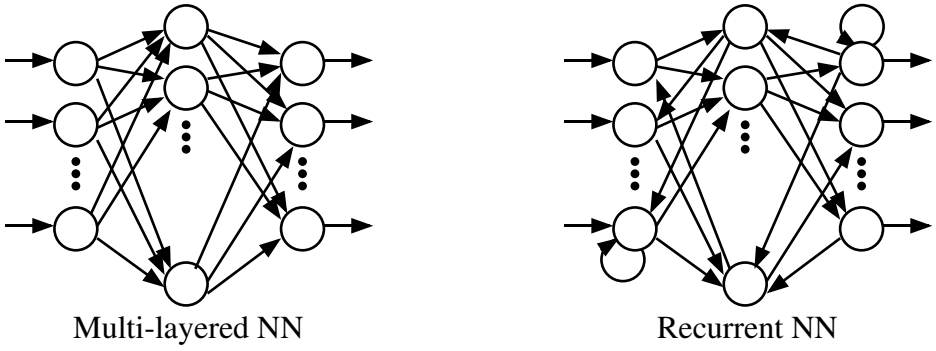


Fig 2.9 Network structure of neural network model.

センサ，音センサなどの連続値の観測情報をそのまま用いることができる．

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{2.17}$$

この人工ニューロンを一つのノードとし，複数結合することによって様々な NN のモデルが構成される．図 2.9 のように，入力ユニットから出力ユニットまで順方向のみに階層的に結合れ，フィードバック結合などの相互結合の形態を持たないような NN モデルを階層型 NN モデル (multi-layered neural network model) と呼ぶ．階層型 NN を用いることで，静的な入出力間の対応関係を表現することは容易であるが，動的なダイナミクスを表現するには工夫が必要である．一方，対象とするモデルの現在の出力が，過去の時系列情報に依存するようなダイナミクスを含む場合において，出力ユニットや中間ユニットから入力ユニットに戻る逆方向の結合を用いる相互結合 NN(recurrent neural network) も盛んに研究されている．中間ユニットから入力ユニットへ出力信号を送るエルマン型や，出力ユニットから入力ユニットへ出力信号を送るジョルダン型，全てのノードが対称的に結合したホップフィールド型などがある．NN 制御を想定すると，NN は，制御対象のダイナミクスや逆システムの構築によく用いられる [5, 10].

階層型 NN を例にとり，対象とするモデルの入出力関係を学習する方法について説明する．適切な結合強度を得るための NN の学習方式としては，教師あり学習 (supervised learning) と教師なし学習 (unsupervised learning) がある．教師あり学習の場合は，NN からの出力と教師信号を比較することによって，その差をできるだけ小さくするよう結合強度の値を変更する．一方，教師なし学習の場合は，理想的な出力は外部から与えられないので，自分自身の評価基準を内蔵しておくことが必要となる．ここでは，教師あり学習の 1 つである逆誤差伝搬法 (back propagation method;BP method) について説明する．BP 法は 1986 年に Rumelhart らによって提案された学習アルゴリズムである．h_i を非減少関数とすると，図 2.10 の順計算は，

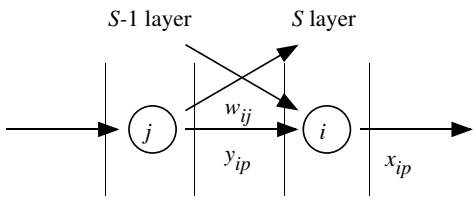


Fig 2.10 Foward calculation.

$$x_{ip} = h_i(z_{ip}) \tag{2.18}$$

$$z_{ip} = \sum_j w_{ij} y_{ip} \tag{2.19}$$

となる．第 p 番目の入力パターンに対する出力層の i 番目のユニットからの出力値を $\bar{x}_{ip}$ とし，教師信号を d_{ip} とする．このとき誤差関数 $E_p(W)$ 及び総誤差関数 $E(W)$ は，

$$E_p(W) = \frac{1}{2}\sum_i (\bar{x}_{ip} - d_{ip})^2 \tag{2.20}$$

$$E_p(W) = \sum_p E_p(W) \tag{2.21}$$

と表される．ただし，W はノード間の結合強度とする．最急降下法（stepest descent method）より勾配は，

$$\frac{\partial E_p(W)}{\partial w_{ij}} = \frac{\partial E_p(W)}{\partial z_{ip}} \frac{\partial z_{ip}}{\partial w_{ij}} \tag{2.22}$$

となり，$z_{ip} = \sum_j w_{ij} y_{ip}$ より，$\frac{\partial z_{ip}}{\partial w_{ij}} = y_{ip}$ となることから，

$$\frac{\partial E_p(W)}{\partial w_{ij}} = \frac{\partial E_p(W)}{\partial z_{ip}} y_{ip} \tag{2.23}$$

と展開することができる．ここで，

$$\frac{\partial E_p(W)}{\partial z_{ip}} = \frac{\partial E_p(W)}{\partial x_{ip}} \frac{\partial x_{ip}}{\partial z_{ip}} \tag{2.24}$$

は，次式を用いて，

$$\frac{\partial x_{ip}}{\partial z_{ip}} = h'(z_{ip}) \tag{2.25}$$

$$\frac{\partial E_p(W)}{\partial z_{ip}} = \frac{\partial E_p(W)}{\partial x_{ip}} h'(z_{ip}) \tag{2.26}$$

となる．今，

$$\delta_{ip} = -\frac{\partial E_p(W)}{\partial x_{ip}} h'(z_{ip}) \tag{2.27}$$

とおくと，

$$\frac{\partial E_p(W)}{\partial w_{ij}} = -\delta_{ip} y_{ip} \tag{2.28}$$

となる．これを次式の更新則に代入して学習を行う．

$$W^{(t+1)} = W^{(t)} - \eta \left. \frac{\partial E_p(W)}{\partial W} \right|_{W=W^{(t)}} \tag{2.29}$$

η は，学習係数（ステップ幅）である．

- 第 s 層第 i ニューロンが出力層に属している場合，出力 x_{ip} は $\bar{x}_{ip}$ に相当する．

$$\frac{\partial E_p(W)}{\partial x_{ip}} = \bar{x}_{ip} - d_{ip} \tag{2.30}$$

  したがって，

$$\delta_{ip} = -\left(\bar{x}_{ip} - d_{ip}\right) h'\left(z_{ip}\right) \tag{2.31}$$

  となる．

- 第 s 層第 i ニューロンが出力層に属していない場合，

$$\frac{\partial E_p(W)}{\partial x_{ip}} = \sum_k \frac{\partial E_p(W)}{\partial z_{kp}} \frac{\partial z_{kp}}{\partial x_{ip}} \tag{2.32}$$

  となり，

$$\begin{aligned} \frac{\partial E_p(W)}{\partial z_{kp}} &= \frac{\partial E_p(W)}{\partial x_{kp}} \frac{\partial x_{kp}}{\partial z_{kp}} && (2.33) \\ &= \frac{\partial E_p(W)}{\partial x_{kp}} h'_k\left(z_{kp}\right) && (2.34) \\ &= -\delta_{kp} && (2.35) \end{aligned}$$

  となる．ここで，

$$\frac{\partial z_{kp}}{\partial x_{ip}} = \frac{\partial z_{kp}}{\partial y_{kp}} = w_{ki} \tag{2.36}$$

  という関係から，

$$\delta_{ip} = h'_i\left(z_{ip}\right) \sum_k \delta_{kp} w_{ki} \tag{2.37}$$

  となる．

NN を用いることにより，非線形な入出力関係を持つ対象モデルを学習することができ，プラントの制御から，移動ロボットやマニピュレータの制御などに適用されている．ただし，結合強度自体から，学習対象のモデルを解釈することは困難であり，ブラックボックスとして扱う必要がる．加えて，学習時の入力データを内分する入力データに対して，出力値は近似されるが，未学習領域に対する入力データが与えられたときの出力値は保証されない．これらの問題を補うために，ファジィ理論を用いた拡張手法として，様々なファジィ・ニューラルネットワーク（fuzzy neural network）手法が研究されている．

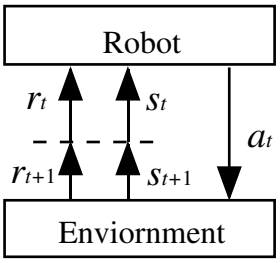


Fig 2.11 Relationship between a robot and the environment.

2.3.4 強化学習

強化学習は，ロボットが環境へ働きかけたことによって得られる報酬や罰の情報のみから，未来に得られる報酬を最大化する方策を見つける学習手法である [11,14]．まず，強化学習におけるロボットの入出力関係を定義する．現在の時刻 t におけるロボットの状態を s_t $(\in S)$ とする．この状態 s_t のもとで，ロボットは，方策 π_t により行動 a_t $(\in A)$ をとる．(方策 π_t は，状態から可能な行動を選択する確率の写像関数であり，$\pi_t(s,a)$ は，もし $s_t = s$ ならば $a_t = a$ となる確率である．) 行動の結果，ロボットは，環境から報酬 r_{t+1} を受け取り，新しい状態 s_{t+1} にいることを知る．ロボットは，未来に得られる報酬の和である期待収益（expected return）R_t の最大化を目的とする．

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{2.38}$$

γ は，減衰率 $(0 < \gamma < 1)$ とする．ロボットは，期待収益を最大化するための行動を方策により選択するが，この選択はロボットの状態の価値や行動の価値に依存する．強化学習では，これらの価値を現在の状態に依存した関数とするために，マルコフ性（Markov property）を仮定する．一般に，次の状態は，過去から現在までのすべての状態と行動に依存していると考えられる．状態 s_t で行動 a_t をとり，遷移確率 $P^{a_t}_{s_t s_{t+1}}$ で状態 s_{t+1} になると考えると，遷移確率は条件付き確率で次のように記述できる．

$$P^{a_t}_{S_t S_{t+1}} = P\left\{s_{t+1} \left| s_t, a_t, s_{t-1}, a_{t-1}, ..., s_0, a_0 \right.\right\} \tag{2.39}$$

ここで，状態がマルコフ性を持つとするならば，次の状態は現在の状態と行動にのみ依存する [11]．

$$P^{a_t}_{S_t S_{t+1}} = P\left\{s_{t+1} \left| s_t, a_t \right.\right\} \tag{2.40}$$

このマルコフ性の仮定により，方策 π に従った時の現在の状態の価値は，状態価値関数（state-value function）として次式のように定義される．

$$V^{\pi}(s) = E_{\pi}\left\{R_t \left| s_t = s \right.\right\} = E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \left| s_t = s \right.\right\} \tag{2.41}$$

E_π は，ロボットが方策 π に従うときの期待値を表す．同様に，方策 π のもとで状態 s で行動 a を取るとき，行動価値関数（action-value function）は，次式で定義される．

$$Q^\pi(s,a) = E_\pi\{R_t | s_t = s, a_t = a\} = E_\pi\left\{\sum_{k=0}^{\infty}\gamma^k r_{t+k+1} | s_t = s, a_t = a\right\} \tag{2.42}$$

ロボットは，この価値と方策に基づいて行動を選択する．代表的な行動選択手法として，価値が最大となる行動を選択するグリーディ手法（$\max_{a\in A} Q(s,a)$）や，確率 $1-\epsilon$ でグリーディな行動をとり，確率 ϵ でランダムな行動を選択する ϵ グリーディ手法などが用いられている．また，行動価値の比によって確率的に行動を選択するソフトマックス手法もよく用いられている．一般に，Boltzmann 分布が用いられ，温度係数 T を導入した次式の確率 $p(a|s)$ により選択を行う．

$$p(a|s) = \frac{e^{Q(s,a)/T}}{\sum_{a_i\in A} e^{Q(s,a_i)/T}} \tag{2.43}$$

温度係数 T が高ければ，すべての行動が同程度に選択されやすくなり，逆に，T が低ければ，価値の高い行動が選択されやすくなる．

強化学習は，動的計画法からの発展として，Sutton による TD 学習や Watkins による Q 学習，Rummery と Niranjan による Sarsa が開発された．また後述する分類子システムからの流れとして，Holland によるバケツリレーアルゴリズムや，Grefenstette による Profit Sharing が開発された．これらは 1980 年中頃から 1990 年中頃にかけて発表されている．また，それ以前の 1973 年には Widrow によって Actor-critic モデルが提案されている．さらに古く 1960 年代には，強化学習と関連が深いと言われている LMS 法やモンテカルロ法などが提案されている [11,14]．上記の様々な学習手法は，その学習構造からブートストラップ型と非ブートストラップ型に分類することができる [14]．前者は，他の状態価値や行動価値に依存して現在の価値を学習する手法であり，TD 学習や Q 学習，バケツリレーアルゴリズム，Actor-critic がこれに該当する．例えば，TD 学習は，次の状態 s_{t+1} での価値 $V(s_{t+1})$ と現在の状態 s_t での価値 $V(s_t)$ との差から学習を行う．

$$V(s_t) \leftarrow V(s_t) + \alpha[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)] \tag{2.44}$$

α はステップサイズパラメータといい学習率を表す．γ は割引率である．式からわかるとおり，報酬が与えられない時でも，次の状態の価値に依存して状態価値は更新され，報酬が得られた状態から次第に価値が伝播される．Sarsa は，上式の学習則を行動価値を用いて，より詳細にした学習モデルであり，次式により行動価値を更新する．

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)] \tag{2.45}$$

Q 学習は，Sarsa のように次状態でとった行動の価値を用いず，方策とは独立に最適行動価値関数を直接近似する．

$$Q\left(s_t, a_t\right) \leftarrow Q\left(s_t, a_t\right) + \alpha\left[r_{t+1} + \gamma \max_a Q\left(s_{t+1}, a\right) - Q\left(s_t, a_t\right)\right] \tag{2.46}$$

Q 学習は，マルコフ性が成り立つ状態で，すべての状態を選択し続けることで行動価値関数が収束することが知られており，数多くの事例が報告されている．Actor-critic は，actor と呼ばれる方策部と，critic と呼ばれる行動の評価部を別々に保持する TD 学習手法である．基本的に，評価部で行動の結果得られた状態との価値の差を TD 誤差 δ_t として定義し，状態価値関数の学習を行う．同時に，その TD 誤差を行動部にフィードバックすることで，行動優先度 $p(s_t, a_t)$ が更新される．

$$\delta_t = r_{t+1} + \gamma V\left(s_{t+1}\right) - V\left(s_t\right) \tag{2.47}$$

$$p\left(s_t, a_t\right) \leftarrow p\left(s_t, a_t\right) + \beta\delta_t \tag{2.48}$$

β はステップサイズパラメータである．行動選択は，行動優先度に基づいてソフトマックスにより行うことができる．

一方，非ブートストラップ型の学習手法は，一連の行動をエピソードとし，そのエピソードに対して評価を行う手法であり，profit sharing やモンテカルロ法，LMS 法などが該当する．ロボットは，評価値をエピソード中に遭遇したすべての状態に対して分配することで，他の行動価値に関連せずに行動価値を学習することができる．例えば，profit sharing であれば，報酬 r_{t+1} が与えられた時刻を $t+1$ としたとき，時刻 $t-j$ における状態 s_i の状態価値関数は，次式によって更新される．

$$V\left(s_i\right) \leftarrow V\left(s_i\right) + \eta\left[\gamma^{t-j} r_{t+1} - V\left(s_i\right)\right] \tag{2.49}$$

エピソード単位で学習を行うため，すべての状態価値が収束するとは限らないが，一旦，高い値で状態価値が収束すると，その状態をとるようにエピソードが生成されやすくなり，ある程度よいパフォーマンスが得られる．ただし，最適性は保証されない．

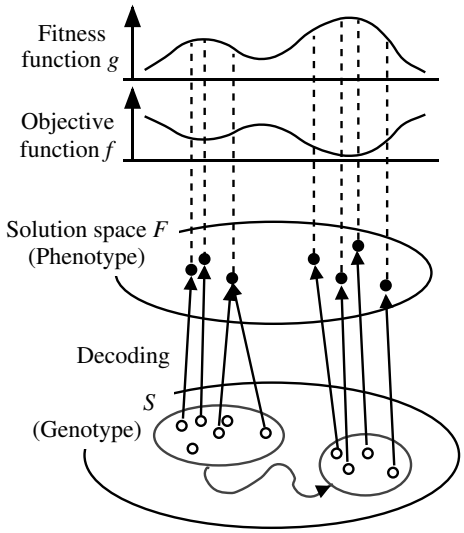
2.3.5 進化的計算

生命の環境への適応を進化として捉える考え方が，Lamarck の「動物哲学」によって提案され，その合理的メカニズムとして C.Darwin と A.R.Wwallace によって自然選択説（natural selection）が提案された．Darwin は自然選択説を「種の起源」により発表した．自然選択説は，異種間の交わりや何らかの原因で変異し新しい新種が発生したときに，その新種がその周りの環境への適応の度合に応じて増殖していくとともに，変異の一部分が子孫に伝えられるというものである．一方，親から子への遺伝については Mendel による研究以来，遺伝学（genetics）や集団遺伝学（population genetics）として研究されてきた．

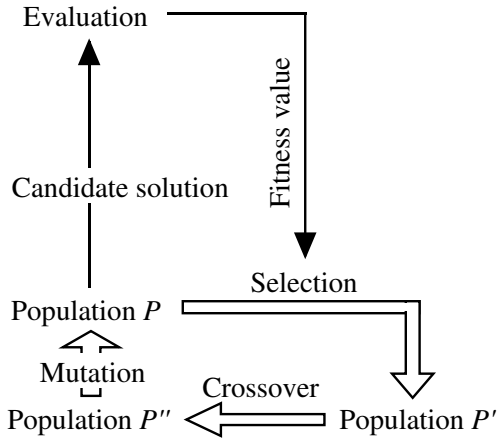
生物は細胞によって構成されており，細胞は分裂することによって自己複製することができる．細胞中の染色体（chromosome）のなかに各生物固有の情報を保持する遺伝子が折り畳まれて存在する．遺伝子の本体は，DNA（リボ核酸）であり，遺伝情報をコード化してもっている．遺伝情報は，細胞分裂の際に DNA が複写されて新しい細胞に入ることによって継承される．しかし，ごくまれに複写に誤りが生じる．これを突然変異（mutation）といい，新しい型の染色体が作られる．また，有性生殖をする多くの生物は同じ遺伝子座をもつ二組の染色体をもっており，生殖の際に相同染色体間で交叉（crossover）による組み換えが起こる．つまり，二組の染色体の相同な位置で DNA が切断され，その前後で組み換えられて新しい染色体ができる．このようにして，できた新しい染色体が，その生物がおかれている環境に適応している場合は，新しい染色体は増殖し，全体として進化していく．

進化的計算は，生物の進化を計算機上で模倣して，適応，学習，最適化などの機能の実現を目的とする手法である．進化的計算の研究では，Holland が中心となって研究してきた遺伝的アルゴリズム（genetic algorithm; GA），Fogel が中心となって研究してきた進化プログラミング（evolutionary programming; EP），Rechenberg，Schwel が中心となって研究してきた進化戦略（evolution strategy; ES）があげられる．それぞれ特徴として，GA は，個体間で遺伝情報を交換する交叉演算を用いて新しい探索点を生成し，補助的に交叉を用いる手法である．EP は，個体集合と選択を用いることについては GA と共通しているが，新しい探索点の生成には突然変異を用い，交叉を用いない手法である．ES は，実数関数の最適化を対象とし，個体表現には実数値をそのまま用いる手法である．また，GA の拡張として不定長の遺伝子表現を木構造を用いて表す遺伝的プログラミング（genetic programming; GP）や，計算機上で生命的な現象の創発を目的とした人工生命（artificial life; A-Life）などの研究も EC に含まれる．以下 GA を例に挙げて説明をする．

1989 年に Goldberg によってアルゴリズムの枠組みが整理された [17]．遺伝情報は，



(a) Solution space and string space

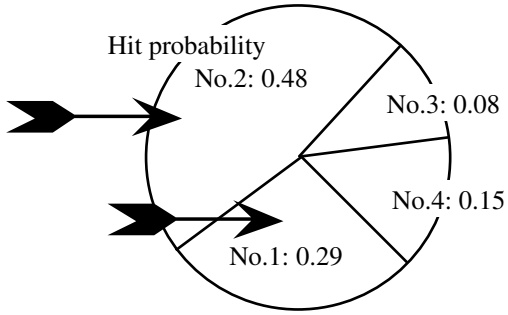


(b) Procedure of GA operation

Fig 2.12 Concept of genetic algorithm.

遺伝子型（genotype）で染色体が保持する．その染色体を個体（individual）といい，個体が多数集まることによって集団（population）をなす．この集団において，遺伝的操作（genetic operation）を行うことによって，次世代に環境（問題）に対する適応度の高い個体を増やし，集団全体を環境に適応できる方向へと進化させる方法論である．簡単に GA の手順を示す．

- Step 0：初期化（initialization）：ランダムに個体を生成し初期個体集団 $P(0)$ を構成し世代 $t = 0$ とする．最終世代を $t = T$ と設定する．
- Step 1：評価（evaluation）：集団 $P(t)$ 内の個体について，その適応度 g を計算する．
- Step 2：選択（selection）：集団 $P(t)$ に選択演算子を適用し，$P'(t)$ を生成する．
- Step 3：交叉（crossover）：$P'(t)$ に交叉演算子を適用し，$P''(t)$ を生成する．
- Step 4：突然変異（mutation）：$P''(t)$ に突然変異演算子を適用し，次世代の集団 $P(t+1)$ を生成する．
- Step 5：判定：$t < T$ ならば $t = t + 1$ としてステップ 1 へ．そうでなければ計算終了．ここで集団内の最大適応度の個体が準最適解となる．



$$P_{selection} = \frac{fitness_i}{\sum_{j=1}^{n} fitness_j}$$

Fig 2.13 A roulette wheel selection.

各遺伝的操作（選択，交叉，突然変異）は，次のような操作である．選択：集団の中から交配するためのペアを選ぶための確率的操作．選択により集団が適応度の高い個体群に収束していく様子を比喩的に選択圧力（selection pressure）と言う．この選択圧力の違いが解の収束性と多様性に影響を与える．選択演算子として，適応度に比例して選択確率を振り分けるルーレットホィール選択（roulette wheel selection）（図 2.13）や最大適応度の個体を交叉や突然変異の対象とせず無条件で残すエリート保存戦略（elitism），適応度に基づく順位付けを行うランキング選択法（ranking selection），個体群から適当な数の個体を抽出し，その中から 1 個体を選択するという操作を数回繰り返すトーナメント選択法（tournament selection）などがある．交叉：2 つの親個体の間で染色体を部分的に組み換えることにより新しい子個体を生成．GA の探索の主推進力となる．交叉演算子には，一点交叉（one point crossover）や多点交叉（multi-point crossover），マスクパターン（mask pattern）を用いた一様交叉（uniform crossover）などや，問題依存で多数の交叉方法がある（図 2.14）．突然変異：染色体上のある確率で選ばれた遺伝子座の値を他の対立する遺伝子に置き換える．交叉に対して補助的な役割を果たすが，突然変異率が高すぎれば個体の遺伝子情報を無視してランダム探索になり，逆に低ければ，探索空間が交叉による組み合わせによる空間内しか探索できない．突然変異演算子には，交換（exchange）や挿入（shift），逆位（inversion）などがある（図 2.15）．

GA の特徴は，最適化問題において最適化変数と評価関数さえ存在すれば，適用が可能であり，高い汎用性をもっていることである．また，比較的少ない計算量で（準）最適解を効率よく求めることができる．GA の応用例としては，スケジューリング問題，画像復元問題，図形配置問題，ロボットの軌道生成，通信ネットワークにおける送信回路発見問題，ニューラルネットワークの最適化，機械学習への適用など多岐にわたる．

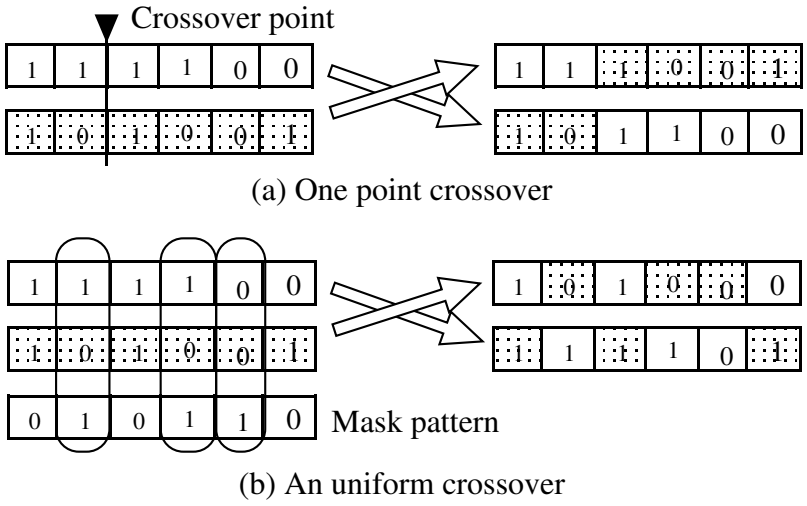


Fig 2.14 Crossover operations.

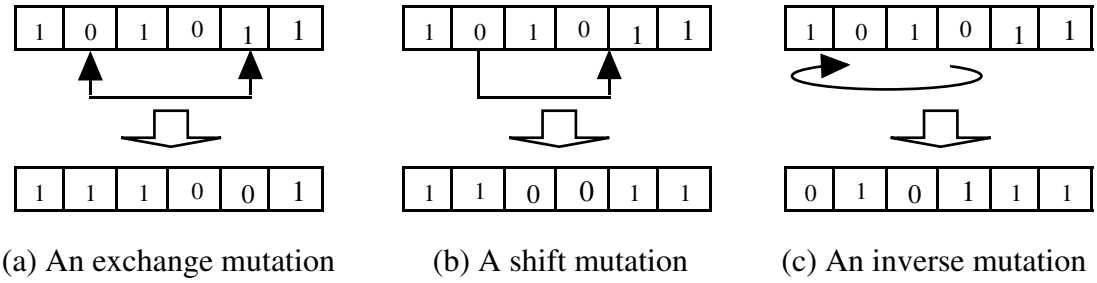


Fig 2.15 Mutation operations.

連続世代モデル

一般によく使われている進化的計算手法は，離散世代交代モデルを仮定したものが多い．離散世代交代モデルは，すべての個体が一斉に子孫を作り，次世代の個体群として入れ替わる．そのため，生成された個体の数に比例して遺伝的操作と評価が行われ，結果的に膨大な計算時間を必要とする．また，今の世代における良い親個体から，次世代に多くのより良い子個体を生成される可能性があるが，逆に，今までよりも評価値の劣る子個体も多く生成する可能性もあるという問題がある．この問題を，ロボットの制御則の獲得に関して考えると，一旦，実行可能な制御則を獲得したにも関わらず，次の世代では，新たに学習し直さなくてはならなくなる可能性が生じる．

一方，部分的な世代交代を可能とした連続世代交代モデルが提案されている [20–22]（図 2.16）．DeJong は，一世代に操作する個体数を generation gap として分割し，部分的に個体群を入れ替える手法を提案し，比較実験を行っている [21]．また，Syswerda は，一世代に少数個体生成し，最小適応度の個体を淘汰する連続世代交代モデルであ

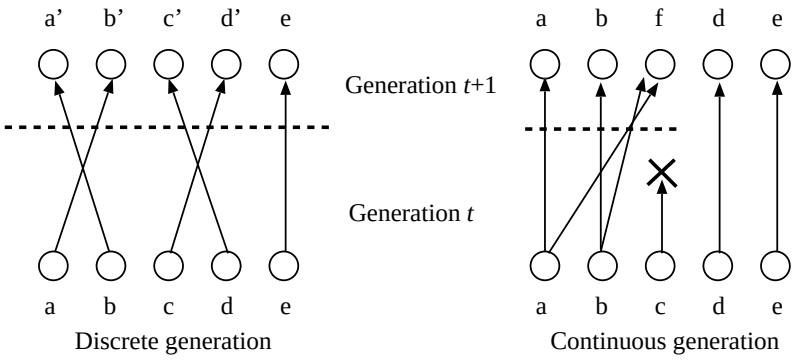


Fig 2.16 Different generation strategy between discrete and continuous one.

る定常状態遺伝的アルゴリズム（steady-state genetic algorithm; SSGA）提案している [22]．SSGA は，実行可能な（より良い評価の）個体を失うことなく，新しい個体を生成させることができるという利点があり，実時間探索が必要なロボットの制御則獲得などに有効であると考えられる．（ただし，連続世代交代モデルにおいても，エリート保存戦略を用いることで，実行可能な解を失うことを防ぐことができる．）また，より良い個体が生成された時に，すぐにその個体を親個体とした子個体の生成が行え，無駄な探索（評価）時間を減らすことができる．さらに，一個体の評価に基づいて，交叉率や突然変異率を変えることができ，局所探索と大域探索のトレードオフを制御でき，評価の構造が未知な問題に対し，より適応的な探索ができうる．

対話型進化的計算

進化的計算手法は，教師データや明示的な報酬構造が与えられない問題に対して，望ましい性能を評価関数として与えるだけで，それを満足する解を探索できる．しかし，設計者とは別の使用者が存在するようなシステムの最適化に関しては，使用者固有の主観的な評価を考慮する必要がある．そこで，人間の主観的評価に基づいてシステムを最適化させる技術として，さまざまな対話型進化的計算（interactive evolutionary computation; IEC）が提案されてきた [23–25, 27–30, 89, 102, 111]．基本的には，図 2.17 に示す通り，個体群の遺伝的操作によって生成された解候補の評価を，人間の主観により行う計算手法である．

IEC の研究は，1986 年の Darkins の Biomorph というフラクタル図形の対話的生成手法が始まりと言われている [23]．その後，人工生命の分野における動植物の形態形成や，顔などの線画や表情生成などのグラフィックアートへの応用として，IEC の研究が数多く行われてきており，最近では，Unemi がメロディの作成に IEC を適用してい

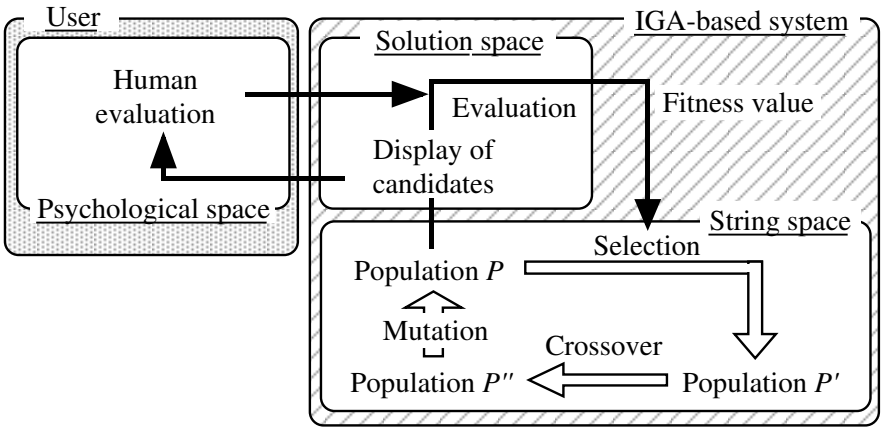


Fig 2.17 Concept of interactive evolutionary computation.

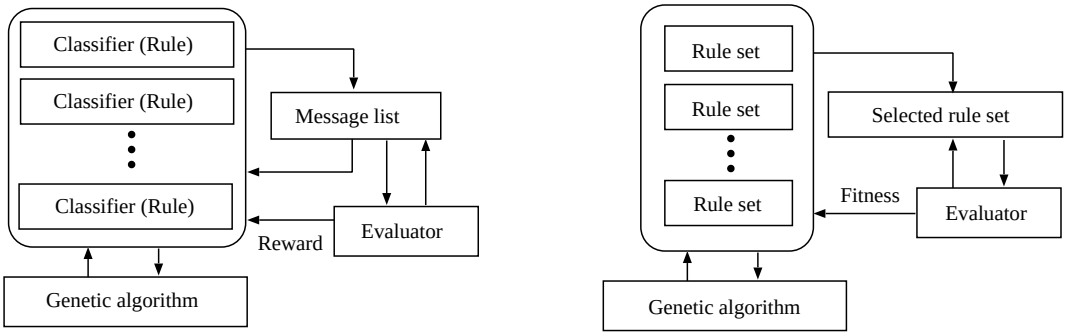
る [23–26]．また，アートの延長として，3 次元 CG ライティングや，吊り橋や車のデザイン支援などにも IEC は適用されている．また本来，使用者の主観的評価がされるべきモノへの応用として，例えば，補聴器の感度調節 [24] や，看護婦の勤務表作成 [30]，人間共存型ロボットの手渡し軌道生成 [89,102,111] など多岐にわたる．さらに，直接人間とは関係しないが，最適化にかかる時間を短縮するために人間のヒューリスティックを導入した進化的探索手法も提案されている．例えば，移動ロボットの制御則獲得 [27] や，オフィスの機器配置 [28]，医療画像の強調化 [29] など．

IEC は，上記のようなシステムが持つ多峰性や多目的性に対して，主観的な評価（選択）を行うことで，解集団を早く最適領域へ収束させることができるという利点がある．その反面，操作者の心的な疲労や，その疲労を軽減するための個体群や世代数の縮小による探索能力の低下といった問題点がある．また，主観的な評価であるが故に，解候補の提示順序に評価が依存してしまうという問題もある．これらの問題に対して，操作者の疲労を軽減するインターフェイスの開発や，IEC とは別に探索を行う EC との融合などが研究されている [23,24]．

2.3.6 遺伝的機械学習

複雑な条件分岐があるような問題や，時系列的な変化を伴う問題などに対して，プロダクションルール集合を自律的に生成する遺伝的機械学習（genetic-based machine learning; GBML）が盛んに研究されている [17,31,32]．

GBML の研究は，1978 年の Holland と Reitman による CS-1（cognitive system one）の開発が最初と言われており，その後，1980 年，Smith による LS-1（learning system one）の開発が行われ，大きく二種類の GBML の流れができた [31,32]．CS-1 の流れは，



(a) Michigan approach　　(b) Pitt approach

Fig 2.18 Genetic-based machine learning.

ミシガンアプローチ（Michigan approach）と呼ばれ，一般には分類子システム（classifier system）として知られている．また，LS-1 の流れは，ピッツアプローチ（Pitt approach）と呼ばれる（図 2.18）．

分類子システムは，プロダクションルール（classifier）を個体，ルール集合を個体群，ルールの信頼度を適応度として一般的な進化的計算手法と対応づけられる．各ルールは，前件部に分類子，後件部にメッセージを保持し，入力情報に対して一致するルールのメッセージを出力する．出力されたメッセージはメッセージリストに送られ，実際に行動出力が生成され，報酬と次の入力情報が得られる．得られた入力情報は，メッセージリストに送られ，各ルールの分類子と比較され，これを繰り返す．分類子は，基本的に {0,1} の 2 値と，ワイルドカードである *don't care*{♯} によ っ て構成される．そのため異なるルール間で同時発火が生じることがあり，信頼度の大きさや分類子の専門性からルール選択を行う手法がとられている．さらに，*Ishibuchi* は，分類子を{0,1,♯} だけでなく，ファジィメンバシップ関数により分類子を構成する手法も開発している [34]．報酬は，行動の度に得られるとは限らず時間遅れが生じうる．そのため，適切に信頼度を更新するために，強化学習で用いられるバケツリレーアルゴリズムや利益共有法などの信頼度割当が適用される．また，ルールの信頼度を適応度として捉え，進化的計算により新しいルールの生成を行う．分類子システムは，一連のメッセージのやり取りから順序関係があるようなタスクにも適用でき，マルチエージェントなどの研究に盛んに用いられている [14,35,36]．

一方，ピッツアプローチは，ルール集合を 1 個体とし，複数のルール集合を個体群とする．そのため，計算の規模としては大きなものとなるが，その他は，進化的計算手法とかわらず，解候補の適応度をそのままルール集合全体の評価値として用いることができ，さまざまな問題に適用が容易である．

頻繁に報酬が得られるようなタスクに対しては，ミシガンアプローチが適していると考えられるが，移動ロボットのナビゲーションなど，一つのシナリオが長く，報酬がなかなかもらえないようなタスクでは，ピッツアプローチを用いてゴール後にルール集合を評価する方がよいと考えられる．二つの GBML は，強化学習に代表される局所的な学習と，進化的計算に代表される大域的な学習の典型的な手法であり，互いの短所を補うべく，融合された手法の開発が望まれている．

2.4 ロボットの制御構造と学習機構

2.4.1 古典的人工知能とロボティクス

古典的人工知能における代表的なロボットは，プロダクションシステムにより構成された組み立てロボットや加工ロボット，検査ロボットなど，主に生産現場で用いられてきた．知識工学の進展により，専門家の知識体系が計算機上に再現できるようになり，簡単な繰り返し作業や，特定のタスクに特化した機能を実現することが，比較的容易に達成でき，工業の発展に大きな影響を与えた．しかし，専門家がもつ知識は，記号的に記述可能な部分と，数値や自然言語で表現できないものが存在し，ロボットの汎化性や適応性が問題となった．センサやアクチュエータなどの要素技術の発展により，これらの問題は改善されてきたが，やはり問題空間が限定された範囲での話である．

古典的 AI に基づくロボットの制御プロセスについて簡単に説明する．まず，作業対象や環境からのセンサ入力（sensing）が行われ，外界の物体･事象をモデリング（modeling）する．そのモデルに基づいて，行動計画（planning）を行う．次に，計画を実行すべきタスク群に分解し実行（task execution）する．最後に，分解した動作群を順次駆動系へと伝達する（motor control）[1]．この制御方法では，あるプロセスで計算時間を多く費やしたり，停止した場合，それから先の制御プロセスを行うことができなくなる．例えば，モデリングの段階で，センサ情報に基づき外界を認識しようとした時，対応する知識が保持されていなければ，モデリングすることはできずタスクの計画はもちろん，なにも行動を起こすことができなくなる．古典的人工知能によるロボットは，限定された問題空間で，かつ時間的拘束がない場合であれば，十分に対応できる．しかし，実環境では，問題空間が膨大であり，かつ有限時間内での行動が要求されるため，古典的 AI によるロボットは機能できない可能性がある．また知識表現においても，実環境ではフレーム問題が生じる．フレーム問題とは，関係のあるもの同士は少数なのに，多数の無関係を記述しなければならない問題である．実環境では知識を分類するフレームが無数に存在し，それらを統べて記述することが不可能である．また，ある

作業目的が与えられたときに，必要か不要かすべてのフレームに対して吟味しなくてはならない．フレーム問題と同様に，実世界で与えられた行為が有効であることを保証する環境を定義しにくいという制御記述問題や，行為に関する暗示的な結論の数が増える分岐問題なども考慮しなくてはいけない．これらの問題の発生原因は，AI が環境と切り離した形で知的システムを構築しようとしたためである．

2.4.2 行動に基づくロボティクス

古典的人工知能によるロボットからのパラダイムシフトとして行動に基づく人工知能（behavior-based AI）がある．その中心となる行動に基づくロボティクス（behavior-based robotics）について説明する [1, 41, 43].

行動に基づくロボティクスは，従来の設計者によるトップダウン的な記述を避け，処理態系の並列化，統合化をはかり，さらには環境との相互作用をはかり，いかに行動させるかを議論してきた．代表的なアルゴリズムに R.Brooks のサブサンプション・アーキテクチャ（subsumption architecture; SSA）がある [40]．Brooks は，古典的 AI を用いて外界をセンサで認識し，そのモデルを内部に構築し，行動計画をたてて，実際に行動を起こす手法が，実環境で動くには困難であることを示唆し，古典的 AI によるロボティクスとは全く異なる手法を提案した．従来の通り機能ごとにモジュール分割するのでなく，ロボットにかせられた課題を達成するための行動ごとに分割することを提唱した．図 2.19 は移動ロボットの例（サブサンプション・アーキテクチャ）である．知覚から実際の動作までをすべての機能を独立に持たせるようにし，並列的に稼動させ，それらの調停により行動をおこす．それにより，ある行動が失敗した場合でも別の行動で補うことができる．また，環境に対するモデルをロボットの内部に持たないために，フレーム問題に陥ることを回避した．行動に基づくロボティクスの特徴として，段階的発達，知的行動の出現，頑健性・適応性が挙げられる [41, 43].

1. 段階的発達：それぞれの行動をモジュール化して扱うため，行動の追加が容易に行うことができる．さらに，その追加により，行動表現が拡大し，知的レベルの発達に繋がる．

2. 知的行動の出現：基本的な行動モジュールにより構成されているにも関わらず，環境に応じた行動を素早くおこすことができるため，高度に知的な行動を行っているように見える．その結果，従来のロボットで扱われていたセンシングやモデリング，プランニングなどの機能が，環境変化にあわせて行動した結果の中に出現すると考えることができ，従来の機能モジュールが必要ではないことを示唆する．

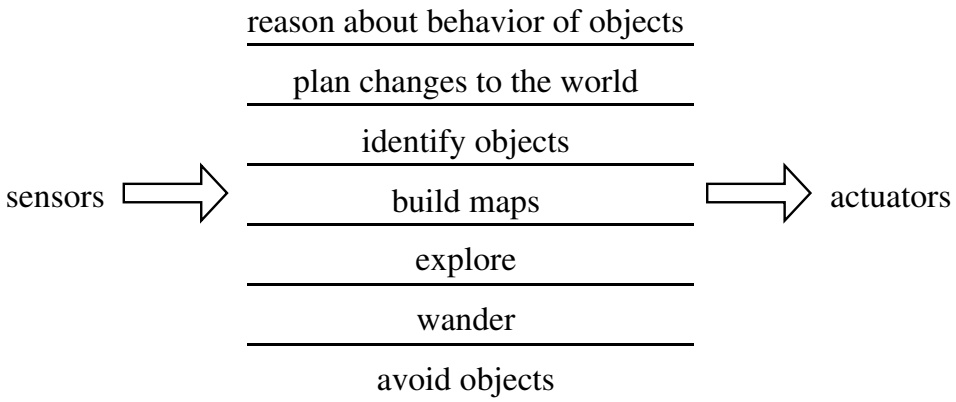


Fig 2.19 Concept of behavior-based robotics.

3. 頑健性・適応性：複雑な実環境をモデリングする必要が無くなるため，行動モジュールの頑健性が高い．また，ある行動モジュールが動作しなくなった場合でも他の行動に影響を与えることがない．

Brooks は，自身の方法論をロボティクスから拡張して，行動に基づく AI として，新しい AI 技術について次のように考察している [41].

1. Situatedness：人工知能は記号化された抽象世界ではなく，現実の世界を対象としなくてはいけない．現実の世界をそのままモデルとして扱うべきである．

2. Embodiment：身体のない知性は，現実に存在しないように，身体がなければ，知能が実現できない．

3. Intelligence：外部とのやり取り（環境との相互作用）により，知的な行動が出現する．

4. Emergence：知能は，どこか一ケ所にあるものではなく，システムの構成要素の相互作用によって創発されるものである．

これらから，認知や行動の主体であるロボットが環境に適応するには，自らの身体により，自らの行為により環境を認知し，行動を起こさなければいけないということを示している．異分野において，哲学者 D.C.Dennett もまた「活動に意味を与えてくれるのは身体である」と述べている [95].

分割統合戦略に基づくロボティクス

目的別に基本行動を設計し，それらを選択／統合する枠組みは，行動に基づくロボティクスだけでなく様々な分野で研究されている．基本的に，これらのスタンスは，難し

い問題を部分問題に分割して解こうという分割統合戦略（divide-and-conquer strategy）に基づいている．制御の分野であれば，マルチコントローラのスイッチング制御，NN の分野であれば，混合エキスパート（mixture of experts）やアンサンブル学習などがあげらる．

分割統合戦略に基づくロボティクスは，大きく二つのアプローチがある [38, 55–61]．一つは，行動切り替え型のアプローチであり，行動モジュールの排他的な選択を行う．もう一つは，行動融合型のアプローチであり，行動モジュールからの出力を融合し動作を行う．行動切り替え型のアプローチとして，Maes は，前述したサブサンプションアーキテクチャの行動間の優先順位を動的に変更するために，行動モジュール間をネットワークで表現し，その中の行動ノード間の活性／抑制により行動を選択する手法を提案している．Zhang らは，目標探索や追従，障害物回避，移動障害物回避により目標値を計算し，ファジィコントローラにより出力する手法を提案している．Bonarini らは，ファジィ if-then ルールを用いて行動モジュールを切り替えているが，ルール条件部に，CANDO（できること）と WANT（したいこと）という関係を導入し，選択の多様性を高める手法を提案している．Pirjanian は，複数の目的に対して構築した行動の選択を，多目的最適化問題として扱う手法を提案している．

行動融合型のアプローチとして，渡辺らは，あいまい行動型制御として，ファジィコントローラによる行動モジュールの構成と，階層的な抑制関係を拡張し，出力方向が同一の場合は動作出力を行動出力の重み付け和で計算する手法を提案している．Saffiotti は，Context-based blending という，ファジィ If-then ルールで構成された調停則による融合手法を提案しており，次ぎに取るべき動作のための調停則を言語的に設計している．

これら分割統合戦略に基づくロボットシステムは，個別の機能としての頑健性や，新しいタスクや環境に対する拡張性に関しては，単一モジュールによるロボットシステムよりも優れているが，各モジュールの設計が個別になされており，モジュール間による一連の動作への影響を考慮していないものが多い．行動の再利用や洗練を考えた場合，それぞれのモジュールが，他のモジュールの機能に依存して，再構成（再学習）する枠組みがなければ，機能だけ肥大になり，使用者が望む性能が発揮できない可能性がある．

複数機能モジュールを持つロボットシステムにおいて，ボトムアップ的に，モジュール間の関係に整合性をつけるようなアプローチとして，川人らによる MPFIM [62–64] や，久保田らによる知覚–行為循環に基づくモジュラー型 NN [88] などがある．MPFIM を強化学習に拡張した MMRL は，連続値を扱う Actor–Critic のモジュールを複数用意したモデルであり，価値の見積もりに依存したモジュール選択や融合を行っている [64, 65]．

価値の見積もり方によって，モジュールの専門性が空間的な状態分割により行われる手法 [65] や，時間的な状態分割により行われる手法 [64] が提案されている．知覚–行為循環に基づくモジュラー型 NN は，知覚–行為モジュールとして NN を用い，動作出力と同時に環境予測を行っている．行動知識からの抽出した入出力関係を複数のモジュールに学習させ，予測誤差が大きくなると特定のモジュールに遷移する手法である．これにより順序関係があるような動作が可能になる．

2.4.3 進化的ロボティクス

行動に基づくロボティクスで示唆されている，環境との相互作用により制御器を獲得する手法として，進化的ロボティクス（evolutionary robotics; ER）がある [42]．進化的ロボティクスは，Cliff らにより提案され，AI，ロボティクス，認知科学，社会学にいたる多くの分野の研究者により，近年盛んに研究されている [42, 66–68, 70]．基本的に，ニューラルネットワークを制御器として用い，その学習に，進化的計算を採用している．ネットワークの結合強度の組を個体とし，目的ごとの評価項目を用意し重み付けした評価関数を用いることで，多目的な動作を獲得することが可能である．例えば，Floreano は，小型移動ロボットを用いて，センサ情報を入力として衝突せずに動き続ける行動を獲得させたり，スパイキング・ニューラルネットワークを導入し，模型飛行船や飛行機が衝突せずに室内を飛び続ける制御器を獲得されたりしている [69]．近藤らは，動的再編成機能を有する神経回路モデルを提案しており，箱押しタスクにおける必要な複数の動作をネットワークの構造を変えることで獲得している [49, 50]．中村らは，把持できる障害物をその障害物の周りを回転し識別する行動の獲得手法を提案しており，ノイズを含む実環境においても，識別能力が頑健であることを示している [67]．

また，ロボットの制御器の獲得のためだけでなく，Lund らは，LEGO を使った教育的エンターテーメントとしての進化的ロボットを提案している．樋口や Thompson らは，電子回路を進化させる進化的ハードウエアの研究を行っている．手首を切断した患者の筋電情報から，義手を動かす制御器を短時間で獲得することに成功している．

これら，進化的ロボティクスは，要求される性能を得るための評価関数を設定するだけでよく，また，複数の評価項目の重み付け和で評価値を計算すると，多目的な制御器が獲得可能である．しかし，獲得した制御器は，要素行動へと分割不可能であり，再利用しにくく，新たなタスク，新たな環境下で，新たに学習し直す必要がある．ただし，本来，進化的計算で用いている解集団の多様性を利用して，新たな環境やタスクに迅速に対応できる手法が考えられる．先行研究として，知覚に基づく遺伝的アルゴ

リズムが提案されており，環境中の障害物密度が異なる環境下でのナビゲーション問題に用いられている．遭遇した障害物の頻度から，個体群の一部を用いて進化させることで，再び同じような障害物頻度が観測された時に，過去に学習した個体が用いられることで，短時間でよりよい解が発見できることが示されている [72]

2.4.4 認知ロボティクス

環境との相互作用と，身体を考慮したロボティクスには，行動に基づくロボティクスや進化的ロボティクスの他に，最近，認知ロボティクスの研究が盛んに行われている．認知ロボティクスとは，認知科学，発達心理，神経科学などの複雑な生命システムの原理を探究する分野から多くのヒントを得て，ロボットを基にした仮説生成と検証を通して，認知・知能システムに対する新しい方法論を確立しようとするものである [47].

認知ロボティクスで，特にキーワードとされているのが「環境との相互作用」と「身体性」である．行動に基づくロボティクスにおいても，それらのキーワードは用いられていたが若干問題がある．それは，環境との相互作用で想起される行動が，私達観測者が与えたフレームで記述していることである．ロボットは環境の状態に応じて設計者が考えた行動モジュールを，優先順位付けされた包含関係により選択している．実環境は問題空間が膨大であると記述したが，それは決して無構造を意味するわけではない．環境が持つ情報は非常に膨大で，かつそのなかで行動する主体にとって重要な情報もまた膨大である．そして，その環境の中で行動する主体が存在することにより，相互の関係に構造が与えられる．それにより，環境が持つ無限の相互作用は，主体と環境との間にある物理的関係や，主体の内部状態，時空間的な文脈により，環境から得ることができる情報，言い換えると，とりうる行動の可能性であるアフォーダンス（Affordance）[94] が限定され，環境と主体との相互作用は拘束される．その相互作用を抽象化し簡単なフレームを与えたものが，行動に基づく AI である．人間の認知や行動はそのようなフレームで表現できるものではなく，さらなる議論が必要となっている．しかし，行動に基づく AI が，従来の AI 手法からの脱却に大きな影響を与えていることには変わりない．認知ロボティクスのアプローチは幾つかあり，行動に基づく AI が環境客体のボトムアップ的アプローチとしたら，認知主体のトップダウン的アプローチとの密な相互作用によるものであると議論されている．谷は，リカレントニューラルネットワーク（recurrent neural network; RNN）とホップフィールドネットワークを用いて，ロボットの認知モデルを構築し，RNN がおこなう予測機能における安定相と不安定相の遷移の関係に，自己意識的な現象が存在するのではないかと議論して

いる．また，浅田らは，身体性とは何か？，相互作用とは何か？という問題に関して，テストベットとしてロボカップ（Robo Cup）を設け，サッカーに限定した中から，まずロボットの認知を考えようという試みを行っている．

認知ロボティクスが盛んに議論され始めたのは，認知を議論するためのツールが揃ってきのと同時に，認知に関する議論の必要性がニーズとして生じためであると考えられる．科学哲学による概念と，種々の学問の融合により，認知ロボティクスは今後発展していくと考えられる．

2.4.5 構造化知能に基づくロボットシステム

知的なシステムを構築しようとした場合，実在する知的なシステムから多くのヒントを得ることができる．実環境の中に存在し，行動する主体として生物を挙げることができる．生物は，個々の細胞が各々の役割を遂行し，情報やエネルギーを相互に伝達・共有することによって生きている．個々の細胞は，特殊な行為や知的な行為を起こしているわけではなく，全体として知的な振る舞いを見せる．つまり簡単な機能モジュールの相互作用により知的な振る舞いが可能となる．知能は，複雑なものからの発現と捉えるのではなく，簡単な機能モジュールで構成する構造の相互作用によって発現すると考えることが重要である．また，知的な主体として自然界に活動するには，機能モジュール間の相互作用だけでは活動不可能であり，環境との密な繋がりも考慮しなくてはいけない．重要なことは，知的な行為の要因が環境の中に存在するということである．より高度な知能を有するシステムほど，環境の中に存在する多くの行為の可能性（アフォーダンス）をピックアップできると考えられる．

生物において，最も知的な振る舞いをしているのは，我々人間であるといわれる．我々の脳のメカニズムのモデルとして，

1. 生理学的モデル：神経系による脳の表現
2. 心理学的モデル：一般化された技量や知識
3. 認知科学的モデル：再帰的な意識

などが挙げられ，構造化知能（structured intelligence）は，これらのモデルを統合することにより高度な知能を実現することが目標である [82]．構造化知能を持つロボットシステムは，対象となるシステムのハードウエアとソフトウエアにより実現される基本的な機構において，それら機構の密接な相互作用により生じる知的な振る舞いにより実現する．構造化知能は，各々の部分が独立に最適化されるのではなく，構造的に相互作用することによって，全体として，高度な知能を獲得していくものである．文

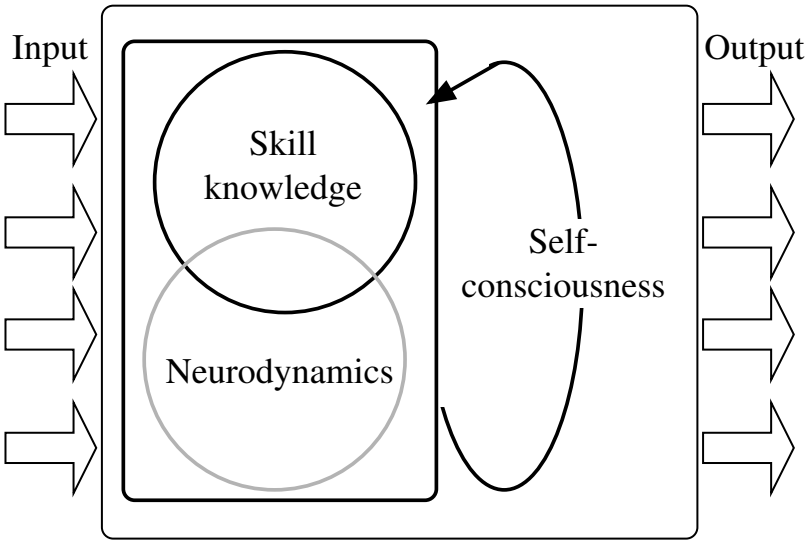


Fig 2.20 Concept of structured intelligence.

献 [88] による報告では，人間–ロボット環境下で，ロボットが環境（人間の接触パターン）に出会うことで，使われた行動知識をその接触パターンに合わせて学習することで，行動知識からコミュニケーションに必要となる知識を抽出し，行為モジュールを学習する手法を提案している．行動知識はファジィルールで構成され，行為モジュールは階層型 NN で構成される．行為モジュールを複数用意し，予測誤差に基づきモジュール間を遷移させることで，行動知識が複数の行為として再構成される．このように環境（人間）との相互作用によって個々のモジュールを構造化する枠組みが構造化知能である．

2.5 結言

本論文で扱う環境の定義を行い，それぞれの環境下で用いるべき学習手法について検討した．定義されている環境の属性について説明し，それらの属性を考慮し，4 つの環境を定義した．4 つの異なる環境条件では，使用できる学習手法と，使用できない学習手法があり，問題のクラスに応じた学習手法の構造化が必要であることを示した．また，周辺研究について古典的人工知能から行動に基づくロボティクス，認知ロボティクス，構造化知能によるロボティクスと説明を行った．本研究で用いる多目的行動調停は，基本的に行動に基づくロボティクスによる目的毎に設計された行動モジュールの融合ではあるが，基本行動間の補完的な役割や行動の再利用のための，基本行動や行動調停則の学習により拡張を行う．これらの拡張の背景には，認知ロボティクスや

構造化知能に基づくロボティクスにおける環境との相互作用や機能モジュールの構造化などの概念がある．

3章で，静的未知環境における移動ロボットのナビゲーション問題に対して，進化的計算手法とデルタルールによる基本行動の学習手法を提案する．4章では，移動障害物を含む動的環境下におけるナビゲーション問題について，局所エピソード学習による行動調停則の適応学習を議論する．5章では，パートナーロボットの構築に関して，人間–ロボット環境を対象とし，対話型進化的計算やNNなどを用いた基本行動獲得と行動調停について議論する．

第3章　未知環境における移動ロボットの基本行動獲得

3.1　緒言

近年，移動ロボットの適用領域が，静的既知な環境から未知な環境に拡大しようとしている．例えば，惑星調査や災害地における被災者の探索などでは，ロボットは事前に環境情報を取得することができず，観測した情報をもとに自律的に移動（タスク遂行）できなければならない．日常の生活環境や工場においても，新しい場所や家具や機械の突然の配置換えに対応しなければならない．また，これらの環境では，同時に満たさなければならない複数の目的（障害物回避，目標追従，探索，採集など）が，存在する．

環境モデルを用いた従来の経路計画は，静的既知な環境では最適な経路を探索することができうるが，未知な環境に対しては大域情報が取得不可能であり適用が困難である．この問題に対して，Brooksが提案したサブサンプション・アーキテクチャ（subsumption architecture; SSA）をはじめとする行動に基づくロボット（behavior-based robotics）は，未知な環境において，タスクを達成するために必要な行動を目的毎に用意し，反射的かつ排他的に切り替えることで，センサ情報のみを用いて，ロボットにとって未知な観測状態に対応できる．しかし，反射行動であるため，複数の目的を同時に達成する必要がある環境では，タスクを達成できないことがある．また，基本行動の設計や行動の重層関係の設定は，設計者が事前に環境をある程度知っている必要がある．さらに，排他的な行動の切り替えは，一連の動作を断続的にする．これらの問題点から，未知環境でタスクを達成できるロボットは，多目的かつ滑らかな動作が必要であり，おかれた環境に適するように基本行動を学習できなければならない．そして，未知環境におけるナビゲーション問題を考えた場合，強化学習で用いるような報酬の構造を事前に与えることができず，タスク遂行時に得られる評価（移動時間や移動距離）によって学習が行われるべきである．

本研究では，移動ロボットのための制御手法として，基本行動（basic behavior）の出力を融合する多目的行動調停（multi-objective behavior coordination; MOBC）を提案し，その有効性を基本行動獲得を通して検討する．各基本行動は，ファジィルー

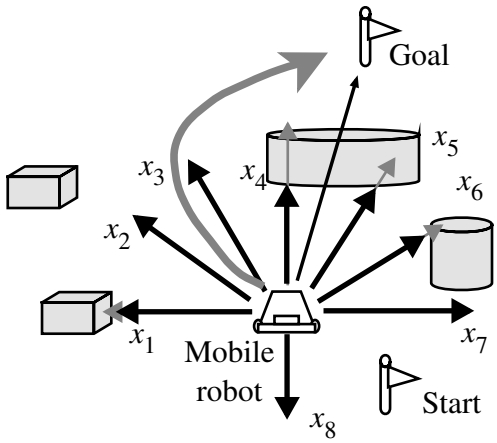


Fig 3.1 A mobile robot in unknown environment including static obstacles.

ル（fuzzy rules）とプロダクションルール（production rules）で表現し，それらルール条件部の組み合わせ最適化と制御出力のチューニングを進化的計算（evolutionary computation; EC）及びデルタルール（Delta-rule）を用いて学習することで，ロボットをおかれた環境に適応させる．本章では，これら提案手法の有効性を，計算機シミュレーションにより，学習速度の改善と未学習環境での実行可能性の観点から検証する．

3.2 移動ロボットの基本行動

移動ロボットは，障害物との距離を計測する複数の距離センサと，ゴール方向を認識するための複数の光センサを装備するものとする（図 3.1）．また，移動機構として独立二輪駆動を仮定し，センシング毎に速度と操舵角を計算し実行することで，障害物が存在する環境下において与えられたタスクを行う．適用環境は，複数の障害物が固定配置された環境とする．ただし事前に障害物の配置情報を知らない静的未知な環境とする．

移動ロボットには，未知環境下で障害物との衝突なしになるべく早く，なるべく最短距離でゴールに向かうというタスクを与える．本研究では，このタスクを達成するために，3 つの基本行動を想定する．

まず，距離センサに基づく基本行動として，障害物回避（collision avoiding）行動と壁面併走（wall following）行動を，簡略型ファジィ推論（simplified fuzzy inference）を用いたファジィコントローラ（fuzzy controller）により構築する．前件部への入力情報を距離センサ情報 $X(t)(x_i(t); i = 1, 2, ..., n)$ とし，後件部の出力を移動ロボットの行動出力 $Y_k(t)$=$(y_1(t), y_2(t))$ とする．行動出力は，速度 y_1 と操舵角 y_2 で構成される．各入力

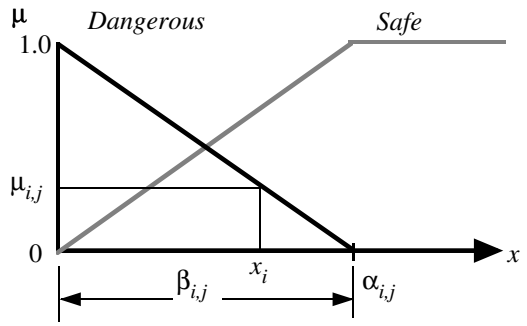


Fig 3.2 Triangular membership functions for condition parts of fuzzy rules.

x_i に対する言語ラベルの組み合わせと理想出力値から，次のようなファジィIF-THENルールを構築する．

$$IF\ x_1\ is\ F_{1,j}\ and\ x_2\ is\ F_{2,j}\ and\ ...\ and\ x_n\ is\ F_{n,j}$$

$$THEN\ y_1\ is\ b_{j,1}\ and\ y_2\ is\ b_{j,2} \tag{3.1}$$

ここで，$F_{i,j}$ は，$i\ (i = 1, 2, ..., n)$ 番目の入力と $j\ (j = 1, 2, ..., m)$ 番目のルールに対するメンバーシップ関数を表し，$b_{j,r}(r = 1, 2)$ は後件部出力パラメータである．

メンバーシップ関数には,「危険」と「安全」の 2 つの言語ラベルを用い，三角型メンバーシップ関数（図 3.2）で表現する．各入力に対するメンバーシップ関数の適合度と，それらを組み合わせた j 番目のルールの適合度は次式により求める．

$$\mu_{F_{i,j}}(x_i) = \begin{cases} 1 - \dfrac{|x_i - \alpha_{i,j}|}{\beta_{i,j}} & |x_i - \alpha_{i,j}| \leq \beta_{i,j} \\ 0 & otherwise \end{cases} \tag{3.2}$$

$$\mu_j = \prod_{i=1}^{n} \mu_{F_{i,j}}(x_i) \tag{3.3}$$

ここで $\alpha_{i,j}$ と $\beta_{i,j}$ は，それぞれメンバーシップ関数の中心値と幅を表す．そして，ファジィルールからの出力 $y_r\ (r = 1, 2)$ は，

$$y_r = \frac{\sum_{j=1}^{m} \mu_j b_{j,r}}{\sum_{j=1}^{m} \mu_j} \tag{3.4}$$

となる．

次に，環境の複雑さに合わせてメンバシップ関数の形状を更新するセンサリネットワーク（sensory network）を導入する．環境の複雑さとは，移動ロボットが遭遇する

障害物の数や間隔に依存するものと考える．センサリネットワークは，ロボットの観測センサ情報に従って，各センサの注意レンジ X_{att} を動的に変化させる手法である．注意レンジは，メンバーシップ関数の幅 $\beta_{i,j}$ に相当し，各センサの注目すべき計測情報の範囲を表す（図 3.3）．注意レンジの拡縮則は，

$$X_{att}(t) = \zeta(t) X_{rng} \tag{3.5}$$

で与える．X_{rng} はセンサレンジであり，計測できる最大距離を表す．また，$\zeta(t)$ は，$[\zeta_{min}, 1.0]$ で値をとるセンサレンジの拡縮率とする．ζ_{min} は最小拡縮率である．拡縮率の更新は次式で行う．

$$\zeta(t+1) = \begin{cases} \gamma^{-1}\zeta(t) & if\ all\ x_i \leq X_{rng}(t) \\ \gamma\zeta(t) & otherwise \end{cases} \tag{3.6}$$

γ は，減衰係数 $(0.0 < \gamma < 1.0)$ である．図 3.3 において，距離 x_1 と x_2 との関係が，メンバシップ関数の形状に依存して大きく異なることを，μ_1 と μ_2 の値の違いから確認することができる．つまり，センサリネットワークを用いない場合では，偏りのある入力データに対して，状態を細分する多くのメンバーシップ関数を必要とし，多くのファジィルールを記述する必要が生じる．そのため環境が複雑になるほど記述すべきルール数が増加する．しかし，センサリネットワークを用いることで，観測される環境の状態に合わせて，注意レンジが動的に拡縮され，メンバーシップ関数の形状の更新を行うことができ，適切に入力情報間に差異を付けることができる．それにより，メンバシップ関数の少なさによる状態空間の粗い分割が補われる．また，センサリネットワークに依存して，基本行動の出力値をスケーリングすることで，障害物が密集し狭い空間では近傍に注意を払いゆっくりと移動し，逆に障害物が近くにあまりなく広い空間では，遠くに注意を払いながら速度を上げて移動できる（図 3.4）．

このファジィコントローラを用いて，障害物回避行動と壁面併走行動を実現する．障害物回避行動は，感知した障害物からなるべく離れようとする行動であり，また，壁

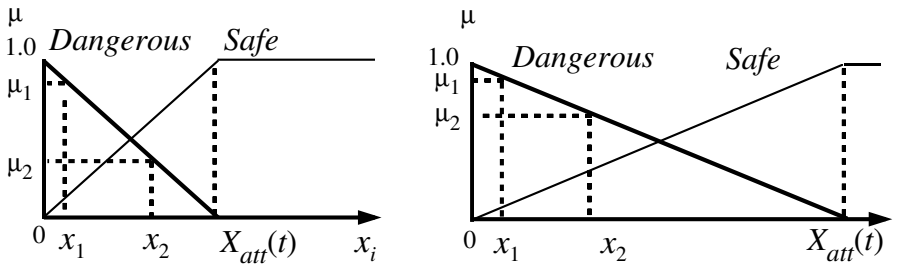


Fig 3.3 Change of triangular membership functions by using sensory network.

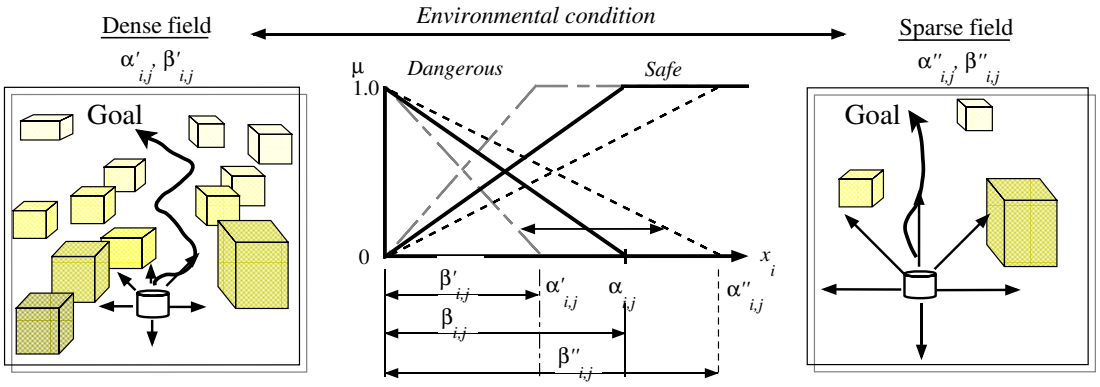


Fig 3.4 Different attention range X_{rng} according to the sparseness of facing obstacles. ($\beta = X_{rng}$)

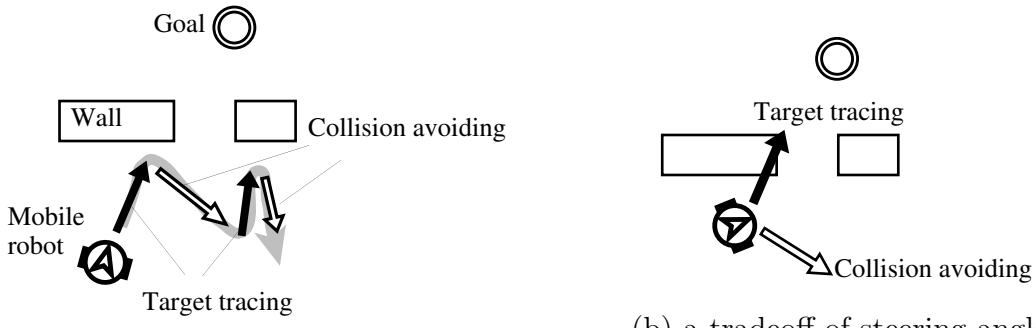
面併走行動は，側面に感知した障害物（壁）を沿うように，壁面と距離を保つような移動を行う．そのため，同じ距離情報を入力されても，出力される行動は異なる．

次に，目標追従（target tracing）行動の構成方法について説明する．目標追従は，ゴールを光源と仮定し，光センサ情報を用いてゴールに向かう行動とする．障害物回避行動と同様に，ファジィコントローラでも構成できるが，今回はよりシンプルに，プロダクションルールにより構成する．ゴール方向と自機の進行方向の差を減らすように操舵角を計算し，操舵角が大きくなればなるほど速度を落とすことで，無駄な大回りを防ぎつつ，ゴール方向に向かって滑らかに方向転換を行う．

基本行動は，ファジィルールやプロダクションルールで構成することができるが，必要とする入出力関係が得られるのであれば，ニューラルネットワークや非線形振動子なども用いることができる．しかし，入出力関係が言語的に（直感的に）分かりやすく，ヒューリスティックが導入しやすい，ファジィルールやプロダクションルールを本研究では採用する．

3.3 移動ロボットのための多目的行動調停

従来，経路計画問題として扱われてきた環境は，既知環境であり，障害物の角（頂点）を結ぶ補助線を引くことでグラフ探索問題として扱ってきた [37]．また，環境のマップ情報からポテンシャル場を仮定しポテンシャルの低いところへ移動する手法が使われてきた [37,39]．しかし，我々は環境の属性すべてを考慮して移動しているわけでも，全体的なレイアウトを考慮しながら動いているわけでもなく，未知な環境において移動できうる．



(a) zigzag action by switching behaviors

(b) a tradeoff of steering angles between different behaviors

Fig 3.5 Problems of conventional behavior-based robotics.

未知環境における代表的な行動制御手法であるサブサンプション・アーキテクチャ[40]を用いたロボットの動作は，観測センサ情報と一対一に対応した基本行動に基づいている．基本行動はレイヤーとして並列に階層付けがされており，ある観測状態の条件が成り立てば上位階層の行動が下位の行動を抑制する．ただし，行動の切り替わりは，環境の状態に依存するため断続的であり，動作間に滑らかさがない．そのため，図 3.5 の (a) のような環境下では，ジグザグな動作が生成され，タスクが達成できない可能性がある．また，行動の学習に関する議論はあまりなされていない．

それに対して，おかれた環境下で，単一の行動モジュールを獲得する進化ロボット（evolutionary robotics; ER）が盛んに議論されており，ロボットの身体的な特性に適した制御器の獲得手法が数多く提案されている [42]．しかしながら，同じセンサ情報に対して相反する動作出力が必要となるタスクにおいて，基本的に ER は適用困難である．例えば，図 3.5 の (b) のような場合,「目的地に向かう」ための動作出力と,「障害物を避ける」ための動作出力との間にトレードオフが生じ，単一の学習器を用いるだけでは行動の学習が困難である．

では，私達はどのように行動しているのか？

未知な環境において，人間は種々の基本的な行動に優先順位をつけ，排他的に選択するということはせず，その基本的な行動の複合により複雑な行為を行なっていると考えられる．それは我々が環境に対して常に多目的であるためである．例えば，移動に関して,「目的地に向かいながら人を避ける」というような「目的地に向かう」と「人を避ける」といった 2 つの行動を同時に行っている．そして，行動間に質的なトレードオフが存在する場合（「目的地に向かう方向」と「人を避ける方向」が相反する場合）においても，現在の状態と最近とった行為に依存して，適切な行為を行っている．ま

たその一連の行為は，反射的ではなく滑らかである．

ロボットも我々人間と共存する環境においては，常に多目的な動作が必要とされるであろうし，環境が複雑であれば，それに対応できる柔軟な動作が必要とされる．しかしながら，ただ単に複雑な動作を学習しようとした場合，実環境に近付けば近付くほど，必要な状態量が無限定に増大し，学習が不可能になる．

生態心理学（ecological psychology）[94] において，J.J.Gibson は,「私たちは，動くために知覚するが，知覚するためにはまた動かなければならない」といっている．このことから，知覚と行為の間にある非分離な構造が指摘され，さらに,「知覚が行為を制約し，行為が知覚を制約する」という下りが象徴するように，それぞれの構造が相互に制約しあうものとして存在すると指摘している．ある姿勢は，見える範囲を制約し，その見えと姿勢に基づき行為を取る．取った行為により新しい姿勢が生成され見え方が制約さる．この一連のプロセスには，文脈的な知覚と行為の依存関係が存在すると考えられる．このことから，本研究では，ある動作により制約された知覚情報によって，ある程度トップダウン的に次の動作を制約する枠組み（制御構造）が必要であると考える．

本研究では，移動ロボットの多目的な動作を実現するために，基本行動を調停する制御構造を提案する．まず，移動ロボットの基本的な行動をファジィコントローラで構築し，センサリーネットワークによる知覚情報を入力として速度と操舵角を各ファジィコントローラから出力する．そして，各基本的行動に対し行動重み（behavior weight; W）（$W(t) = (w_1(t), w_2(t), ..., w_K(t))$）を与え，各基本行動の出力に重み付けを行う．環境情報 $X(t)$ に対する各基本行動の出力結果を $Y_k(t)(k = 1, 2, ..., K)$ とすると，行動調停した結果，動作出力 $A(t)$ は，以下のようになる．

$$A(t) = \frac{\sum_{k=1}^{K} w_k(t) Y_k(t)}{\sum_{k=1}^{K} w_k(t)} \tag{3.7}$$

K は，基本行動の総数とする．この重み付け平均された出力に基づき，移動ロボットは動作を行う．また，行動重み $W(t)$ を更新することで，環境の変化に対応した動作を行う．行動重みの更新には次式を用いる．

$$w_k(t+1) = \frac{w_k(t) + G_k(X(t))}{\sum_{k=1}^{K} (w_k(t) + G_k(X(t)))} \tag{3.8}$$

ここで，$G_k(X(t))$ は k 番目の行動の環境情報 $X(t)$ に対する重み更新関数であり，知覚と動作の間の文脈を考慮したプロダクションルールで構成される．

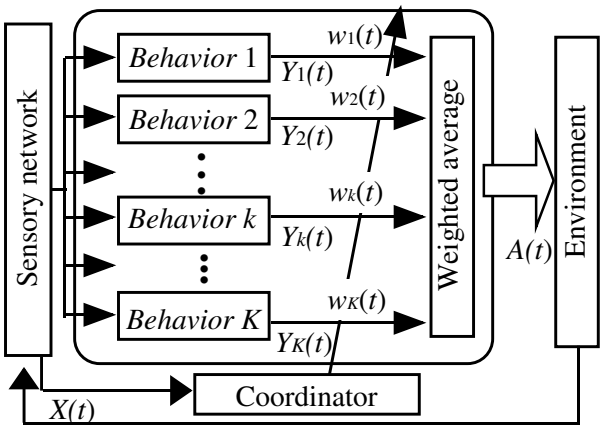
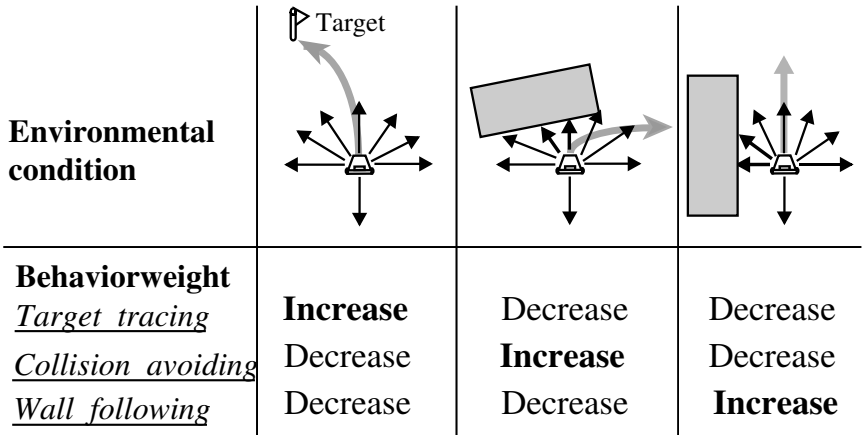


Fig 3.6 Concept of a mobile robot with multi-objective behavior coordination.



Environmental condition			
Behaviorweight			
Target tracing	**Increase**	Decrease	Decrease
Collision avoiding	Decrease	**Increase**	Decrease
Wall following	Decrease	Decrease	**Increase**

Fig 3.7 Examples of behavior coordination rules of MOBC.

3つの基本行動 ($K = 3$)，目標追従行動，障害物回避行動，壁面併走行動を考えた場合，行動重み更新則として,「前方方向の距離センサが感知すれば障害物回避行動の重みを増加させ，他の行動重みを減少させる」であったり「側面の距離センサが反応すれば，壁面併走行動の行動重みを増加させ，他の行動重みを減少させる」といった簡単なルールが構築できる（図3.7）．ここで，逐次行動重みを更新するため，主行動の遷移時に多目的な動作が実現できる．更新ルールの詳細は，実験にて説明する．

3.4 進化的計算とデルタルールによる基本行動の構造最適化と学習

本節では多目的行動調停を用いた移動ロボットの行動獲得手法を提案する．スタートからゴールまで移動する一回の試行を1シナリオとする．1つのシナリオが終わる度にオフライン的な評価をし，定常状態遺伝的アルゴリズム（steady-state genetic algorithm; SSGA）[22] を用いてファジィコントローラの構造最適化を行う．SSGA は，一世代につき最小適応度の極少数の個体を淘汰し，新しい個体を生成する最小の連続世代モデルであり，実行可能な個体を保持しやすい．ここでのオフライン的な評価とは，実行した解候補の評価をシナリオ終了時に逐次行うことをいう．また，1つのシナリオの中で，オンライン学習としてデルタルールを用い，スタートからゴールへ移動している間に，ファジィコントローラの後件部出力値の教師あり学習を行う．

まず，SSGA によるファジィコントローラの構造最適化に関して説明する．基本的に，ピッツアプローチ [31,32] の枠組みで，If-then ルール集合で構成されたファジィコントローラの学習を行う．比較的小規模のルール集合でファジィコントローラを構築するために，個体表現は最大ファジィルール数を固定し，(3.1) 式のファジィルールと，それに対するバリディティ r_{valid} {0:使用，1:無視 } を用いて構成する（図 3.8）．各ルールの各センサ入力 x_i $(i = 1, 2, ..., n)$ に対するメンバーシップ関数の組み合わせを {0：$dangerous$，1：$safe$，2：$don't\ care$} で表現し，出力を実数値で設定する（図 3.8）．

SSGA の手順としては，**（*1*）最小適応度個体の除去**：最も適していない個体を淘汰する．**（*2*）エリート交叉**：最良個体とランダムに選んだ個体とで多点交叉を行い，新しい個体を生成する．**（*3*）突然変異**：生成された個体のルール毎のバリディティに対してビット反転させ，前件部のファジィルールの組み合わせに対してメンバーシップ単位で入れ換えを行う．後件部出力値は，実数値を用いており，次式に示す各世代 s 間の相対的な適応度の差に基づき適応型突然変異を行う．

$$b_{j,r}^{k}(s+1) = b_{j,r}^{k}(s) + \epsilon_r \frac{fitness - minF}{maxF - minF} N(0,1)$$

$$(j = 1, 2, ..., m;\ r = 1, 2) \tag{3.9}$$

ここで，ϵ_r と r は，重み係数と動作出力の個数をそれぞれ表し，$fitness$，$minF$，$maxF$ は，それぞれ遺伝的操作を行った個体の評価値，個体群中の最小評価値，個体群中の最大評価値とする．また，$N(0,1)$ は，平均 0，分散 1 の正規乱数を表す．これら突然変異は，各遺伝子座に対して，突然変異率に従い実行される．**（*4*）評価**：(2)(3) の遺伝的操作を受けた個体を用いてタスクを 1 シナリオ実行する．そのシナリオにおける，ゴールまでの時間ステップ P_{time} と移動した距離 P_{length} に基づき，次式より評価値を計

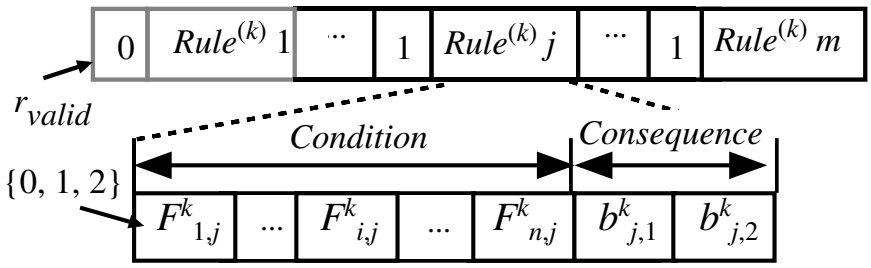


Fig 3.8 Representation of a candidate fuzzy controller.

算する．

$$fitness = \omega_1 P_{time} + \omega_2 P_{length} \tag{3.10}$$

ω_h $(h = 1, 2)$ は各評価に関する重みとする．この (1) から (4) の繰り返しプロセスは，(3.10) 式の最小化問題として帰着される．

次に，デルタルールによるオンライン学習について説明する．基本行動の r 番目の出力信号の総誤差 $E_r(t)$ を，多目的行動調停における行動重みを考慮し，次式のように設定する．

$$E_r(t) = \sum_{k=1}^{2} E_r^k(t) = \sum_{k=1}^{2} \left\{ \frac{1}{2} \frac{w_k}{\sum_{k=1}^{3} w_k} \left(d_r^k(t) - y_r^k(t) \right)^2 \right\} \tag{3.11}$$

ここで，$d_r^k(t)$ は，k 番目の行動に関する教師値である．後件部出力値の更新式は，

$$b_{j,r}^k(t+1) = b_{j,r}^k(t) - \sigma \frac{\partial E_r^k(t)}{\partial b_{j,r}^k(t)} \tag{3.12}$$

とする．σ は，学習率を表す $(\sigma > 0)$．ここで，右辺第二項の偏微分は，

$$\frac{\partial E_r^k(t)}{\partial b_{j,r}^k(t)} = \frac{\partial E_r^k(t)}{\partial A_r^k(t)} \frac{\partial A_r^k(t)}{\partial b_{j,r}^k(t)} = -\frac{w_k(t)}{\sum_{k=1}^{K} w_k(t)} \left(d_r^k(t) - \frac{\sum_{j=1}^{m} \mu_j^k(t) b_{j,r}^k(t)}{\sum_{j=1}^{m} \mu_j^k(t)} \right) \frac{\mu_j^k(t)}{\sum_{j=1}^{m} \mu_j^k(t)} \tag{3.13}$$

となるため，(3.12) 式は次式のようになる．

$$b_{j,r}^k(t+1) = b_{j,r}^k(t) + \sigma \frac{w_k(t)}{\sum_{k=1}^{K} w_k(t)} \left(d_r^k(t) - \frac{\sum_{j=1}^{m} \mu_j^k(t) b_{j,r}^k(t)}{\sum_{j=1}^{m} \mu_j^k(t)} \right) \frac{\mu_j^k(t)}{\sum_{j=1}^{m} \mu_j^k(t)} \tag{3.14}$$

ここで，(3.14) 式に示されるように，行動重みが学習係数をスケーリングする．教師値は，図 3.9 と 3.10 に示すように生成される．障害物回避行動の教師値は，障害物の密

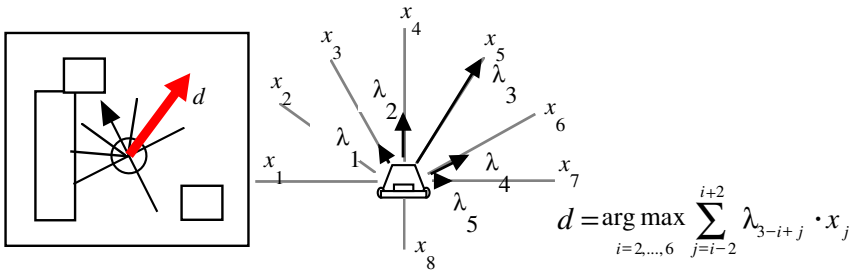


Fig 3.9 Teaching direction for collision avoiding behavior according to the state evaluation of the sensing direction x_i by weight coefficients: $\lambda_i = \{0.1, 0.3, 1.0, 0.3, 0.1\}$.

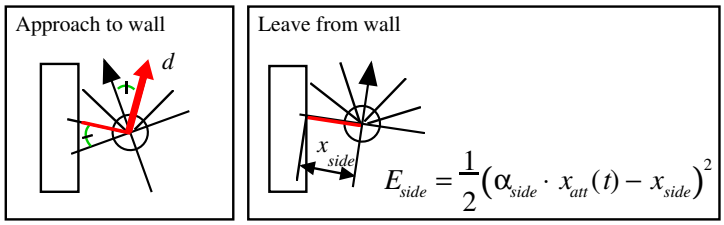


Fig 3.10 Teaching direction and error value for wall following behavior. Left: teaching direction to avoid wall. Right: error value to keep distance.

度が粗な方向とする．具体的には，2 近傍のセンサを考慮した 5 つのセンサ重み付け平均を計算し，もっとも値の大きなセンサの方向を教師方向とする．壁面併走行動では，ロボットが壁に近づく場合と，遠ざかる場合で教師値が異なる．前者は，計測センサ情報が最も短いセンサと，側面のセンサとのなす角を進行方向に対する回避角として教師値にする．後者の場合において，側面のセンサのみが反応する場合，適切な教師角度が計算できないため，教師値を用いず，側面のセンサの距離を一定に保つように誤差関数を構成し学習を行う． これらの教師値は，理想的な出力として設定するが，ロボットの身体性の観点から，真に適切であるかは解らない．そこで，学習係数を世代毎に減少させることで，学習初期のみにデルタルールを用い，実行可能な基本行動の迅速な獲得を試みる．このように，行動獲得には，理想的なルール構造が事前に設計できない部分に対しては SSGA を適用し，理想的な出力値を学習できる部分に対してデルタルールを適用することで，おかれた環境に適応できるファジィコントローラの獲得を行う．

3.5 計算機シミュレーション

計算機シミュレーションの設定条件について説明する．ロボットのサイズを直径 10[pixel] とし，環境を 500[pixel] × 500[pixel]，距離センサの注意レンジ X_{att} は，

$$30[pixel] \leq X_{att} \leq 90[pixel]$$

とする（図 3.1，図 3.11）．ただし X_{att} の初期値は 60[pixel] とする．距離センサの個数は 8（$n = 8$）とする．基本行動は目標追従行動 $(k = 1)$，障害物回避行動 $(k = 2)$，壁面併走行動 $(k = 3)$ の 3 種類をとる $(K = 3)$．各基本行動の出力は，移動ロボットの速度 v と操舵角 $\Delta\theta$ とする．ロボットにはスタートからゴールまで障害物に衝突することなしに最短移動時間，最短移動距離で移動することをタスクとして与える．行動重みの変化量 $G_k(X(t))$ は，表 3.1 の条件を満たす箇所の足し合わせにより計算される．ただし，x_i は $[0, 1]$ に規格化され，0 は接触，1 は無反応を意味する．例えば，真正面のセンサ x_1 のみが反応した場合，目標追従行動の行動重みに 0.035 が加算され，障害物回避行動には 0.2 が加算される．逆に壁面併走行動の行動重みは 0.15 だけ減じられる．

本シミュレーションでは次項に関して実験を行う．

1. 基本行動獲得における動作の滑らかさ
2. 複数未知環境における基本行動獲得の適応速度
3. 未学習環境における実行可能性

1 に関しては，多目的行動調停則を状況の知覚モジュールとして考えた場合における知覚と行動の接続に関して，単一未知環境における学習を通して検証する．2 に関しては，多目的行動調停の制御構造の有効性を検証するために，調停則を従来の切り替え型とした移動ロボットを用意し，複数未知環境における行動獲得を比較して，提案手法の有効性を検証する．3 に関しては，2 で獲得した基本行動を用いて，未学習環境において多目的行動調停と行動切り替え型とで，どのように異なる特徴があるかを検証する．

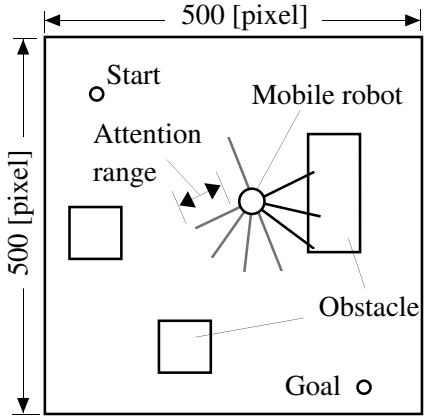


Fig 3.11 Simulation environment including static obstacles.

Table 3.1 Setting parameters of the behavior coordination rules. ($L = 0.95$).

G_k \ X	$x_4 < L$	$x_3, x_5 < L$	$x_2, x_6 < L$	$x_1, x_7 < L$	$x_1 \sim x_7 > L$
1	0	0	0	0	+0.005
2	+0.2	+0.15	+0.1	-0.05	0
3	-0.15	-0.1	+0.05	+0.1	0

3.5.1 基本行動獲得における動作の滑らかさ

多目的行動調停則は，基本行動をどのように調停するかをルール化したものであり，直面する環境の状態を解釈してトップダウン的に調停していると考えることができる．本節では，行動獲得によって，この多目的行動調停に基づく移動ロボットがどのような基本行動を獲得するか？また，どのような動作を生成できるようになるか？を検証する．

各基本行動（障害物回避，壁面併走）に対して解候補個体を各 50 個用意し，SSGA の 500 回評価とオンラインでのデルタルールにより行動獲得を行った．ただし，ファジィコントローラの初期化は全くランダムに行うものではなく，ある程度のヒューリスティックに基づいて設定し，それに摂動を加えることで，初期個体を生成している．

図 3.13 にロボットがおかれた環境を，図 3.12 に 500 回評価における評価値の履歴を示す．図 3.13 の (a) は，学習初期の軌道であり，ゴールに到達すらできず障害物に衝突していることが確認できる．図 3.13 の (b) は，500 回評価後の軌道であり，ゴールに到達でき且つ，滑らかに移動できているように見える．このことは，図 3.12 に示す評価

値の推移からも確認でき，500 回評価を通して，なるべく早く，なるべく最短距離で移動するための基本行動が獲得できていることがわかる．次に，図 3.13 の (b) に示される軌道における行動重みの変化を図 3.14 に示す．基本的に突然センサレンジ内に入ってくる障害物に対して障害物回避行動の行動重みを増加させ，ある程度回避しつつ壁面併走行動の行動重みを増加させることで障害物に沿うように移動する．そして，環境の中頃にある壁面に対しては，高い壁面併走行動の行動重みで調停することで，ゴールに向かいつつ壁面を沿うような多目的な動作が生成されている．

この獲得された基本行動と多目的行動調停の接続に関してさらに検討するために，学習初期の基本行動と学習後の基本行動を，図 3.15 に示す単純な環境で比較する．それぞれの行動重みの変化を図 3.16 に示す．非常に簡単な障害物の配置のため，学習初期においても基本的にゴールまでの移動は達成されているが，学習後の軌道と大きく異なることが確認できる．中盤の壁面に対して，学習初期のロボットは，壁面併走行動と目標追従行動の行動重みを交互に増加させ，その結果として図 3.15 の (a) に示される通り，ジグザグの動作になった．これは，側面のセンサが反応しつづけることにより壁面併走行動の行動重みが増加する一方，壁面併走行動が移動ロボットの運動特性や環境の特徴に適していないため，壁面併走行動が異なる行動（ここでは目標追従）をすぐに接続しようとするためである．逆に，行動獲得後のロボットは，壁面併走行動が獲得されているため，壁面併走行動の行動重みが大きな時には，壁面に沿うように動作を生成し，見え方を一定に保つことで，壁面併走を継続させている．この結果として，壁面に対してほぼ真っすぐに並進することができ，無駄な操舵が削減され滑らかな移動が可能になっている．

このように，多目的行動調停は，トップダウン的に簡単な文脈を与える手法であり，進化的計算による行動獲得を通して，その文脈に適した行動を獲得できると考えられる．また，多目的行動調停は，動作の多様性を保持しつつも，その状況で必要であろう多様性を限定し，知覚できる情報と動作との間に整合性を保てることが明らかになった．

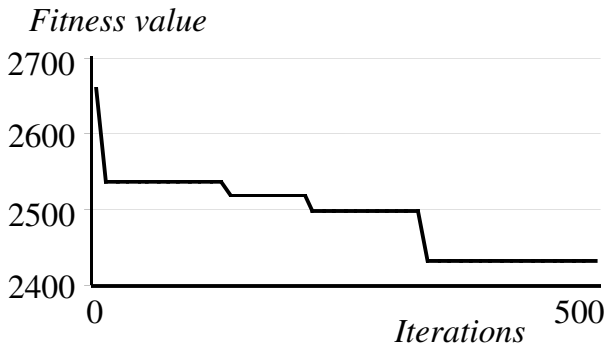
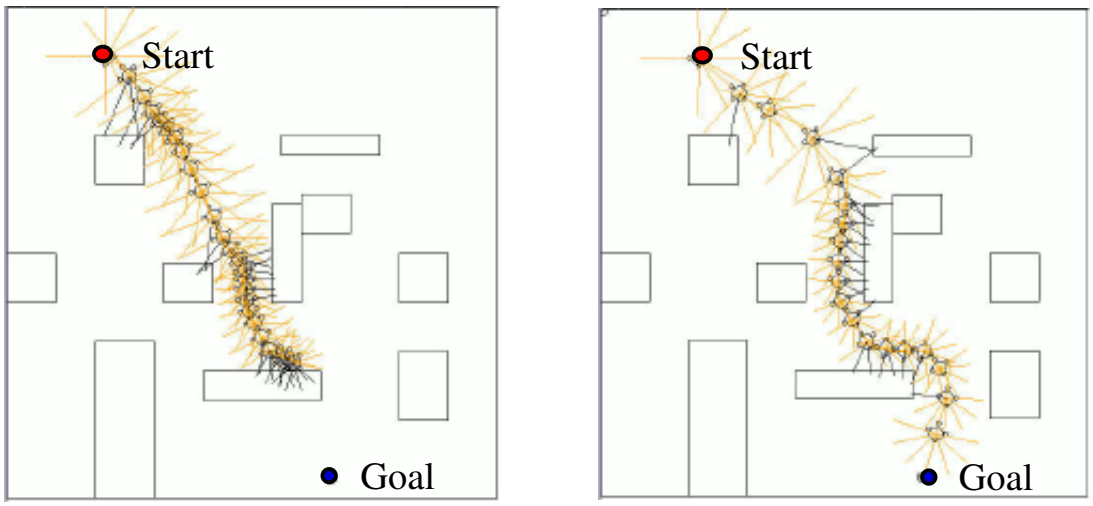


Fig 3.12 History of fitness value



(a) First iteration (b) Final iteration

Fig 3.13 The trajectories of a mobile robot at first and final iterations.

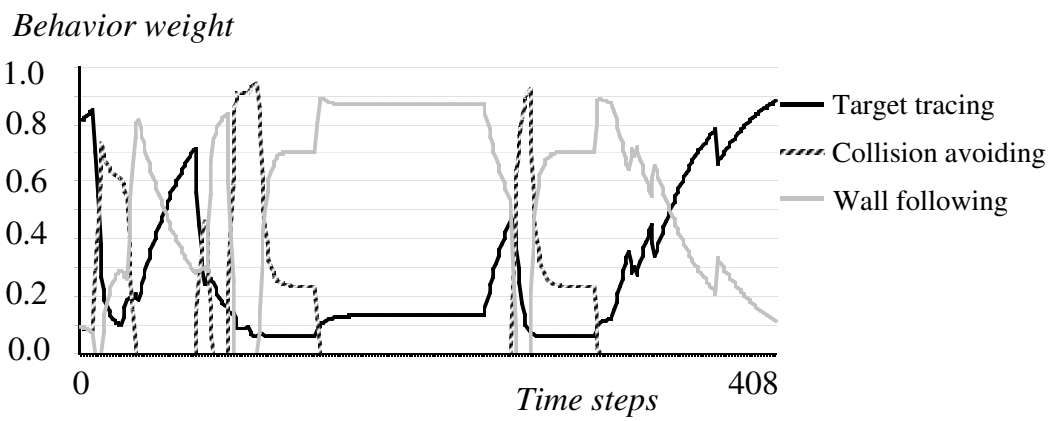
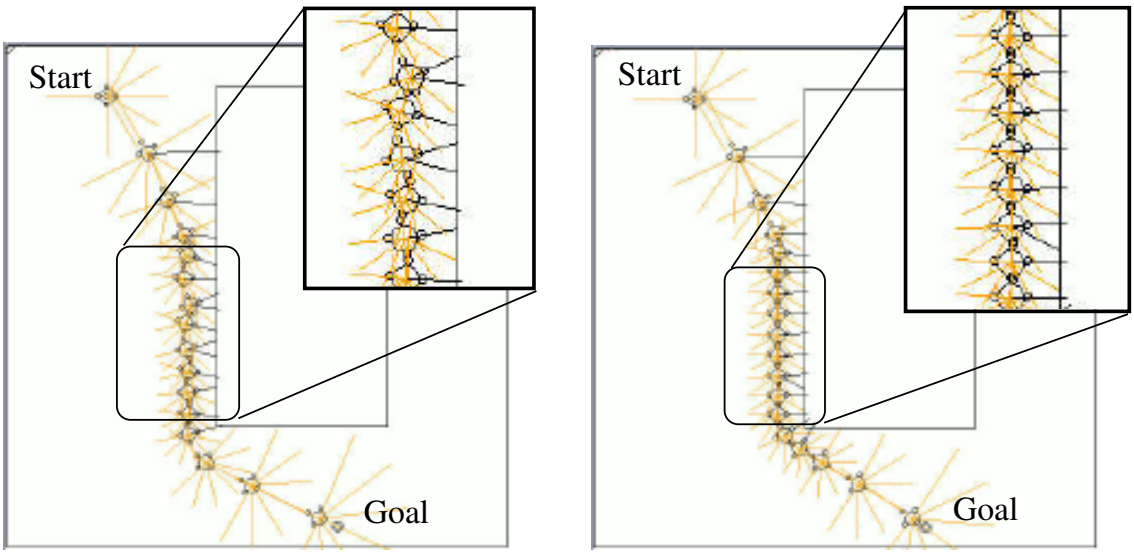
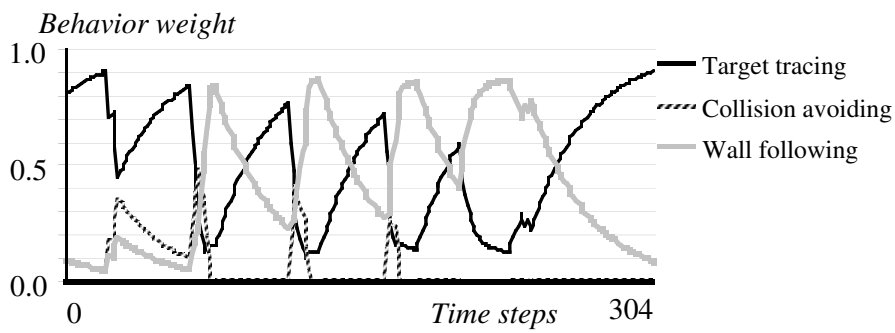


Fig 3.14 Change in behavior weights of the robot with acquired behaviors.

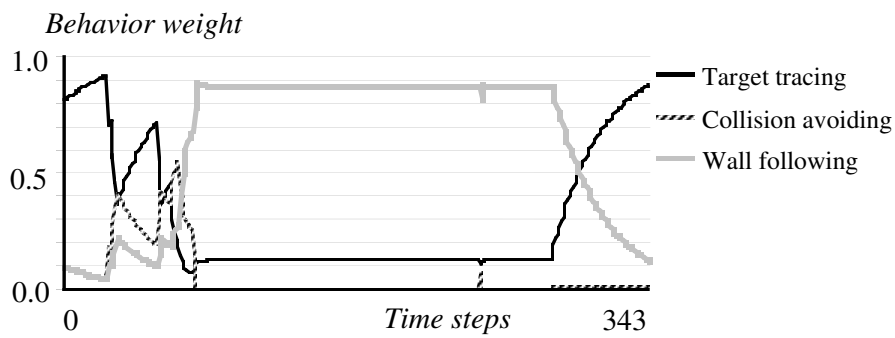


(a) using initial behaviors (b) using acquired behaviors

Fig 3.15 Comparision of trajectories between initial and acquired behaviors.



(a) using initial behaviors



(b) using final behaviors

Fig 3.16 Change in behavior weights between initial and acquired behaviors.

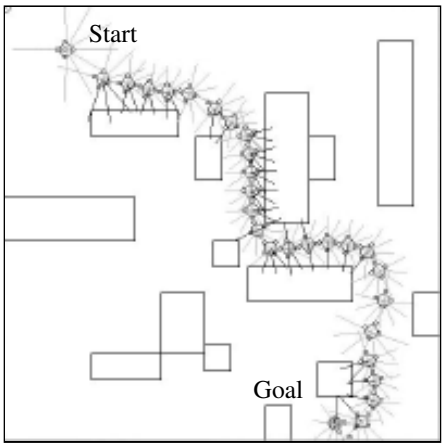
3.5.2 複数未知環境における基本行動獲得の適応速度

提案手法の有効性を，従来の優先度固定の行動調停則（切り替え型）を用いた移動ロボットの行動獲得と比較し検討する．障害物配置が異なる 5 つの環境において，ロボットにタスクを実行させ，解候補ファジィコントローラを評価する．各環境での評価値の総和を解候補ファジィコントローラの評価値として用いる．SSGA のパラメータ設定は，個体数を 50，評価回数を 2000 回，突然変異率を各遺伝子座に対して 0.01 とし，各行動調停則に対してそれぞれ 10 回シミュレーションを行う．比較対象とする行動切り替え型は，2 タイプ用意する．一つは，目標追従，障害物回避，壁面併走の 3 つの基本行動を切り替える SSA3，もう一つは，目標追従と障害物回避の 2 つを切り替える SSA2 である．構造をシンプルにするために，これら二つの行動切り替え型手法は，サブサンプション・アーキテクチャにならい，三行動の場合であれば，前方に障害物があれば障害物回避行動を選択し，前方に障害物がなく側面のみ反応する場合に壁面併走行動を選択する．そして，どのセンサも障害物を感知しない場合，目標追従行動を排他的に選択するものとする．二行動の場合は，障害物を感知するしないで，障害物回避行動と目標追従行動を切り替える．

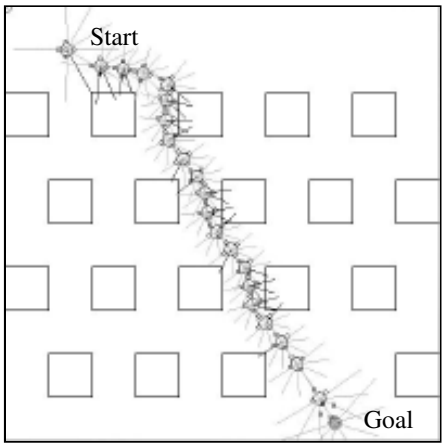
図 3.17〜図 3.20 に，シミュレーション結果を示す．図 3.17 は，実験を行った 5 つの環境の障害物配置を示し，移動ロボットの軌道は，多目的行動調停を用いた行動獲得の最終世代のものを表す．シミュレーションに用いた環境は，異なる特徴をもつ障害物配置になっており，例えば，環境 2（図 3.17(b)）であれば，細かな切り返しが必要であったり，環境 5（図 3.17(e)）であれば，ネズミ返しを乗り越える必要があったりする．2000 回評価を行うことで，どの行動調停則を用いたロボットもタスクが達成された．図 3.18 に，各手法における評価値の履歴を示す．これは，2000 回評価のシミュレーションをそれぞれ 10 回づつ行ったときの平均を表す．多目的行動調停を用いたロボットが行動切り替え型のものと比べて収束が早く，より良い評価値が得られていることが確認できる．

図 3.19 と 3.20 に環境別でまとめた評価値の履歴と，総操舵角の履歴を示す．総操舵角を，シナリオ実行中における操舵角の絶対値の総和と定義し，少ないほど無駄な操舵が無く，滑らかに移動していることを表す．ただし，この評価指標は，SSGA の評価関数に含まれておらず，行動獲得で得られた基本行動に基づく動作が，どれだけ滑らかなものであるかを比較するためだけに用いる．図 3.19 と 3.20 より，どの環境においても，多目的行動調停を用いることで，動作に滑らかさがでていることが確認できる．評価値の履歴に関しても，およそどの環境においても多目的行動調停を用いた方が，収束が早いことが確認できる．特に，環境 5 に関して，行動切り替え型と多目的行動調停の評価値が大きく差が出ていることから，ネズミ返しを含むような複雑な環

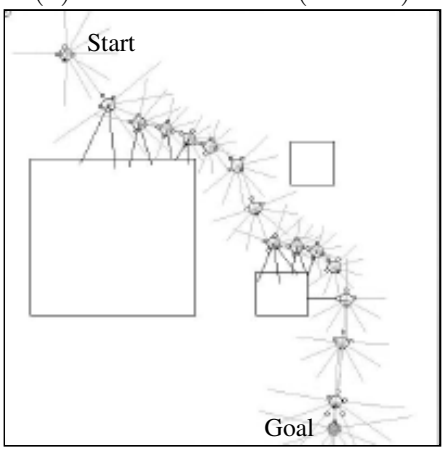
境において，より多目的行動調停の有効性が発揮される．これは，多目的行動調停における動作の多様性の影響と考えられる．逆に，環境 2 のような，小さな障害物が密集する環境下においては，最終的に三行動切り替え型より評価値が劣っている．これは，多目的行動調停における基本行動間の滑らかな主行動の遷移故に生じる，主行動の遷移時間の遅れが原因だと考えられる．



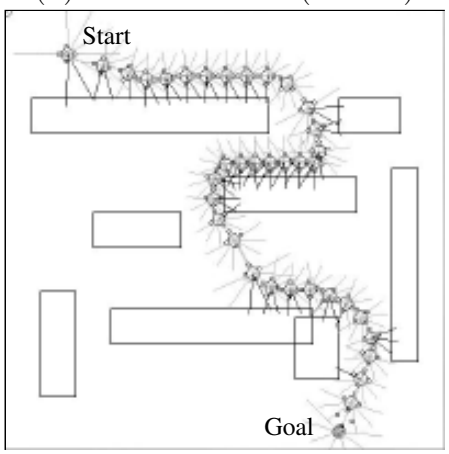
(a) Environment 1 (Env. 1)



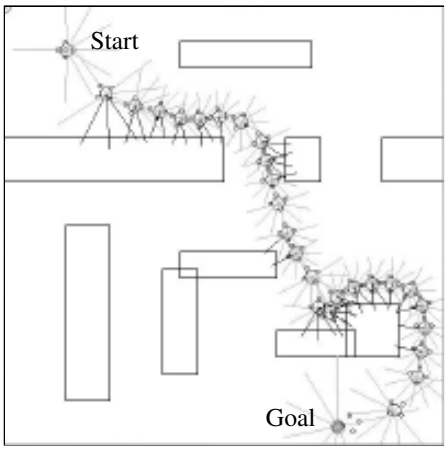
(b) Environment 2 (Env. 2)



(c) Environment 3 (Env. 3)



(d) Environment 4 (Env. 4)



(e) Environment 5 (Env. 5)

Fig 3.17 Trajectories of a mobile robot at final iteration.

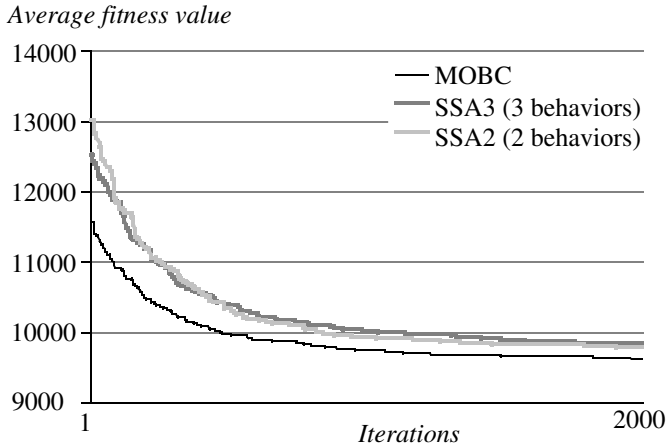
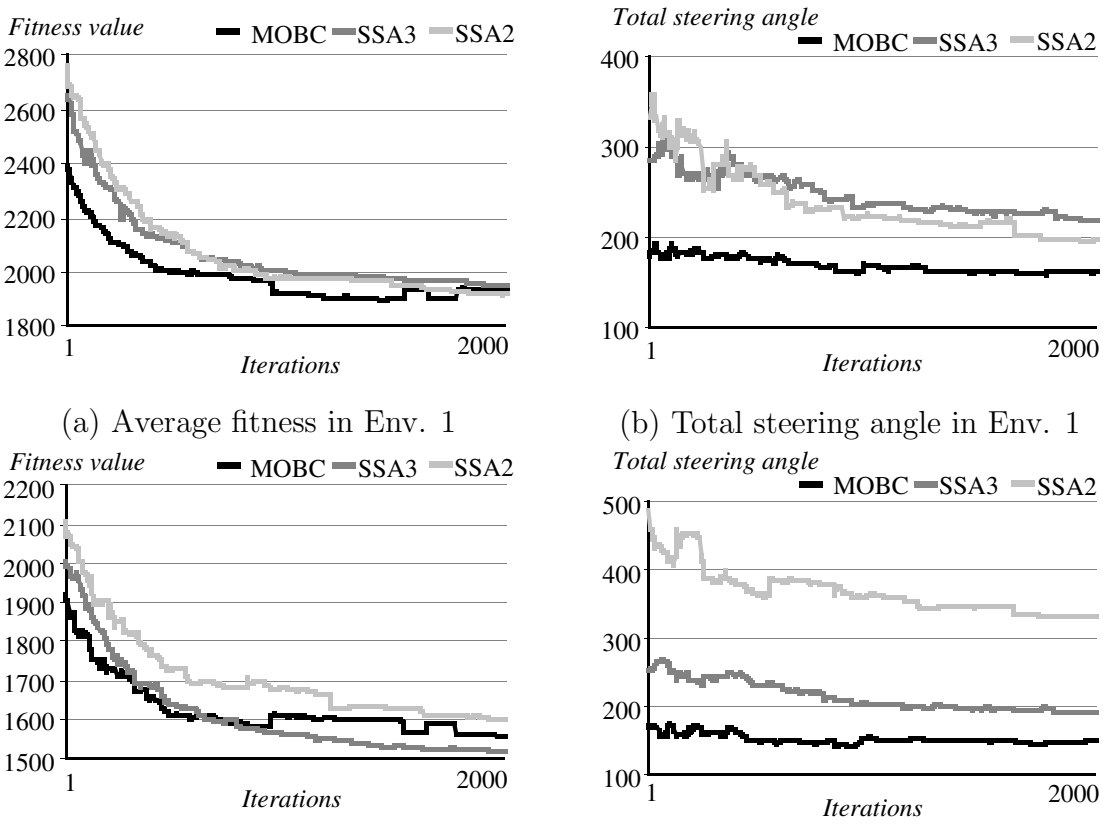


Fig 3.18 Comparison of average fitness values among MOBC, SSA3, and SSA2.



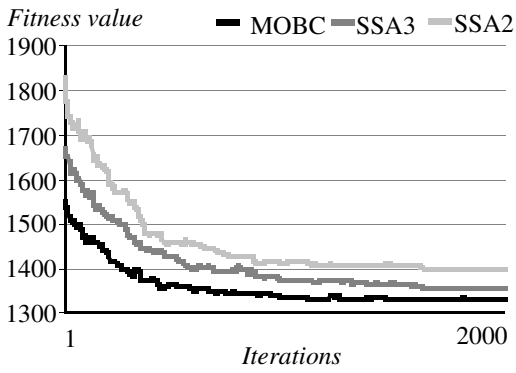
(a) Average fitness in Env. 1

(b) Total steering angle in Env. 1

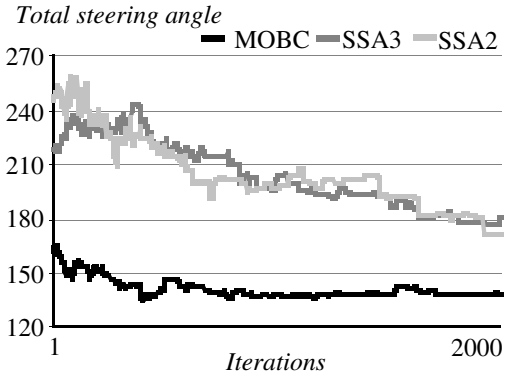
(c) Average fitness in Env. 2

(d) Total steering angle in Env. 2

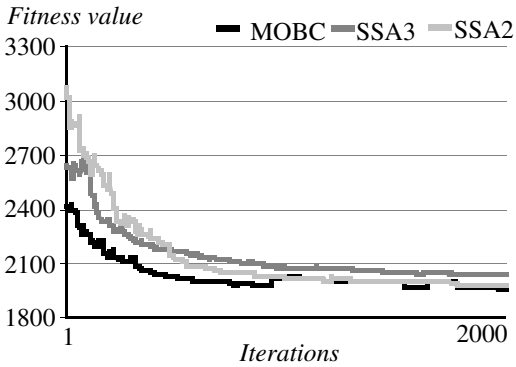
Fig 3.19 Histories of fitness values and total steering angle in each environment.



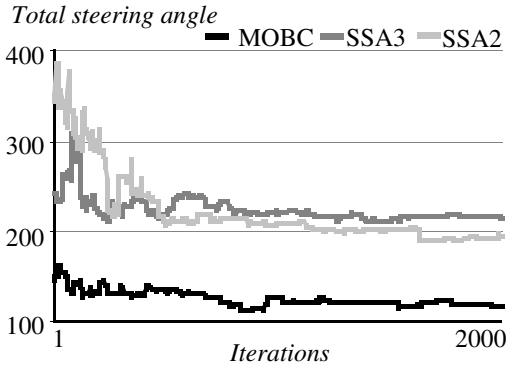
(a) Average fitness in Env. 3



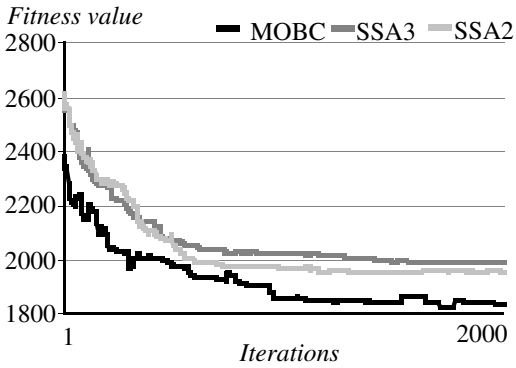
(b) Total steering angle in Env. 3



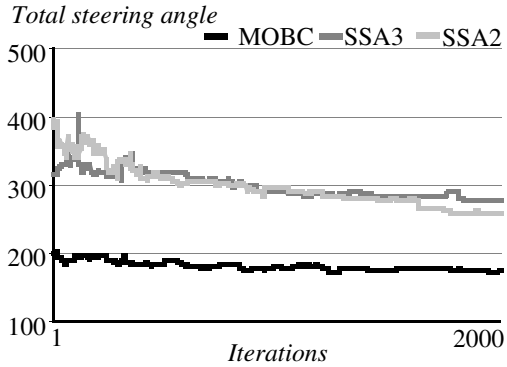
(c) Average fitness in Env. 4



(d) Total steering angle in Env. 4



(e) Average fitness in Env. 5



(f) Total steering angle in Env. 5

Fig 3.20 Histories of fitness values and total steering angle in each environment.

3.5.3 未学習環境における実行可能性

獲得した基本行動が，未学習環境においても対応できるかを検証する．行動獲得時に用いた環境とは異なる環境（Env. 6-10）を用いて，行動調停則毎にシミュレーションを行った．表 3.2 に各環境における評価値と総操舵角を，図 3.21〜3.23 に各環境における移動ロボットの軌跡を示す．表中のハイフンは，タスクが達成できず障害物に衝突したり，デットロックに陥ったことを表す．多目的行動調停は，獲得した基本行動を用いて，すべての環境 6-10 において，行動切り換え型より滑らかに移動できていることが表から確認できる．評価値に関しては，環境 6 と環境 8 を除いては多目的行動調停の方が良好な結果がでている．次に，図 3.21〜3.23 に示す軌道の中で特徴的なものを 2 つ比較する．

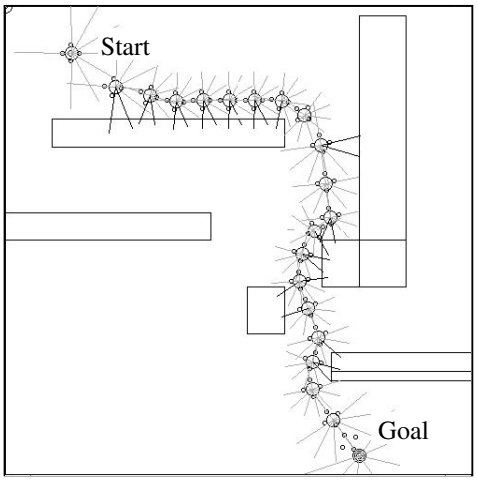
(1) 環境 7 における多目的行動調停と二行動切り替え型との比較：迷路の通路のような環境において，多目的行動調停を用いることで，ロボットが無駄なく滑らかに移動していることが図 3.21(b) より確認できる．行動切り替え型に関しては，三行動の場合タスクが達成されておらず，二行動の場合であれば壁面に沿って大きく回り込むような軌道を描いている．そこで，この環境 7 における多目的行動調停の行動重みの変化と，二行動切り替え型の行動切り替えの変化を比較する（図 3.24）．図 3.21(b) 中の A の辺りにおける行動重みを図 3.24(a) の A に示す．障害物回避しながら壁面を沿い，そして目標に向かうような多目的な動作を行っていることが解る．この多目的性が保たれた行動重みの配分が，A 付近の通路出口において，壁から離れると同時にゴール方向を向く切っ掛けになっていると考えられる．逆に，二行動切り替え型に関しては，通路中から障害物回避行動の行動重みが 1 になっており，壁面を併走し続けることで，図 3.23(b) のような無駄な軌道をとることが確認された．またこの結果より，二行動行動切り替え型において獲得された障害物回避行動は，壁面併走行動のような行動として獲得が行われ，センサが反応している時に壁に沿うように移動することが確認できる．つまり学習環境に特化した基本行動として獲得され，壁面併走が有効であろう環境 6 において，多目的行動調停よりも評価値が高くなっていると考えられる．

(2) 環境 9 における多目的行動調停と三行動切り替え型との比較：図 3.21(d) の B 付近の軌道が示す通り，多目的行動調停を用いることで，移動ロボットは狭路も通ることができる．この時の行動重みの変化を図 3.25(a) に示す．狭路にさしかかる前は，壁面併走行動を主行動とした多目的な動作を行い，その多目的さ故に狭路への移動が可能なっている．通路が狭いために，主行動が障害物回避行動に変わるが，結果として，狭路を抜けることによって，より良い評価値が得られている．逆に，三行動切り替え型においては，図 3.25(b) の C に示す通り，壁面から隙間にかけて，頻繁に主行動を切り替えていることがわかる．どの行動も狭路を進むにはいたらず，結果として回り道

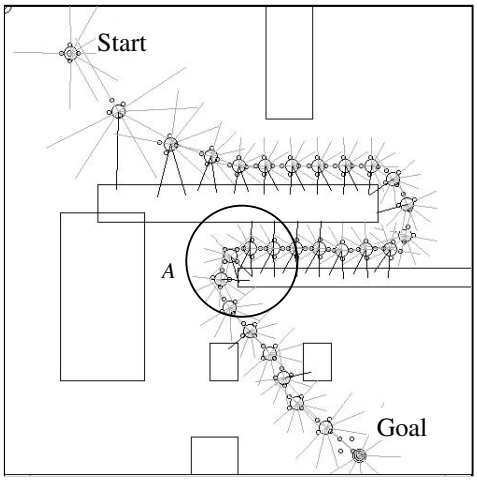
Table 3.2 Simulation results in unknown environments different from Env.1 - 5.

Env.	Criteria	MOBC	SSA3	SSA2
Env.6	Fitness	2441	-	2101
	Total steering angle	172.4	-	172.9
Env.7	Fitness	3276	-	3553
	Total steering angle	194.0	-	217.3
Env.8	Fitness	1746	2530	1669
	Total steering angle	166.9	298.2	182.3
Env.9	Fitness	1914	2496	2397
	Total steering angle	144.9	336.2	226.6
Env.10	Fitness	2006	2585	2467
	Total steering angle	150.3	336.9	215.3

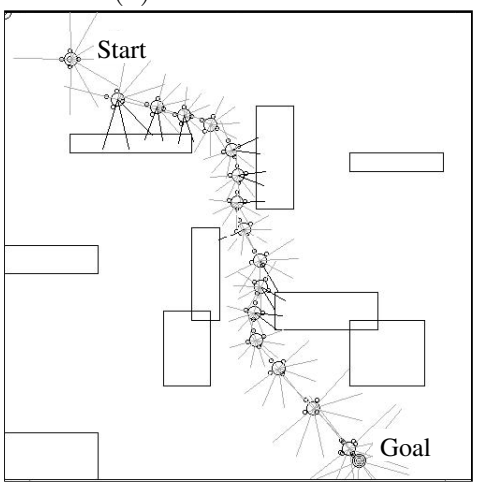
するような軌道が確認された．行動切り換え型は，選択された行動の出力のみが動作出力として直接用いられるため，各基本行動の適用範囲が限定され，学習環境に大きく依存して構造最適化が行われる．一方，多目的行動調停を用いると，動作出力はすべての基本行動の融合であるため，基本行動の特性は保持され構造最適化が行われると考えられ，その結果，未学習環境においても滑らかな動作，多目的な動作が可能となっている．



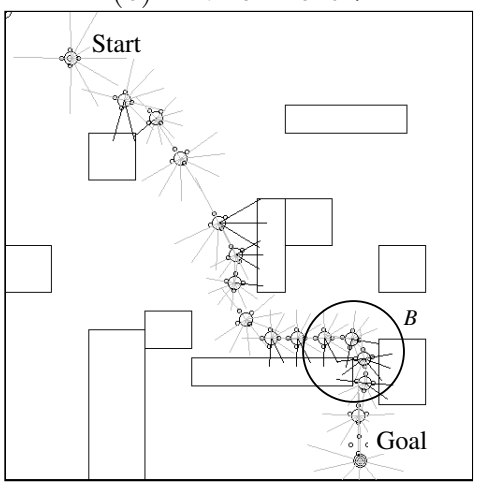
(a) Environment 6



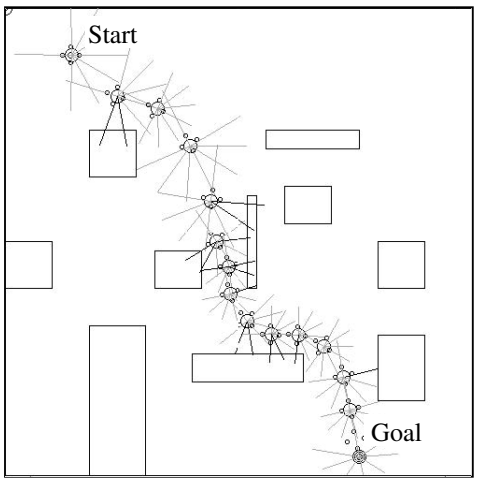
(b) Environment 7



(c) Environment 8

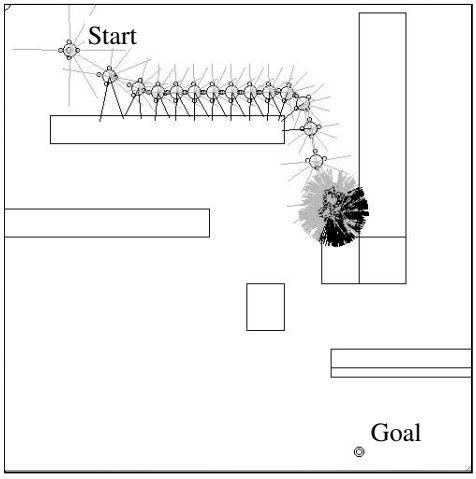


(d) Environment 9

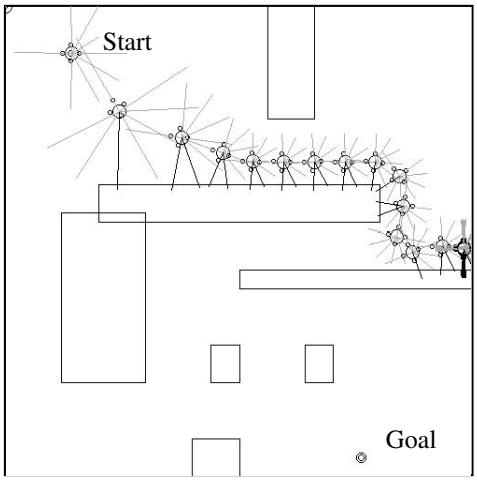


(e) Environment 10

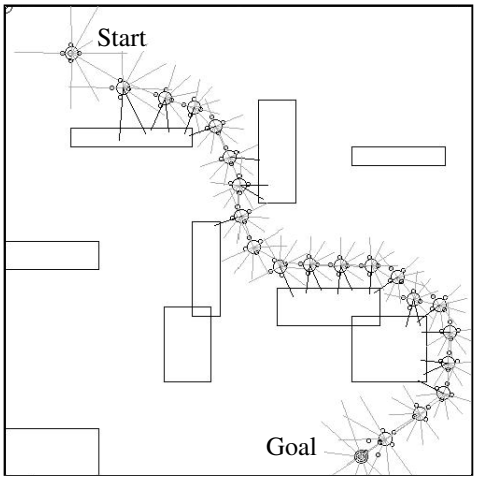
Fig 3.21 Trajectories of the robot with MOBC in unknown environments.



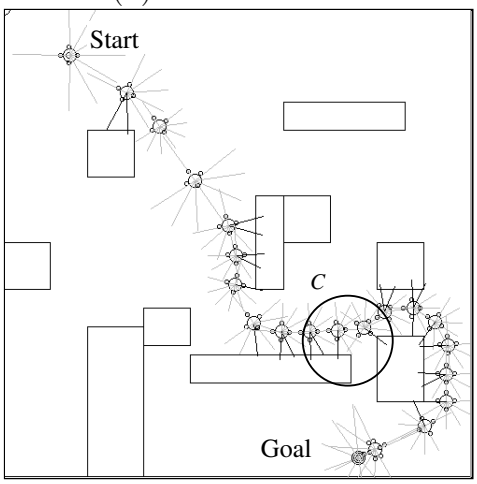
(a) Environment 6



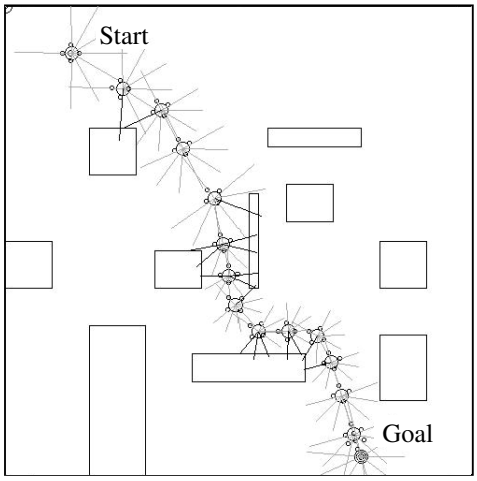
(b) Environment 7



(c) Environment 8

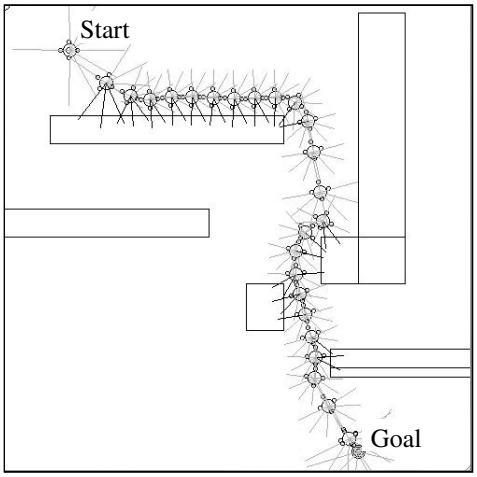


(d) Environment 9

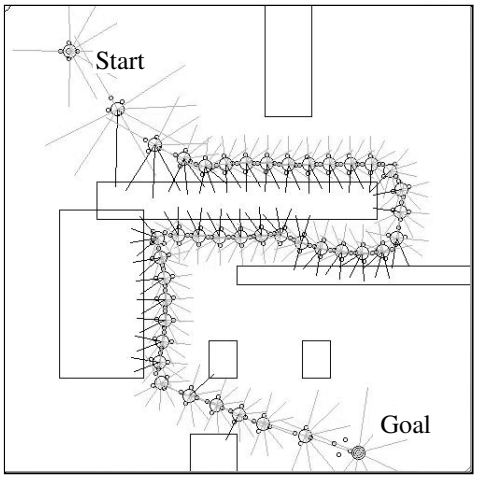


(e) Environment 10

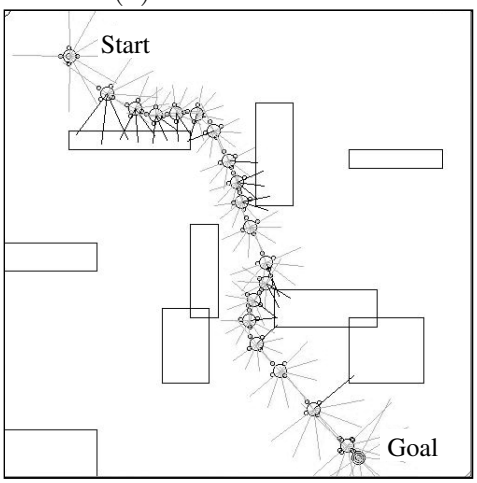
Fig 3.22 Trajectories of the robot with SSA3 in unknown environments.



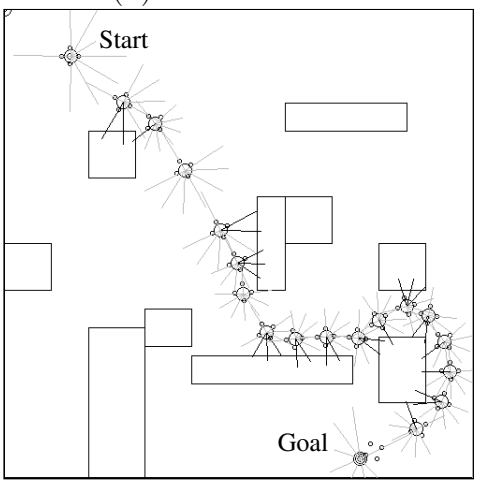
(a) Environment 6



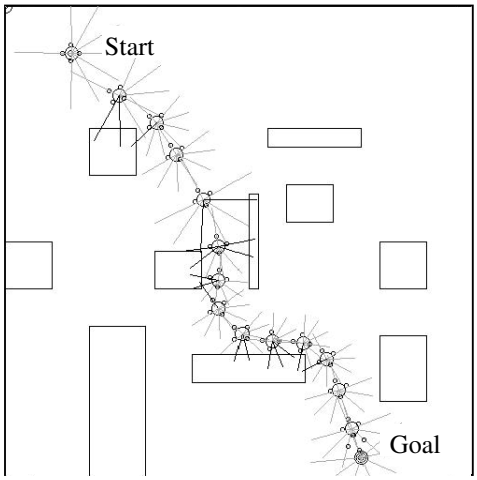
(b) Environment 7



(c) Environment 8

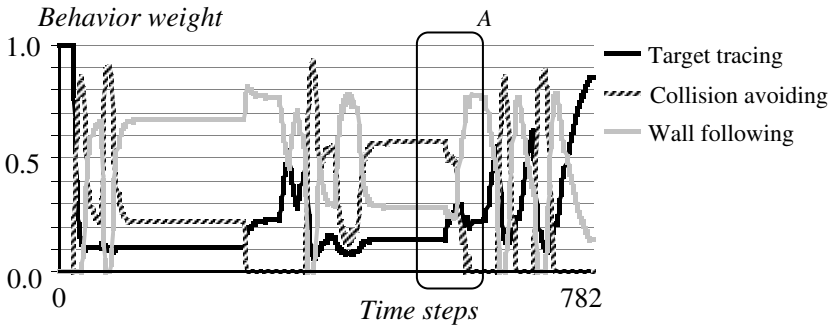


(d) Environment 9

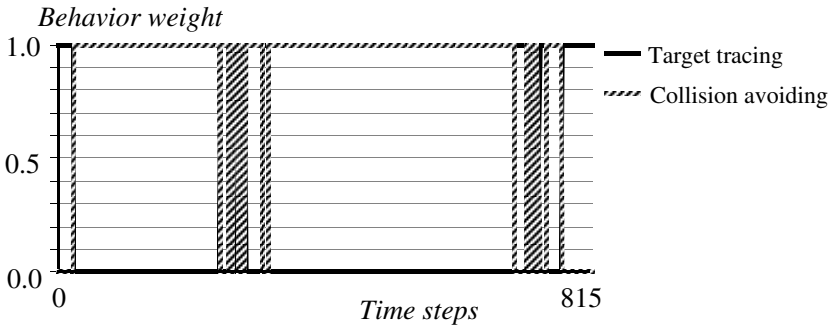


(e) Environment 10

Fig 3.23 Trajectories of the robot with SSA2 in unknown environments.

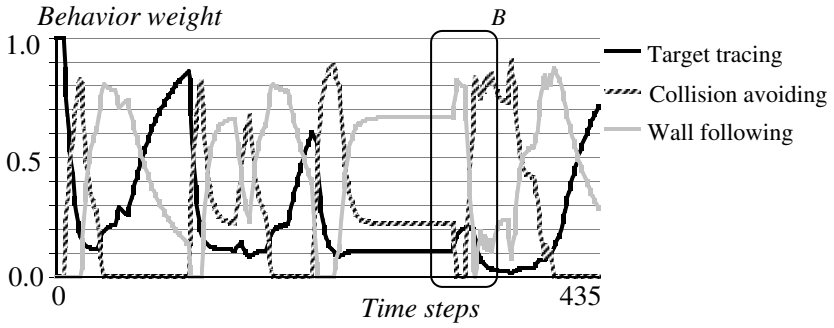


(a) MOBC

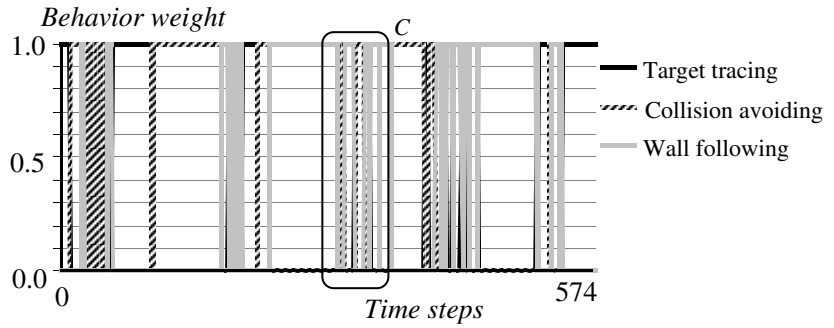


(b) SSA2

Fig 3.24 Change in behavior weights in Env. 7.



(a) MOBC



(b) SSA3

Fig 3.25 Change in behavior weights in Env. 9.

3.6 結言

本章では，静的未知環境におけるナビゲーション問題に対して，移動ロボットの基本行動の獲得を，センサ情報の時系列を考慮したトップダウン的な多目的行動調停則のもとで行った．行動調停則により役割が割り当てられた基本行動の学習を通して，障害物回避行動と壁面併走行動を獲得し，センサ情報に対して滑らかな主行動の遷移が可能となることを示した．また，一連の動作は，無駄な切り返しが少ない滑らかな動作として獲得されることを示した．

次に，複数の静的未知環境下で基本行動の学習を行い，時系列のセンサ情報を考慮しない反射的な行動切り替え型ロボットと比較することで，多目的行動調停を持つロボットがより短い世代で，最短距離及び最短時間に近い性能を達成することができることを示した．

さらに，未学習の未知環境下においても，多目的行動調停に基づくロボットが，学習環境と同様の動作を行うことができることと，反射的な行動切り替え型ロボットのタスク達成度が低いことを示した．

多目的行動調停を用いたロボットシステムの行動獲得は，学習環境に特化されるのではなく，観測情報に依存してトップダウン的に与えられる役割に対して特化される．またその役割の度合い（行動重み）は，時系列センサ情報により変化するため，単に行動を切り替えるよりも役割が多様である．そのため，未学習の静的未知環境に対しても，環境の障害物の配置や疎密さなどが大きく変わらないような環境であれば，問題なく対応できることが明らかになった．つまり，調停則の設計の段階でのヒューリスティクスから，環境が大きく異ならない限りにおいて，学習環境と同様の性能が得られ，基本行動の再利用性が高いことが明らかになった．

本章でとりあげた環境は未知であるが障害物が移動しない静的な環境を対象としている．実際，我々の生活している環境は，静的な要素も多く含まれるが，移動障害物となりうるものも同時に存在している．今後の課題として，そのような動的な障害物を含む環境下における適応学習を検討する必要があり，次章にて議論する．

第4章　動的環境下における移動ロボットの局所エピソードに基づく環境適応

4.1　緒言

前章でとりあげた未知環境は，固定障害物の配置が事前に知らない静的未知環境である．その静的環境下において，基本行動の出力を融合する多目的行動調停を提案し，学習速度や未学習環境への対応能力などから有効性を示した．しかし，我々の生活環境を見渡した時，家具や建物など静的な部分はたくさんあるが，他の人間や乗り物など様々な移動物体も同時に存在している．これら移動物体を含む観測情報は，多岐にわたり未知な遭遇パターンも多い．そこで，本章では，移動ロボットの適用範囲の拡張を目指して，移動障害物を含む動的環境下におけるロボットの環境適応に関して議論する．

移動障害物を含むナビゲーション問題に関する研究は，大きく 3 つのアプローチにわけることができる．

1. 大域情報を用いた意思決定.
2. 局所情報から特定の移動障害物の動き方を予測した動作生成.
3. 局所情報のみから観測情報の変化に対応した動作生成.

アプローチ 1 に関して，例えば，Fujimura らは，位置や速度が既知な移動障害物に対して，時間ごとに空間をセル分割し探索する手法を提案している [73]．また，人工ポテンシャル場 [39] を用いた手法として，杉山らは，ポテンシャル場の局所デットロックを回避する方法として，障害物の中心に渦の吹き出しを仮定した流体力学的ポテンシャルを提案し，移動障害物回避に適用している [75]．ただし，移動障害物とロボットの位置関係により渦の回転方向を事前に設定する必要がある．Brock らは，ポテンシャル法を基本とした Elastic Strips という弾性的な軌道情報を用い，移動障害物の経路への出入りに対し，実時間で再経路計画を行う手法を提案している [74]．これは，進行方

向に対する障害物回避能力と実行可能性は高いが，経路以外からの移動障害物に対応できない．また，井上ら [77] や藤澤ら [76] は，移動障害物の短期的な軌道予測と実時間探索を用いた手法を提案している．これらは，予測誤差が大きくなったときに，大きく回避するような動作を生成したり [77]，探索途中の実行可能解を使用し移動障害物による実時間性に対応する [76]．これら大域情報を用いたアプローチは，障害物の区別と正確な位置やベクトル情報など，対象の属性を用いるため鳥瞰的な観測機構が必要となる．さらに，これらの手法を実環境に適用しようとする場合，ロボットの意思決定に必要な情報と不必要な情報を，障害物との距離や属性からすべて抽出しなければならないという問題がある．

アプローチ 2 に関して，新井らは，赤外線通信を用いて，近くにいるロボットと通信を行い，向きや進行方向を共有し，自ら計測した距離情報とで，状態空間を構成し強化学習により回避行動を学習させる手法を提案している [78]．Fiorini らは，衝突が生じるであろう Velocity Obstacle というベクトル情報から，回避可能な経路を予測する手法を提案している [79]．これらの手法は，近傍情報から意思決定を行うが，大域情報のアプローチと同様に，対象とする移動障害物や他のロボットの対応付けのため，個々の障害物に ID などを常に設ける必要がある．

アプローチ 3 に関しては，例えば Brooks の SSA [40] に基づいて，移動障害物に対応できるように基本行動を設計することで，事前に想定できうる移動障害物にのみ対応可能となる．ただし，移動障害物を含む環境の特性が事前に知っている必要がある．それに対し，環境の特徴を学習する手法として，Chang らは，距離センサ情報から次に観測されるセンサ情報を NN を用いて予測し，移動障害物に対応する手法を提案している [80]．さらに，北川らは，静的環境下で距離センサ情報の予測モジュールを学習し，動的環境下で予測誤差が大きく生じる時，移動障害物に遭遇したと仮定し，多目的な動作から緊急回避動作に切り替える手法を提案している [113]．これらの手法は，移動障害物に対処するために，ロボット自体が単目的化（移動障害物回避のみの動作）してしまうという問題がある．そのため，タスクが複雑になれば，タスクを達成できない可能性が高くなる．つまり，タスク達成に必要な多目的性を保持しながら，移動障害物を回避する動作を学習する必要がある．

アプローチ 1 や 2 に比べ，アプローチ 3 は，事前設備や設定条件が少なく未知環境に導入しやすい．ただし，行動を移動障害物に適応しようとした場合多くの学習時間が必要になるという問題がある．そこで，本研究では，動的環境下において，多目的性を保持しつつ移動障害物を回避できる動作の迅速な学習手法を検討する．

本章で扱う動的環境を図 4.1 に示す．移動障害物や他のロボットが存在する環境内をスタートからゴールまでなるべく短い移動時間および移動距離で向かうことをタスク

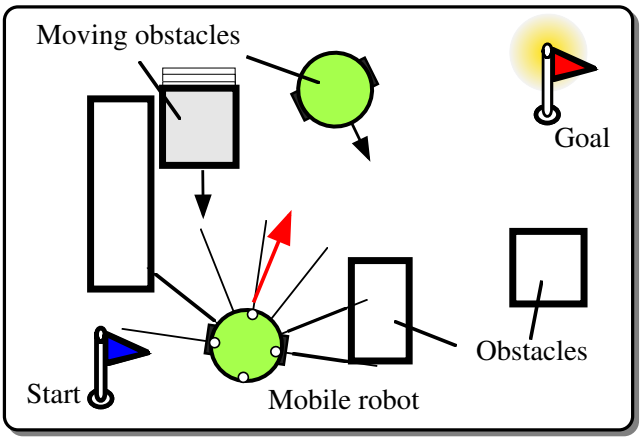


Fig 4.1 Dynamic environment including static and moving obstacles.

として与える．このタスクにおいて，スタートからゴールまでの試行をシナリオとする．ロボットは，近傍の障害物と距離を計測できるセンサと，ゴール方向が計測できる光センサを装備しているものとする．

基本的に，タスク達成の善し悪しは，オフライン的に一連のシナリオを評価することで決まる．しかし，移動障害物の移動方向や遭遇時間，遭遇場所などが既知である場合，シナリオを評価することは可能であるが，未知な場合，遭遇の仕方が多岐にわたることを考慮すると，シナリオを一意に評価することが難しい．そこで，静的環境下で学習した基本行動や行動調停則を用いて，移動障害物との遭遇に対して適応できるよう部分的に行動調停則を改善する手法を提案する．

短期的に過去の観測情報と動作出力の情報を保持する局所エピソード（local episode）を導入し，移動障害物との衝突に対して，過去にとった動作を局所エピソードから参照し，制御器を学習する手法を提案する．この学習手法を多目的行動調停則の適応学習に適用し，静的環境下において獲得された行動調停則を，動的環境で直面する状況に合わせて部分的に改善することを試みる．前段階として，行動調停則を局所的に調節しやすくするために，行動調停則をファジィルールで置き換え，静的環境下で獲得する．次に，動的環境下での多目的行動調停則の適応学習について検討し，局所エピソードを用いた学習手法を提案する．これら提案手法を計算機シミュレーションおよび，実ロボットにより実験を行い有効性を検証する．

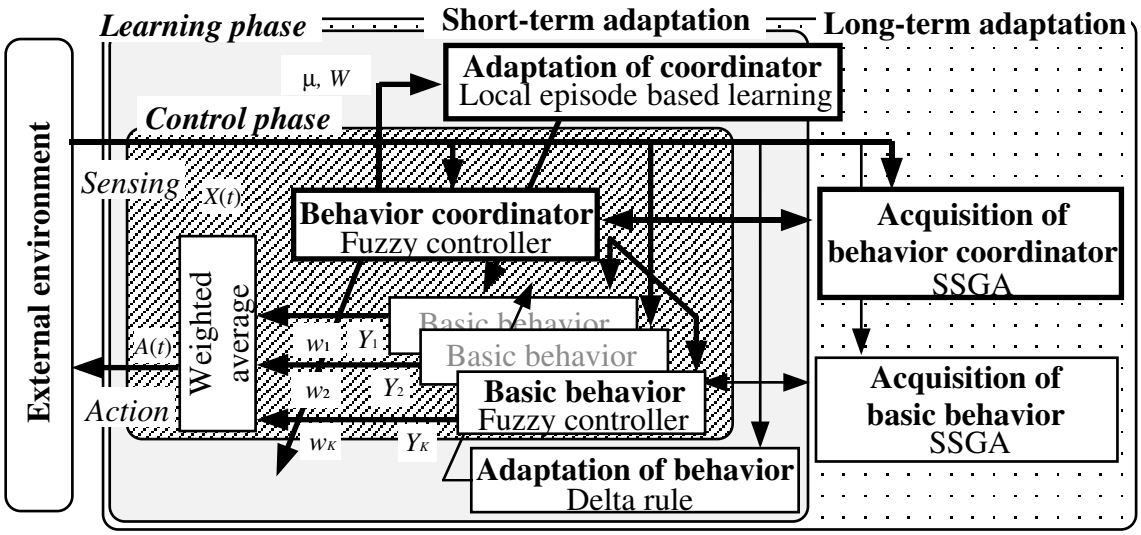


Fig 4.2 Total architecture of the multi-objective behavior coordination with local episode-based learning.

4.2 多目的行動調停に基づく移動ロボットの構造化学習

図 4.2 に，MOBC を用いた移動ロボットの全体概念図を示す．まず図中斜線部の制御フェーズ（control phase）に関して説明する．外部環境の状態 $X(t)(x_i(t); i = 1, 2, ..., n)$（障害物との距離や，目標方向）をセンシングし，基本行動毎に行動出力 $Y_k(t)$=($y_1^k(t)$, $y_2^k(t)$) $(k = 1, 2, ..., K)$ を計算する．行動出力は，y_1 を速度，y_2 を操舵角とする．そして，各基本行動に対して行動重み $W(t)$ =($w_1(t)$, $w_2(t)$, ..., $w_K(t)$) を割り当て，重み付け平均により動作出力 $A(t) = (a_1(t), a_2(t))$ を生成する．

$$A(t) = \frac{\sum_{k=1}^{K} w_k(t) Y_k(t)}{\sum_{k=1}^{K} w_k(t)} \tag{4.1}$$

K は基本行動の総数とする．基本行動は，前章で説明した通り，目的に応じてファジィルールやニューラルネットワーク，プロダクションルール等を用い構成することが可能である．

3 章で用いたプロダクションルールによる行動調停則は，非常に簡略化され，静的環境下では適用も容易である．しかし，移動障害物に対応させようとして設計した場合，タスク達成を軽視した，回避重視の行動重みの配分になりかねない．そこで，局所的に行動調停則の更新を行うための前段階として，状態空間をファジィメンバシップ関数で分割し，より詳細な状態分割を行う．そして，不要な状態分割を避けるために，進化的計算により静的環境下で獲得を行う．ファジィルールとして行動調停則を表現す

ることで，これまで用いてきたプロダクションルールの明示性を失うことなく，より詳細な行動重みの更新が行えると考えられる．

多目的行動調停則は，後件部が実数値で表される簡略型ファジィ推論（simplified fuzzy inference）を適用する [7]．入力情報 $I(t)$ は，現在の外界センサ情報 $X(t)$ と 1 ステップ前のセンサ情報 $X(t-1)$ との差 $\Delta X(t)$ とで定義する．j 番目のファジィルールは，入力 I_i とその入力に対するメンバシップ関数 $F_{i,j}$ と，後件部実数値 $q_{k,j}$ により次のように構成する．

$$IF\ I_1\ is\ F_{1,j}\ and, ..., and\ I_i\ is\ F_{i,j}\ and, ..., and\ I_{2n}\ is\ F_{2n,j}$$
$$THEN\ w_1\ is\ q_{1,j}\ and\ w_2\ is\ q_{2,j}\ and, ..., and\ w_K\ is\ q_{K,j}$$

j 番目の調停則ファジィルールの適合度 μ_j は次式で計算される．

$$\mu_j = \prod_{i=1}^{2n} \begin{cases} 1 - \dfrac{|I_i - \alpha_{i,j}|}{\beta_{i,j}} & |I_i - \alpha_{i,j}| \leq \beta_{i,j} \\ 0 & otherwise \end{cases} \tag{4.2}$$

$\alpha_{i,j}$ と $\beta_{i,j}$は, i 番目の入力に対するルール j のファジィメンバシップ関数の中心値とその幅とする（図 4.3)．また，k 番目の行動に対するルール j の後件部出力値を $q_{k,j}$ とすると，各基本行動の行動重みは次式により計算され，次ステップの行動重み $W(i+1)$ に用いられる．

$$Q_k = \frac{\sum_j^m \mu_j q_{k,j}}{\sum_j^m \mu_j} \tag{4.3}$$

$$w_k\,(t+1) = \frac{Q_k}{\sum_k^K Q_k} \tag{4.4}$$

m は調停則ファジィルールの総数を表す．

次に，図 4.2 の学習フェーズ（learning phase）について簡単に説明する．未知環境では事前に正確な教師値を与えることができないため，シナリオ終了時に得られる評価項目，例えば移動時間や移動距離から，学習対象の基本行動を評価する．学習には進化的計算手法の一つである定常状態遺伝的アルゴリズム（steady-state genetic algorithm; SSGA）用いる．これらの学習を通して，MOBC の動作の多目的性や，新たな未知環境に対する汎化性を議論してきた．同様に，行動調停則の学習として，行動調停則のファジィルールの前件部の組み合わせ最適化と後件部のチューニングを SSGA により行う．

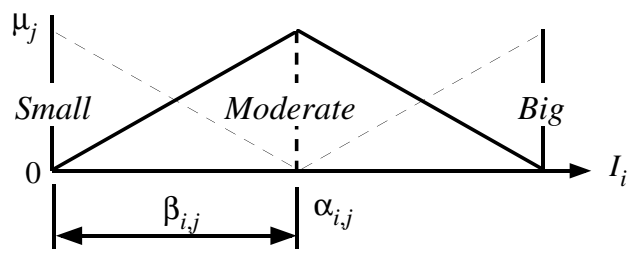


Fig 4.3 Fuzzy membership functions for the behavior coordination.

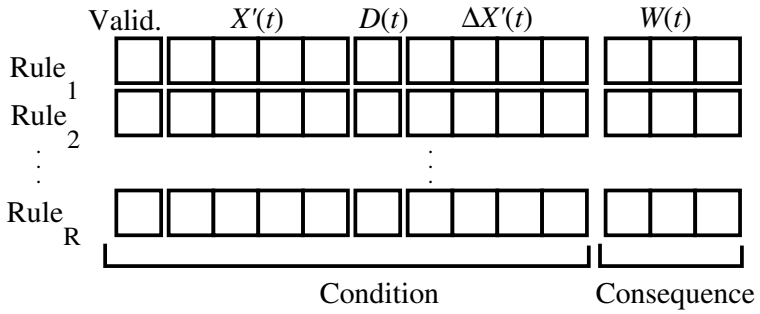


Fig 4.4 Coding of behavior coordination rules as an individual for SSGA.

4.3 静的環境下での多目的行動調停則の獲得

静的環境下における多目的行動調停則の獲得は，3 章で説明した基本行動の獲得と同様の枠組みで行う．まず，行動調停則を解候補として，コーディングする（図 4.4）．個々のルールは，ルールを使用するかしないかの Validity と，前件部のメンバシップ関数の組み合わせ，そして後件部の行動重み出力値でコーディングされる．メンバシップ関数の組み合わせは，図 4.3 に示される 3 つのメンバシップ関数と，どんな入力に対しても適合度 1 を返す *don't care* で表現される．また，後件部出力値は，実数で表す．

SSGA の手順として，(1) 個体群の初期化：前件部の組み合わせと後件部出力値を乱数で初期化する．(2) 初期評価：各個体を用いてタスクを実行し，評価値を計算する．(3) 最小適応度除去：もっとも評価値の悪い個体を一つ淘汰する．(4) 交叉：エリート交叉を用いて，新しく個体を一つ生成する．(5) 突然変異：新しく生成した個体に対して，遺伝子座単位で，メンバシップ関数の入れ替えや，摂動を加える．(6) 評価：新しく生成した個体を用いてタスクを実行し，シナリオを評価し，評価値を計算する．各評価項目は，基本行動の獲得と同様に，ゴールまでの時間ステップ P_{time} と移動した距離 P_{length} に基づき，次式より行う．

$$fitness = \omega_1 P_{time} + \omega_2 P_{length} \tag{4.5}$$

$\omega_h\,(h=1,2)$ は各評価項目に関する重みとする．終了条件がみたされるまで，(3) から (6) を繰り替えす．

4.4 多目的行動調停則の局所エピソード学習

これまで静的未知な環境下において，基本行動や行動調停則が，大域的な学習手法である SSGA やデルタルールにより獲得できることを示してきた．しかし，明示的に与えられる評価関数は，局所的な状況を考慮しておらず，また静的環境において理想的なデルタルールの教師値は，動的環境で適さないことが多い．それゆえ，静的環境で学習したロボットは，動的環境下において移動障害物と容易に衝突しうる．本研究では，動的環境において，移動障害物との衝突の原因が，近い過去にとった動作にあるものと仮定する．近い過去の動作は，一定ステップ数の観測情報と動作出力情報で構成され，局所エピソードとして短期記憶に保存される．衝突時にこの局所エピソードを参照し，過去にとった動作に関連する意思決定ルールを調整することで，ロボットは移動障害物に対応できる．

多目的行動調停において，行動調停による動作は，行動重みの変化に依存する．この依存関係から，衝突前の近い過去における行動重みの更新遅れに，移動障害物との衝突の原因があると仮定する．この仮定を基に，行動重みの更新遅れに関連する行動調停則のファジィルールを局所的に更新し，同じような状況下での移動障害物との衝突を回避する．現在の時間 t から過去に遡る一定ステップ数 τ_{end} までのファジィルールの適合度 $\mu(\tau)$ と行動重み $W(\tau)(\tau=1,2,...,\tau_{end})$ から，局所エピソードを構成する．この局所エピソードが，スタートからゴールまで逐次更新され，ロボットは常に一定の過去の情報を保持する．そして，衝突時の行動重みの更新遅れを補うために，衝突時の行動重み W^* を局所エピソードにフィードバックする．この学習は，行動調停則のファジィルールの後件部出力値を調整する次式により行う．

$$q_{k,j} \leftarrow q_{k,j} + \gamma^{\tau}\eta\left(w_k^* - w_k\left(\tau\right)\right)\frac{\mu_j\left(\tau\right)}{\sum_j \mu_j\left(\tau\right)}$$
$$(t=1,2,...,\tau_{end};\ 0<\gamma\leq 1) \tag{4.6}$$

η と γ を，それぞれ学習係数と減衰率とする．衝突時の行動重みは，移動障害物に非常に接近した状態での緊急回避的な行動重みになっている．これを衝突前に遡って教師値とすることで，衝突時と同じような遭遇パターンにおいて，より早い段階で移動障害物に対する回避が可能になると考えられる．また，衝突時の行動重みを教師値とすることによって，外部から教師値を与える必要がなく，学習の指向性が得られ，学習回数が減少すると考えられ，移動障害物に迅速に対応できることが期待できる．

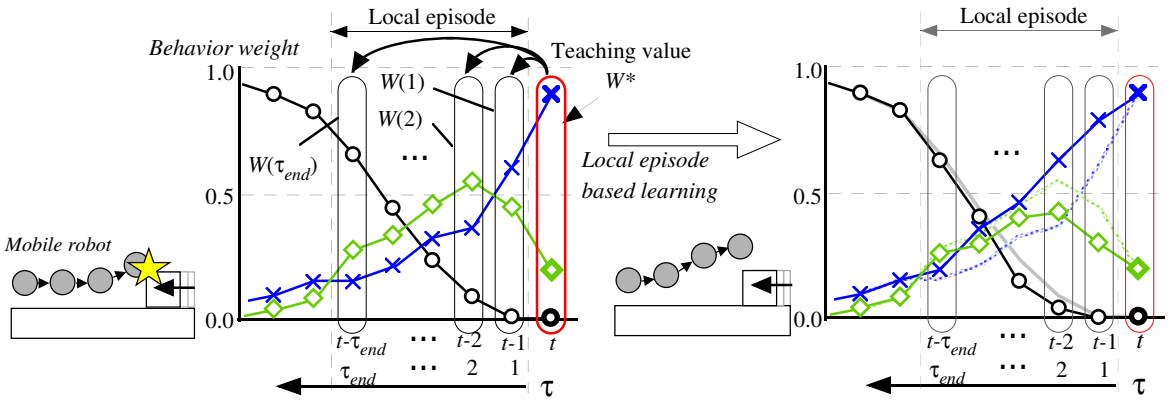


Fig 4.5 Concept of local episode-based learning for mobile robots.

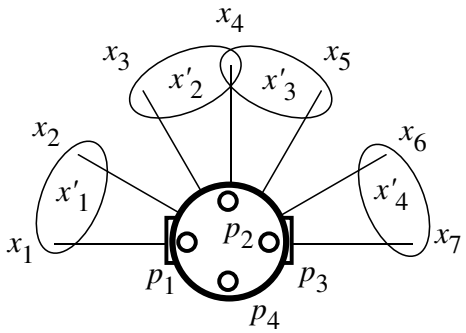


Fig 4.6 Sensor arrangement of the mobile robot.

4.5 計算機シミュレーション

移動ロボットは，図 4.6 に示す 7 つの距離センサ $X(x_1, x_2, ..., x_7)$ と，4 つの光センサ $P(p_1, p_2, p_3, p_4)$ を持つものとする．ロボットに与えるタスクは，障害物が配置された環境内を，なるべく短い移動時間と移動距離でゴールに向かうこととする．このタスクを達成するために，3 つの基本行動を用意する．一つは，光源をゴールと仮定した，光センサからの情報を用いた目標追従行動であり，プロダクションルールで設計され，速度と操舵角を出力する．他の二つは，距離センサによる障害物回避行動と壁面併走行動であり，距離センサ情報を入力とするファジィコントローラにより設計され，速度と操舵角を出力する．実数で表されるファジィルールの後件部は，それぞれの基本行動の特徴に合わせてヒューリスティックに設定する．基本行動の特徴として，障害物回避行動は，障害物を回避するために，操舵角が大きく，速度が遅くなるように後件部を設定する．壁面併走行動は，側面の x_1 や x_2，x_6，x_7 のセンサのみ障害物を感知するとき，障害物回避よりも速く壁面と平行に移動（併走）できるように設定す

る．多目的行動調停則は，簡単化のため，縮約した距離情報 $X'(t)$ とその変化 $\Delta X'(t)$ $(= X'(t) - X'(t-1))$ 及び，正規化した光センサ情報 p' を入力情報 $I(t)$ とする．縮約は次式により行う．p_{max} は，光センサ情報の最大値とする．

$$x'_1 = \frac{1}{2}(x_1 + x_2) \tag{4.7}$$

$$x'_2 = \frac{1}{2}(x_3 + x_4) \tag{4.8}$$

$$x'_3 = \frac{1}{2}(x_4 + x_5) \tag{4.9}$$

$$x'_4 = \frac{1}{2}(x_6 + x_7) \tag{4.10}$$

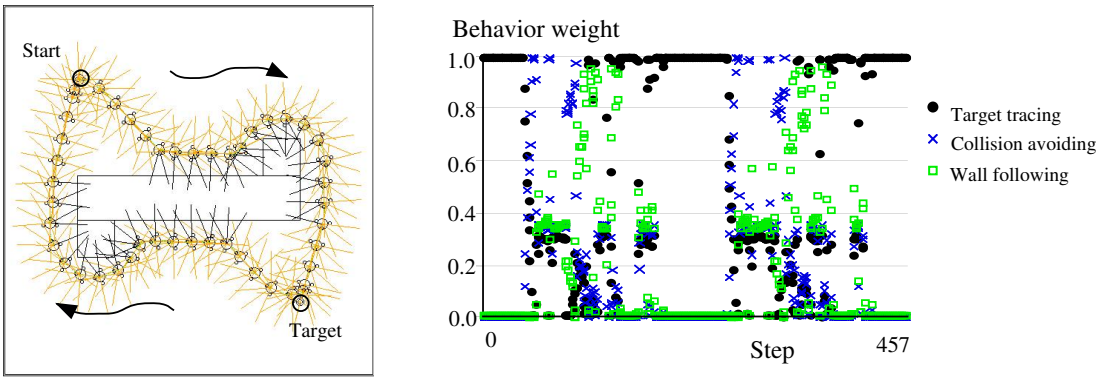
$$p' = \frac{p_1 - p_3 + p_{\max}}{2p_{\max}} \tag{4.11}$$

4.5.1 静的環境下における行動調停則の獲得

まず，複数の静的環境下において，行動調停則の獲得を行う．固定障害物の配置が異なる 3 つの環境を用意した．個体の評価は，各環境におけるシナリオの評価値の総和として計算する．個体は，50 個のファジィルールをそれぞれ保持するものとする．個体数 50 個，10000 回評価（200 世代）にて行動調停則を学習させた．実験結果を図 4.7 から図 4.9 に示す．それぞれ，最終世代の軌道と，そのときの行動重みの変化を表す．

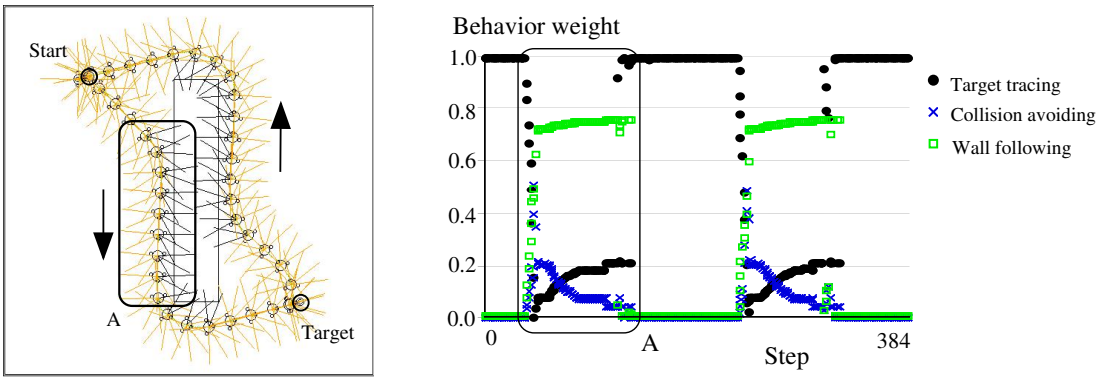
ロボットが学習した行動調停則の特徴を結果から検証する．壁面が続く図 4.8 の A のあたりにおいて，壁面併走行動を主に使用するような行動重みの更新が行われていることが確認できた．また，図 4.9 の B のあたりにおいて，障害物が密集した場所から少し開いた空間への移動に対して，徐々に壁面併走行動の重みをあげて，目標追従に切り替えるような行動重みの変化が確認できた．前者のような行動重みの変化は，ヒューリスティックルールでも設計できていたが，後者は，設計者が構築したヒューリスティックとは異なる行動重みの更新であり，スタートからゴールまでのシナリオの評価により，状況の一連の流れを考慮した行動調停則が獲得されていることが解る．静的環境下で獲得された行動調停則は，42 個のファジィルールで構成され，ロボットは静的環境下で基本行動を調停し多目的な動作を実現しタスクを達成した．

シミュレーション実験に用いた行動調停則の一覧を表 4.1 に示す．ただし，複数の静的環境で獲得されたルール 42 個の内，あまり使用されていないルールは除外してある．表中の S，M，B は，メンバシップ関数の形状を意味し，それぞれ *small*，*moderate*，*big* の言語ラベルに対応する．また $\sharp$ は，*don't care* を意味する．なお，今回用いた行動調停則は一例であり，異なる静的環境下では，異なる調停則が獲得されうる．



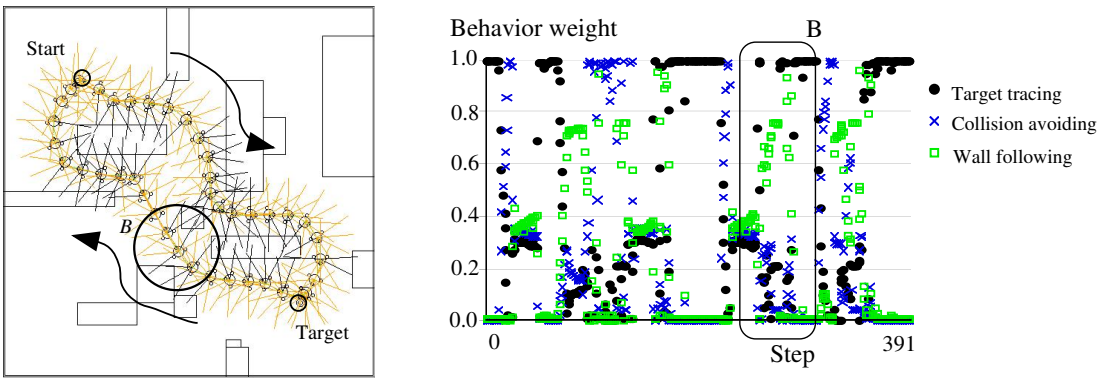
(a) Trajectory (b) Change in behavior weights

Fig 4.7 Trajectory and change in behavior weights in static environment 1.



(a) Trajectory (b) Change in behavior weights

Fig 4.8 Trajectory and change in behavior weights in static environment 2.



(a) Trajectory (b) Change in behavior weights

Fig 4.9 Trajectory and change in behavior weights in static environment 3.

Table 4.1 Coordination rules acquired in static environments. "S", "M", and "B" mean linguistic label *small*, *moderate*, and *big*, respectively. "♯" means wild card called *don't care*.

Rule	x'_1	x'_2	x'_3	x'_4	p'	$\Delta x'_1$	$\Delta x'_2$	$\Delta x'_3$	$\Delta x'_4$	q_1	q_2	q_3
1	♯	♯	S	♯	♯	♯	S	♯	♯	0.16	0.85	0.59
2	♯	♯	♯	♯	♯	♯	♯	S	M	0.36	0.64	0.18
4	♯	M	B	♯	♯	♯	♯	M	♯	0.76	1.00	0.94
7	♯	♯	♯	S	♯	♯	♯	M	S	0.80	0.98	0.55
9	S	M	♯	♯	B	S	♯	S	♯	1.00	0.02	0.09
11	♯	♯	♯	♯	♯	♯	♯	S	♯	0.48	0.90	0.54
12	♯	♯	♯	S	♯	♯	♯	♯	S	0.96	0.05	0.00
17	♯	♯	♯	♯	♯	♯	♯	S	♯	0.14	0.99	0.49
18	B	♯	♯	♯	♯	♯	S	♯	♯	0.43	0.50	0.65
20	S	♯	♯	S	♯	♯	♯	♯	♯	0.44	0.00	0.00
22	♯	♯	B	♯	♯	♯	♯	♯	♯	0.73	1.00	0.98
24	M	♯	♯	S	♯	♯	♯	♯	S	0.96	0.40	0.55
26	♯	♯	♯	♯	♯	♯	♯	M	S	0.00	0.04	0.91
29	♯	♯	M	♯	♯	♯	♯	♯	♯	0.01	0.99	0.00
30	M	♯	♯	♯	♯	S	S	♯	♯	0.73	0.68	0.32
31	♯	♯	♯	♯	♯	♯	M	S	♯	0.88	1.00	0.83
32	♯	M	♯	♯	♯	♯	♯	M	♯	0.00	0.13	0.31
35	♯	♯	B	♯	♯	♯	S	♯	♯	0.68	0.00	1.00
37	B	S	♯	♯	♯	♯	♯	♯	♯	0.27	0.05	1.00
38	♯	♯	♯	♯	♯	♯	S	♯	♯	0.18	0.18	0.59
40	B	♯	♯	♯	B	♯	M	S	♯	0.70	0.08	0.09
41	♯	♯	S	♯	M	M	♯	♯	♯	1.00	0.00	0.78

4.5.2 動的環境下における局所エピソード学習

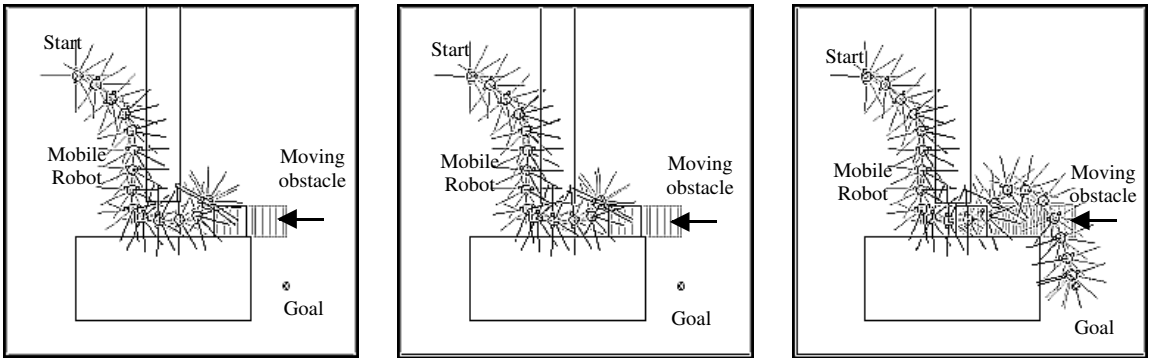
対象とする動的環境は，一定の速度で移動する障害物が存在する環境とする．移動障害物の進行方向が異なる 3 つの Case を比較することで，調停則の更新がどのように行われるかを検証する．なお，局所エピソードの長さは 5step に設定した．

Case1 として，移動障害物が壁に沿ってロボットに向かってくる場合を想定する．図 4.10 にロボットの移動軌跡，図 4.11 に行動重みの変化を示す．試行 1 において，ロボットが移動障害物を避けきれず，ロボットの側面に障害物が衝突しているのが確認できる．このときの最終的な行動重み（図 4.11(a)A の右端）を見ると，壁面併走行動が優先されていることが確認できる．よって図 4.10(a) のような状況であれば，壁面併走行動が優位に機能すると仮定し，その行動重みを衝突以前の過去の状態（局所エピソード）にフィードバックする．図 4.10 と 4.11 に示す通り，わずか 2 回の局所エピソード学習で移動障害物を回避することができた（図 4.10(c)）．行動重みの履歴から，ロボットが x_1 や x_2 あるいは x_6 や x_7 のセンサで障害物を感知した場合，壁面併走行動を優先するように行動調停則が改善されている．これは，壁面併走行動が障害物回避行動よりも移動速度が速いという特徴からも適していると考えられる．ここで，局所エピソード学習で更新された特徴的な行動調停則のファジィルールを図 4.12 に示す．図中左側は条件部を表し，縮約距離情報 x'_3 が増減せず，x'_4 が減少する場合，つまり，障害物が右側面に近付いてくる状態に最も発火することを意味する．図中右側は，後件部出力値の試行毎の変化を表す．局所エピソード学習により壁面併走行動に対する出力値が増大し，他の出力値が減少していることが確認できる．図 4.11 及び図 4.12 から，静的環境下で獲得された多目的行動調停則の多目的性を保持しつつ，移動障害物に対応するために局所的に，調停則が改善されていることが確認できる．

同様に，移動ロボットに対して，移動障害物が右側壁面から飛び出してくる Case2 において，図 4.13 に示すように，ロボットは移動障害物の前を通過する直前に側面から衝突された．図 4.14(a) の A 付近において，衝突時に壁面併走行動の行動重みが大きいことから，局所エピソード学習により，それ以前の状態に対して壁面併走行動を取りやすくすることが学習され，図 4.14(b) の B 付近に示すように，移動障害物に対して，なるべく壁面併走行動を維持するような調停則として改善されていることがわかる．これは，図 4.15 に示す改善された特徴的な調停則ファジィルールの後件部出力値の変化からも確認できる．

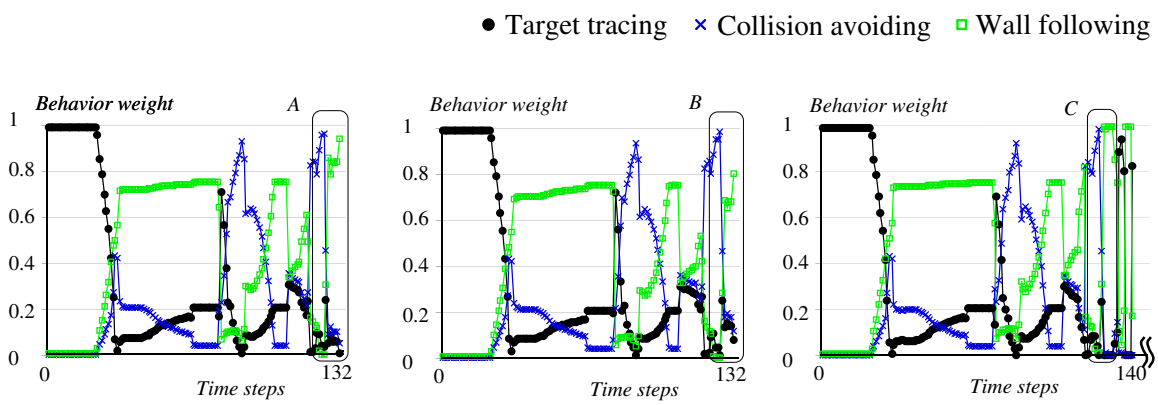
3 つ目の Case として，移動ロボットに対して，移動障害物が左側から，右の壁側に移動する場合の結果を図 4.16，4.17，4.18 に示す．この Case では，これまでとは異なり，右側正面で障害物と衝突していることが図 4.16(a) より確認できる．ここで衝突時の行動重みは，障害物回避行動が最も大きく（図 4.17(a)A），局所エピソード学習に

より，移動障害物に対して，大きく回避するように障害物回避行動の行動重みが増加するようになった（図 4.17(b)B）．これは，改善された特徴的な調停則ファジィルールからも確認することができる（図 4.18）．前件部は，縮約距離情報 x'_2 のセンサが，そのレンジの中間に障害物を感知し，x'_3 が変化しない状態を表す．このルールは，正面に感知した障害物の距離がかわらないときに発火する．局所エピソード学習における，障害物回避行動に対する後件部出力値の増加は，速度を落としてでも大きな操舵角を出す方が良いであろう図 4.16(a) のような状況に適していると考えられる．また，この Case においても，多目的な動作が保持されつつ，局所的に行動調停則が改善されていることが確認できる．



(a) 1st trial (b) 2nd trial (c) 3rd trial

Fig 4.10 Trajectories at each trial in case 1.



(a) 1st trial (b) 2nd trial (c) 3rd trial

Fig 4.11 Changes in behavior weights at each trial in case 1.

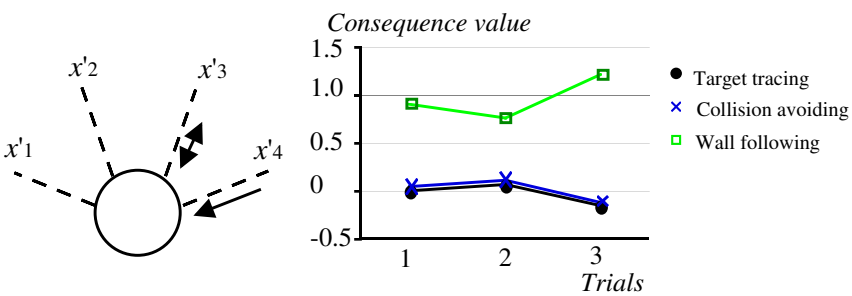
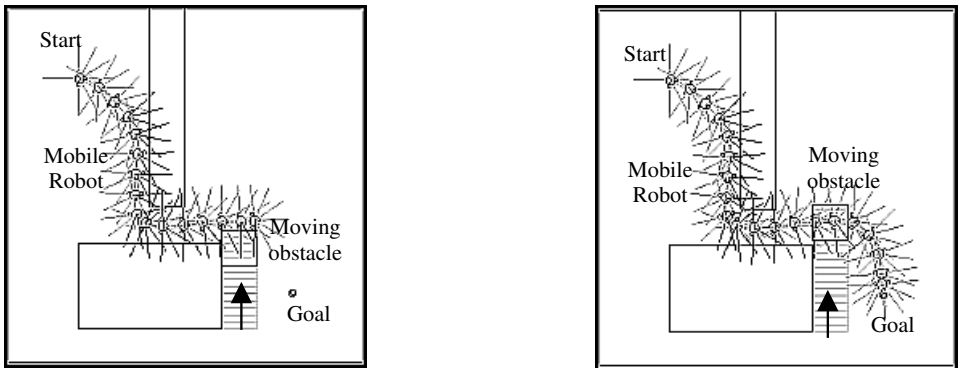
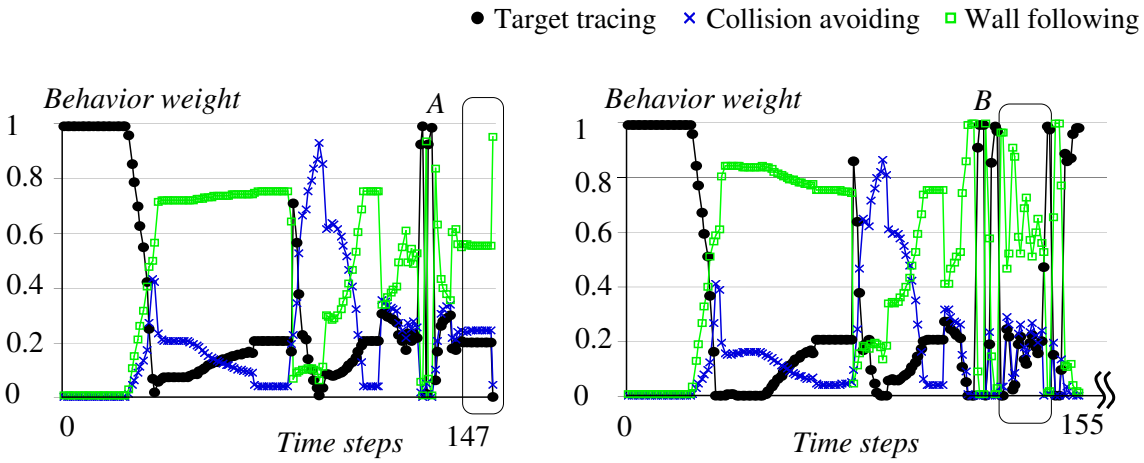


Fig 4.12 The 26th coordination rule and changes of their consequence parts in case 1.



(a) 1st trial (b) 2nd trial

Fig 4.13 Trajectories at each trial in case 2.



(a) 1st trial (b) 2nd trial

Fig 4.14 Changes in behavior weights at each trial in case 2.

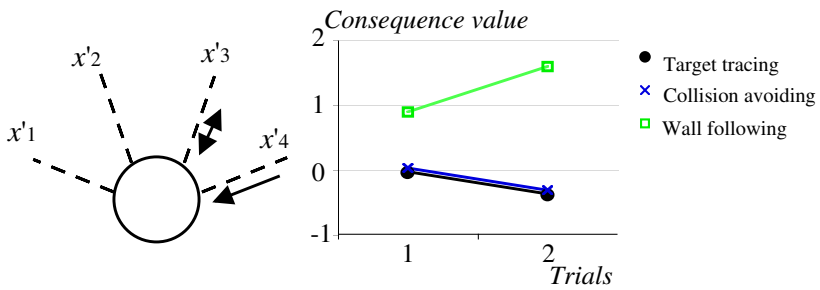
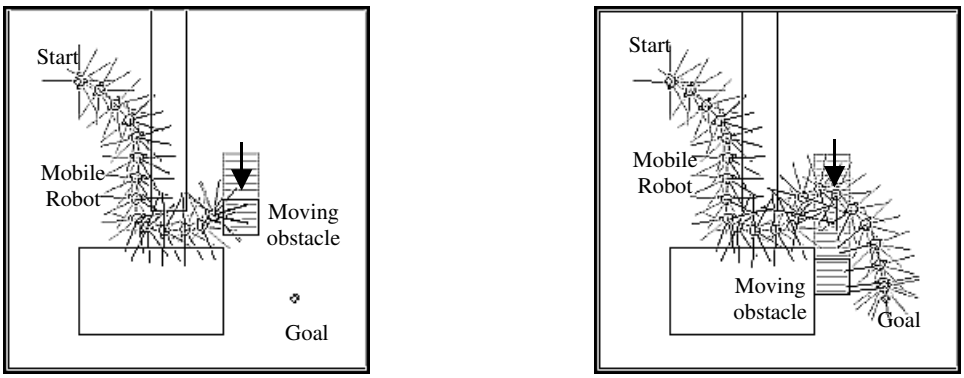
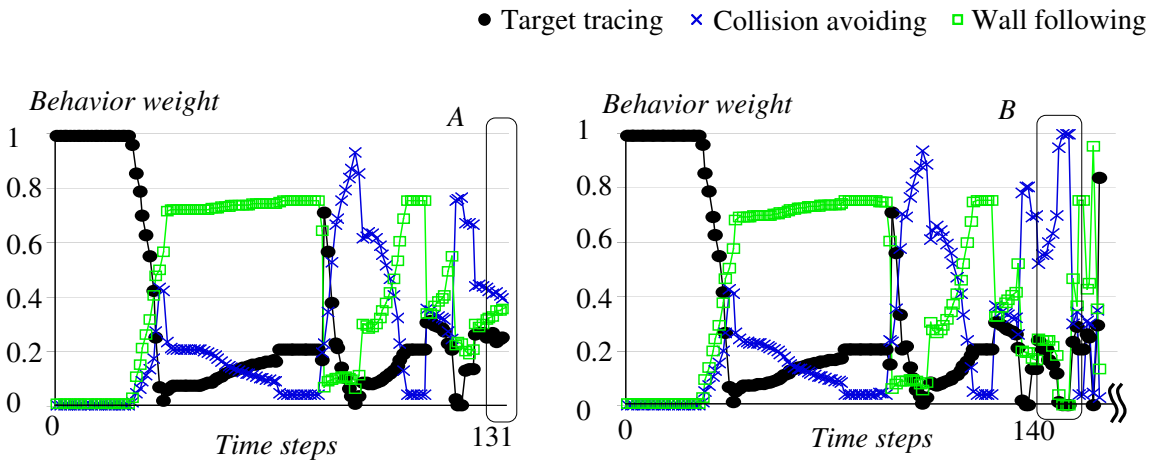


Fig 4.15 The 26th coordination rule and changes of their consequence parts in case 2.



(a) 1st trial (b) 2nd trial

Fig 4.16 Trajectories at each trial in case 3.



(a) 1st trial (b) 2nd trial

Fig 4.17 Changes in behavior weights at each trial in case 3.

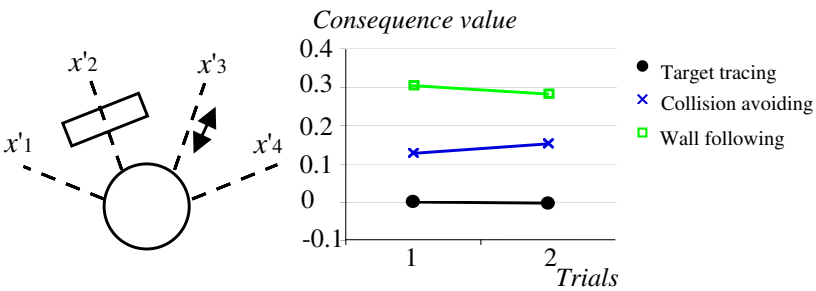


Fig 4.18 The 32nd coordination rule and changes of their consequence parts in case 3.

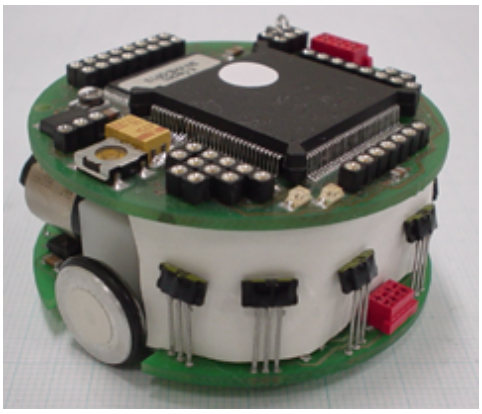


Fig 4.19 Khepera robot.

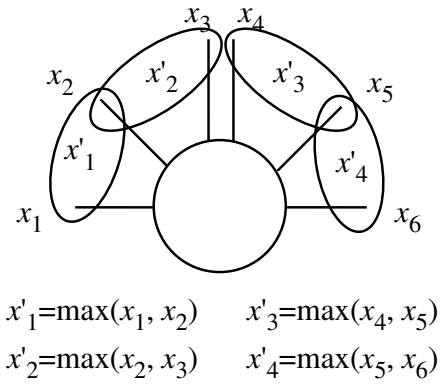


Fig 4.20 Sensor arrangement of Khepera robot.

4.6 実機による実験

局所エピソード学習の有効性を，実機移動ロボット Khepera を用いて検証する（図 4.19）．Khepera は前面から側面にかけて 6 つの赤外線センサを持ち，1〜5[cm] の範囲で障害物との距離が計測できる（図 4.20）．簡単化のため，基本行動として，エンコーダを用いた自己位置認識による目標追従行動と，距離情報による障害物回避行動を用いる．多目的行動調停則には，図 4.20 に示す縮約センサ情報を用い，入力情報として，縮約距離 $x'_i(t)$ とその変化 $\Delta x'_i(t)(= x'_i(t) - x'_i(t-1))$ を用いる．前件部には図 4.21 に示す 3 つの三角型メンバシップ関数を用いた．ただし，後件部出力値に関しては，基本的に目標追従行動の行動重みが 1 になるように設定する．例外として前件部の距離情報に対して，言語ラベル Small を含むルールを障害物回避行動の行動重みが 1 になるように設定する．これにより，Khepera は近傍に障害物を感知しなければ，目標位置に向かうようになり，障害物を感知した場合は，速度を落とすことで障害物を回避することができる．つまり，静的環境下において，この調停則の初期化は，最小限の事前設計と言える．

実験は，2 台の Khepera（$KheperaA, KheperaB$）を図 4.22 に示す環境において，それぞれ，ターゲット T_{A1} と T_{A2}，T_{B1} と T_{B2} を周回するようにタスクを与えた．基本行動や行動調停に関しては 2 台の間で違いはなく，互いに避けあうことができる．ただし，センサレンジ内に突然相手 Khepera が入ってくる場合において行動重みの更新遅れにより衝突する．そこで，$KheperaA$ に局所エピソード学習を導入し，どのような回避動作が学習されるかを検証する．実験では局所エピソードの長さを 3[step] とした．また，衝突時に緊急回避として後進することで，ロボット間でデットロックになることを防ぐ．この後進の瞬間に局所エピソード学習を実行する．

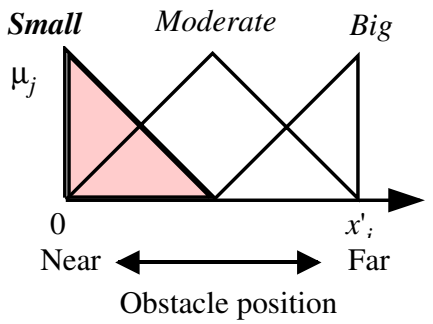


Fig 4.21 Initialization of fuzzy if-then rules.

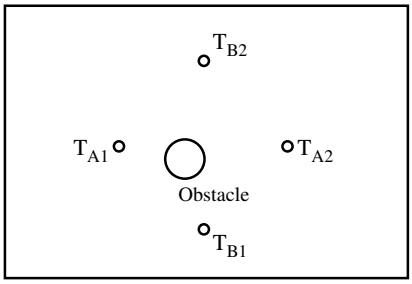
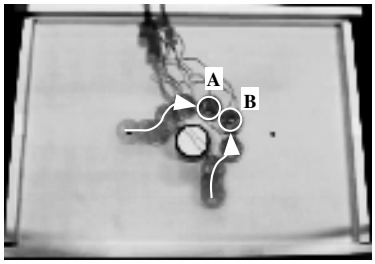
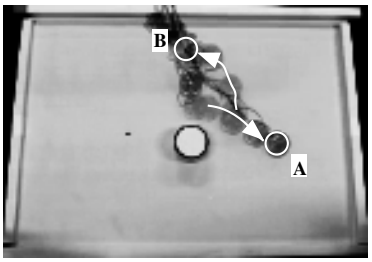


Fig 4.22 Layout of targets and a obstacle.

図 4.23 は，スタートしてから $KheperaA$ が最初に T_{A2} に到達するまでの軌跡である．写真は一秒間隔のスナップショットを重ね合わせている．T_{A1} から T_{A2} まで移動したときの行動重みを図 4.25 に示す．図中 C は，$KheperaA$ が B に衝突した 5[s] あたりの行動重みであり，行動重みの更新が間に合っていないことが衝突の原因と考えられる．このように，スタートから $KheperaA$ は，2 回 $KheperaB$ と衝突し，2 回の局所エピソード学習を行った．図 4.24 は $KheperaA$ が，4 往復目の帰路 T_{A2} から T_{A1} に移動したときの軌道である．図中 45[s] と 50[s] のときに，感知した障害物に対して障害物回避行動の行動重みを早い時期から増加させ，$KheperaB$ が通り過ぎるのを待つような振る舞いが確認できた（図 4.26，D と E の辺り）．スタートしてからターゲット間を 4 往復する間，ロボット同士の遭遇が 5 回確認されたが，最初の 2 回の衝突後，後半 3 回の遭遇は無事回避された．すべての遭遇パターンは似ているが同じものはなく，局所的な適応学習の有効性が示された．

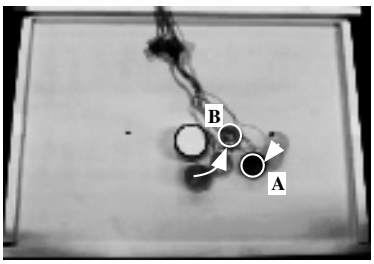


(a) 0[s] - 5[s]

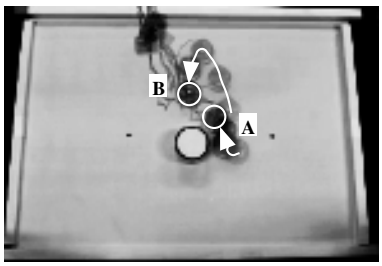


(b) 6[s] - 11[s]

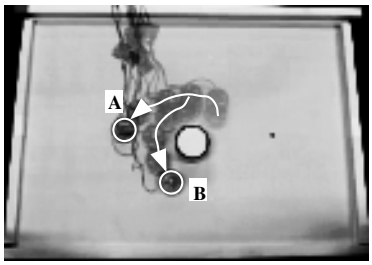
Fig 4.23 Trajectories of Kheperas (0[s]-11[s]).



(a) 41[s] - 45[s]



(b) 46[s] - 50[s]



(c) 51[s] - 56[s]

Fig 4.24 Trajectories of Kheperas (41[s]-56[s]).

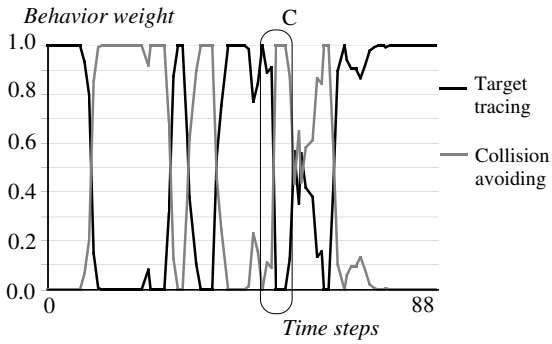


Fig 4.25 Change in behavior weights of Khepera A (0[s]-11[s]).

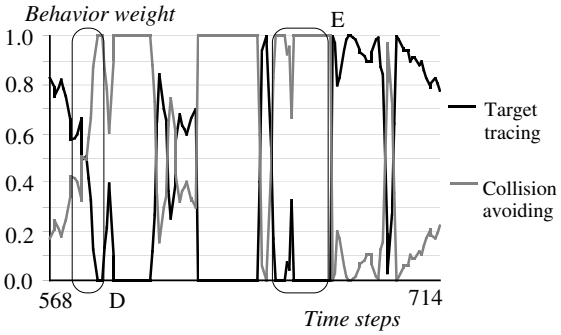


Fig 4.26 Change in behavior weights of Khepera A (41[s]-56[s]).

4.7 結言

移動ロボットにおける多目的行動調停の学習に，局所エピソードを用いた自己参照的な適応学習手法を提案した．提案手法は，明示的な評価関数や，設計者からの教師信号を必要とせず，迅速に移動障害物に対応できることを計算機シミュレーションとKheperaを用いた実ロボットの実験により示した．具体的には，静的環境下で獲得した調停則の多目的性を維持しつつ，移動障害物に対応するために局所エピソード学習を導入し，遭遇した移動障害物に合わせて，行動調停則のファジィルールを部分的に更新し適応できることを示した．提案手法は，この部分的な更新により，基本行動の再学習を回避でき，獲得済み基本行動を再利用できるよう行動調停則を改善できる手法であるといえる．

実環境において，移動障害物との接触のパターンは多岐に渡り，またその接触頻度もそれぞれ異なる．そのため，少ない学習回数で移動障害物に適応できる多目的行動調停則の局所エピソード学習は，実環境においても調停則の学習に有効であると考えられる．しかし，局所的に調停則を改善するため，シナリオ全体としての評価を悪くする場合もある．つまり，おかれた環境と与えられたタスクに適した，シナリオ評価を用いた学習（進化的学習）と，局所エピソードを用いた学習（局所エピソード学習）とを相互に関連づける学習機構が必要であり，今後検討していく予定である．

第5章　実環境下におけるパートナーロボットのための多目的行動調停

5.1　緒言

近年，人間と共生するための様々なロボットが開発されている [84–89]．例えば，日常生活におけるコミュニケーションを解析するために神田らによって開発された Robovie がある．人間とコミュニケーションを行うために十分な表現能力を持つため，認知科学的な知見を得るための実験が可能となる．Robovie は，特定の状態で特定の動作を行う 100 個の状況依存モジュールと，それらが使われる順序を記述する 800 本のエピソードで作り込まれており，様々な状況下で動くことができる [84,85]．しかし，人間とのコミュニケーションにおける行動の獲得やエピソードの学習などは行っていない．セコムは，食事介護が必要な障害者のために，食べ物をすくい口元に運ぶアームロボット「マイスプーン」を開発している [86]．障害者が自ら操作するための簡単なインターフェイスが設計されているが，その操作やロボットアームの動き方に最初に慣れる必要がある．

人間社会において必要とされるロボットの能力として，まず安全に動けるということが重要であるが，さらに，使用者とのコミュニケーションを通して，使用者が好む行動を獲得することが必要とされる．例えば，アームを装備する介護や秘書ロボットを考えると，口元に食べ物を運ぶことや書類を手渡しするといったタスクが想定できる．これらのタスクにおいてアームの先端の動き方や終端位置などは，従来の軌道生成手法を用いると，使用者に恐怖やいら立ちを与えかねない．そこで本研究では，単一の手渡し動作に関して，対話型遺伝的アルゴリズム (interactive genetic algorithm; IGA) [23,24] を用いた軌道生成について議論している [101,102,111]．使用者の評価モデルを状態価値関数により近似し，使用者が好むであろう軌道をロボットが提示し，使用者がそれを評価する．これを繰り返すことで使用者にとって評価の高い軌道を獲得する．

さらに，進化的計算を実行している間に，良好な関節角度列を教師値として，基本行動を NN により獲得することで，他の基本行動と協調した多目的な動作を試みる．動作生成には，多目的行動調停を適用し，人間とロボットとの距離の変化に合わせて，基

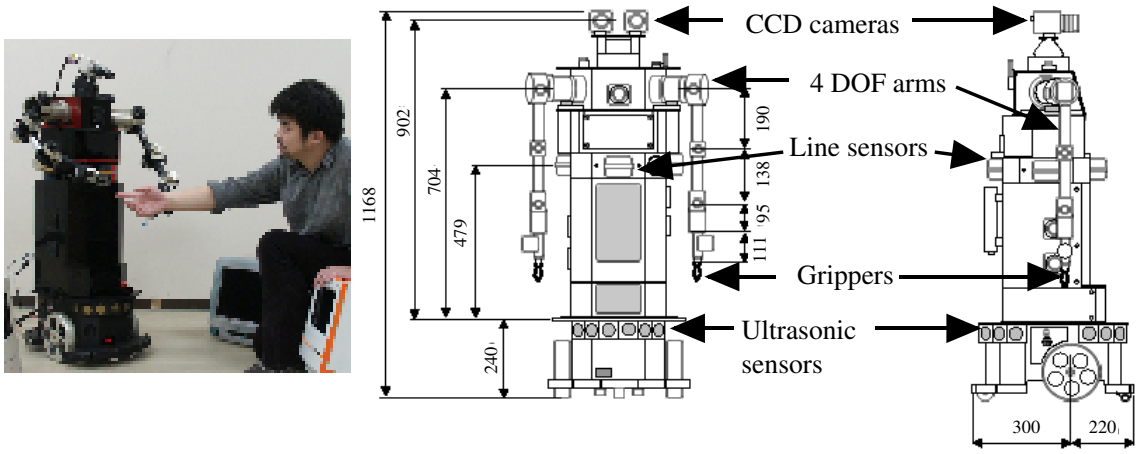


Fig 5.1 Human-friendly robot Hubot.

本行動に対する行動重みを更新することで，異なる接近の仕方で異なる動作が出来ることを示す．また，手渡し動作時の人間との距離を学習する方法を提案する．

5.2 パートナーロボットの構成

パートナーロボットを構築するために，本研究では，図5.1の人型ロボットHubotを採用する．Hubotは，グリッパを含む6自由度のアームを2本持ち，その他にも，CCDカメラを載せるパンチルトと，2次元平面を移動する台車で構成される．外界センサとして，CCDカメラ，赤外線ラインセンサ，グリッパ部赤外線センサ，台車の超音波センサを有する．これらのアクチュエータとセンサは2台の組み込みPCで制御される．

アームのモデル化

アームを用いた動作のためのモデル化を行う．グリッパの閉開および手首の回転を除く4自由度の関節角度で姿勢θを構成する．

$$\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^T \tag{5.1}$$

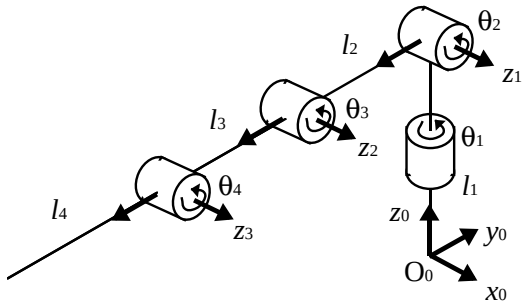


Fig 5.2 Link configuration of Hubot arm.

Table 5.1 Link parameters of Hubot arm.

Link No.	θ_n	d_n	l_n	α_n
1	θ_1	d_1	0	90
2	θ_2	0	l_2	0
3	θ_3	0	l_3	0
4	θ_4	0	l_4	0

この関節角を用いて，ベース位置から手先位置（グリッパ位置）を計算する．図 5.2 と表 5.1 で表されるアームの同次変換行列は，

$$
\begin{aligned}
{}^{i-1}_{i}A &= Rot\,(z,\,\theta_i)\,Trans\,(0,0,d)\,Trans\,(l,0,0)\,Rot\,(x,\alpha_i) && (5.2)\\
&= \begin{bmatrix} \cos\theta_i & -\sin\theta_i\cos\theta_i & \sin\theta_i\sin\alpha_i & l_i\cos\theta_i \\ \sin\theta_i & \cos\theta_i\cos\alpha_i & -\cos\theta_i\sin\alpha_i & l_i\sin\theta_i \\ 0 & \sin\alpha_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} && (5.3)
\end{aligned}
$$

と表すことができる．Rot と $Trans$ はそれぞれ，回転変換行列と並列変換行列を意味する．この同次変換行列を用いて，ベース座標フレーム $\{L_R\}$ からみたアーム先端の座標フレーム $\{L_N\}$ の同次変換行列 ${}^{R}_{N}T$ は，次式のように表される．

$$
{}^{R}_{N}T = {}^{R}_{1}A^{1}_{2}A^{2}_{3}A^{3}_{N}A \tag{5.4}
$$

Hubot のアームは，位置制御を用いており角度情報から姿勢を生成し，ある姿勢から次の姿勢への移動は PID 制御により行われる．

5.3 対話型進化的計算に基づく軌道生成

使用者に適した手渡し動作を獲得するために，使用者の評価モデルを同定しつつ，対話的に軌道を生成する手法を提案する．図 5.3 に，軌道生成の一連の流れを示す．まず，軌道生成の前段階として，赤外線ラインセンサからの距離情報と，CCD カメラの時系列画像情報から，対面する人間を検出する．そして，検出した人間の位置から，なるべく前方方向に軌道が生成されるようにボーナスゾーンを設定する．

軌道生成は，関節角度情報を個体とする進化的計算により行う．内部評価（inner evaluation）として，計算機シミュレーションにより，個体の遺伝的操作と軌道の生成及び評価を繰り返えす．一定回数の評価の後，使用者に提示する軌道をボルツマン選択により選択し，実際にアームを動かし使用者に評価（outer evaluation）してもらう．そして，使用者が提示された軌道に対して与えた評価値に基づき，評価モデルを更新し，内部評価の評価関数に反映させる．評価モデルは，ファジィ状態価値関数（fuzzy state value function）を用いて構築し，更新には利益共有法（profit sharing plan）を用いる．内部評価の繰り返しは，使用者の評価のモデルに基づき，より好まれるであろう軌道の探索を行い，外部評価の繰り返しは，状態価値関数を洗練（詳細化）する．

以下，使用者の検出とボーナスゾーンの設定，手渡し動作獲得のための対話型進化的計算，使用者の評価モデルの学習について順に説明する．

使用者の検出とボーナスゾーンの設定

使用者の位置を考慮しないアームの軌道は，使用者がロボットに合わせる必要があるだけでなく，人間に危害を加える可能性もある．そこで，使用者にとって使いやすい軌道が，使用者とロボットとの間にあると仮定し，赤外線ラインセンサとカメラ画像から使用者の位置を検出する．赤外線ラインセンサは，2 度おきに 90 方向の距離が計測でき，ロボットの前方 180 度内にある障害物の距離が計測できる．ただし，障害物と人間とを区別することが難しい．そこで，現在の画像と一時刻前の画像を差分することで，画像の中から変化のある方向を計測する．その変化のある方向を使用者がいる方向と仮定し，赤外線ラインセンサ情報と合わせて，使用者の位置を特定する．

使用者の中心座標 o” とロボットの中心座標 o’ との中間領域 $BZ(M)$ に，手渡し易い目標領域があるもの仮定し，その領域に到達できる軌道を生成し易くする為のボーナスゾーンを設定する (図 5.4)．ボーナスゾーンを大きく外れる遠い位置を通る軌道や，使用者に近づきすぎる軌道は，内部評価によって淘汰される．

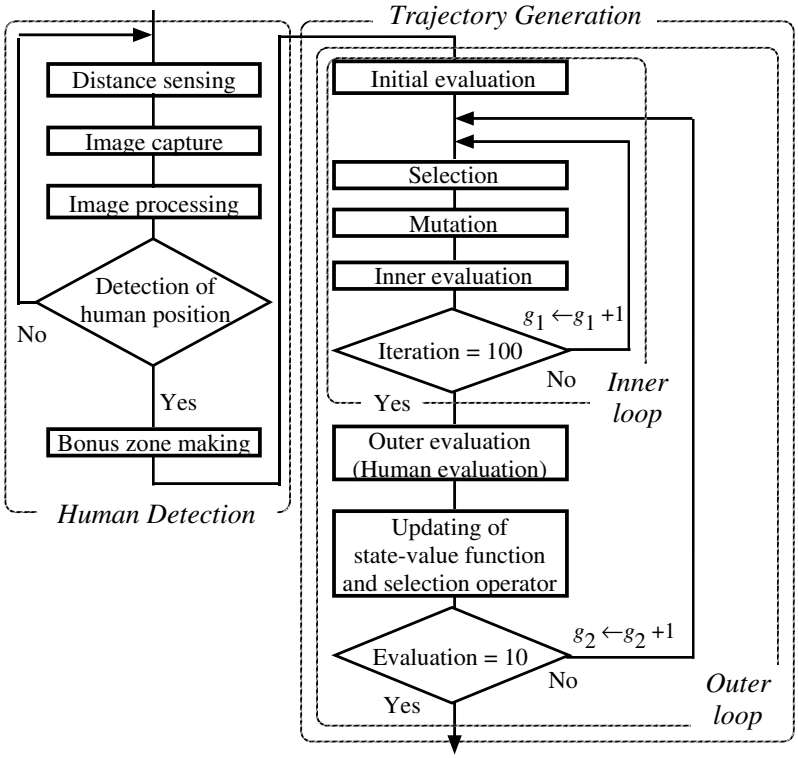


Fig 5.3 Flowchart of calculation process for human-friendly trajectory generation.

$$B_t = \begin{cases} 1 & if\ l(t) < D(t) \\ \dfrac{D(t)}{l(t)} & otherwise \end{cases} \tag{5.5}$$

$D(t)(t = 1, 2, .., M)$ はボーナスゾーン $BZ(t)$ の半径，$l(t)$ をアームの手先位置からボーナスゾーンの中心までの距離とする．このボーナスゾーンの中心は，目標位置とアームの初期手先位置とを結ぶ直線上とし，徐々に半径を狭めるものとする．

進化的計算による軌道生成

これまで，アームにおける Point-to-Point の軌道生成に関して盛んに研究されており，逆運動学を解く方法や GA を用いた方法など提案されている [92]．GA による軌道生成は，必要な性能を評価関数として設計し，順モデルの計算結果を評価するだけで行うことができる．解候補である GA の個体は，時間ステップ $t\ (t = 1, 2, ..., M)$ におけるアームの各関節に対する角速度 $\theta\ (\theta_1, \theta_2, .., \theta_N)$ を遺伝子としてコーディングする．アームの自由度を N，中間姿勢を M 個とすると，1 個体は $M \times N$ の実数値で構成できる (図 5.5)．この個体群に対して，遺伝的操作を加え手渡し動作を生成する．本稿では，一世代に最小適応度個体のみを淘汰する定常状態遺伝的アルゴリズム (steady-state

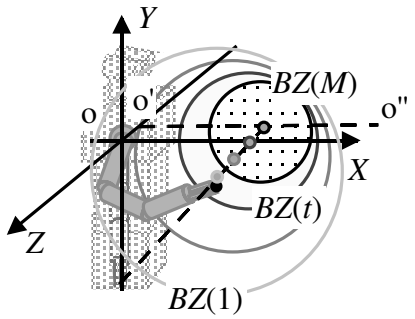


Fig 5.4 Setting of bonus zone.

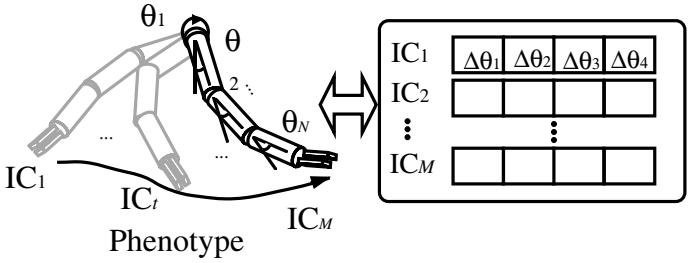


Fig 5.5 Coding of manipulator configuration.

genetic algorithm; SSGA) [22] を適用する．最小適応度個体を淘汰するため，一旦実行可能な解が求まれば，それを保持あるいは改善することができ，使用者を危険にさらすことがない．選択にはボルツマン選択を用い，最良個体とのエリート交叉により新しい個体を生成する．ボルツマン選択は，評価値と温度係数 T に依存した確率 p_j により行う．

$$p_j = \frac{\exp(fitness_j/T)}{\sum_{k=1}^{L} \exp(fitness_k/T)} \tag{5.6}$$

ここで，L は解候補の総数を表す．温度係数 T を高くすることで，よりランダムに近い選択を行い，低くすることで，確定的なグリーディ選択を行う．突然変異には，実数値によるコーディングのため適応型突然変異を用いる．評価関数は，使用者の評価をモデル化したファジィ状態価値関数の値と，ボーナスポイント，さらに不要な関節回転を削減するための項によって構成される．以上をまとめて，j 番目の個体に対する評価値を次式の評価関数により計算する．

$$fitness_j = \omega_1 \sum_{t=1}^{M} S_t + \omega_2 \sum_{t=1}^{M} \gamma^{M-t} B_t + \omega_3 \frac{1}{MN} \sum_{t=1}^{M} \sum_{i=1}^{N} \frac{1}{1+|v_{t,i}|} - P \tag{5.7}$$

ω と γ は，重み付け係数及び割引率とする．また，P はペナルティ項であり，アームがロボット自身に衝突することを罰する．この軌道生成は最大化問題となる．

外部ループである IGA において，上述の SSGA を用いて一定世代探索を行った個体群から，ボルツマン選択により使用者に提示する軌道を選択する．選択した軌道を提示し，使用者に 1 から 10 ポイントで評価をしてもらう．この使用者の評価値の履歴から，ボルツマン選択における温度係数を変更することで，探索戦略の切り替えを行う．

$$T(g_2+1) = \begin{cases} T_{min} & if\ R(g_2) > R(g_2-1) \\ T(g_2)\,d & otherwise \end{cases} \tag{5.8}$$

T_{min} と $R(g_2)$，d はそれぞれ，温度係数の最小値，g_2 回目の外部評価を正規化した値 $(0.0 \leq R \leq 1.0)$，変数 $(1 < d)$ とする．使用者に提示された軌道が前の軌道よりもよく，高い評価を与えた時には，温度係数を下げ選択圧を高くし，逆に，前回の評価よりも悪くなった時は，徐々に温度係数が高くなるようになっている．つまり，より良い評価が続けば，局所探索となり軌道のファインチューニングを行い，評価が悪くなることによって，大域的な探索を行うようになる．さらに，ファジィ状態価値関数が評価値に基づき学習することで，より詳細な使用者の評価モデルを学習する．ファジィ状態価値関数については，次項で説明する．

人間の評価モデル：ファジィ状態価値関数の学習

人間の評価は曖昧であり，実際に見た軌道しか評価できない．また，評価の繰り返しにより，使用者の疲労が生じる．そこで，アームの手先座標 (x, y, z) を入力として用い，前件部を三角型ファジィメンバーシップ関数の組み合わせにより構成し (図 5.6)，後件部を実数値とする簡略化ファジィ推論により状態価値 S_t を見積もる．j 番目のファジィルールの前件部適合度を $\mu_{j,t}$，後件部出力値を s_j とすると，状態価値 S_t は次式のように計算できる．

$$S_t = \frac{\sum_{j=1}^{J} s_j \mu_{j,t}}{\sum_{j=1}^{J} \mu_{j,t}} \tag{5.9}$$

ただし，J はファジィルールの総数とする．

使用者は，提示された軌道に対して数値で評価を行い，その評価値を用いてロボットは状態価値関数を学習する．学習には，利益共有法 [15,16] を適用し，人間の評価値を [0,1.0] に正規化した値 R を報酬として用いる．

$$s_j \leftarrow s_j + \alpha^{M-t} \tau \mu_{j,t} (R - S_t) \tag{5.10}$$

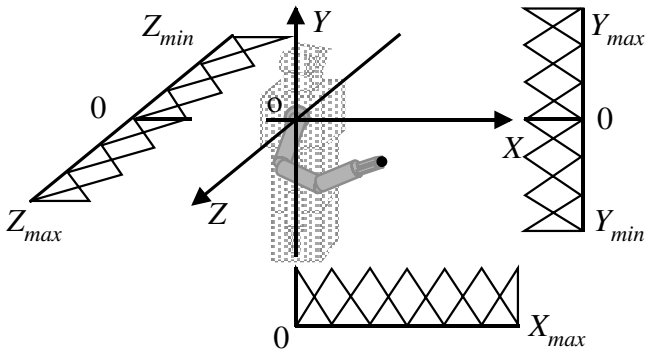


Fig 5.6 State space decomposition for the fuzzy state value function.

α と τ は，それぞれ減衰係数 $(0 < \alpha \leq 1.0)$ と学習係数とする．与えられた報酬によって，中間姿勢と最終姿勢の手先位置に報酬が分配される．この状態価値関数の値が内部評価の評価関数に反映されることで，使用者の評価が高くなるであろう軌道が探索される．

5.4 ニューラルネットワークを用いた基本行動の学習

使用者にとって評価の高い軌道には，その使用者固有の何らかの傾向があり，これらを抽出し基本行動として獲得することは，ヒューマンフレンドリなロボットとして様々なタスクに再利用できうる．本節では，IGA により生成された軌道から，NN を用いて基本行動を学習する手法を説明する．

一定の評価が得られた GA の個体から，アームの姿勢 $\theta(\theta_1, \theta_2, \theta_3, \theta_4)$ に対する変化量 $\Delta\theta(\Delta\theta_1, \Delta\theta_2, \Delta\theta_3, \Delta\theta_4)$ の組みを教師値として生成し，一定個数保持する．ここで，GA で生成される軌道から学習できる行動は，用意する NN の構造に依存するが，手渡し動作において，急激な方向転換や周期的な動きがないものとし，本稿では階層型 NN を適用する．入力を N 個の関節角 θ，出力を N 個の変化量 $\Delta\theta$ として構成する．学習には慣性項付き逆誤差伝播法を用いる．教師値 θ^* との誤差 E を，

$$E(t) = \frac{1}{2}\sum_{i=1}^{N}\left\{\theta_i^*(t) - \theta_i(t)\right\}^2 \tag{5.11}$$

と表し，NN の結合強度を慣性項付き BP で学習する．内部ループでの進化の過程の間，オンラインで教師値を作成し学習することにより，可動範囲内において，汎化性のある実行可能な基本行動モジュールが獲得できると考えられる．これは，評価値が同じ場合であっても，少しづつ異なる経路を通る軌道が生成されるためである．

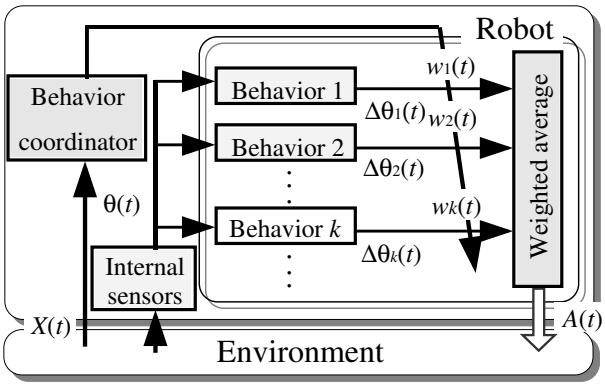


Fig 5.7 MOBC for Hubot arm.

5.5 多目的行動調停に基づく行動制御と学習

コミュニケーションにおいて，我々人間は，互いの距離やこれまでの動作系列から次に取るべき行動が制限されていると考えられる．またその動作は，複数の目的を満たすような複合的な動作である場合が多い．挨拶を例にとれば，遠い間合いでは手をあげ，近くに居る時は握手したり，さらに接近していれば肩を叩いたり包容したりする．これらの行動は単独で実行される場合もあるが，距離に依存して，手を挙げて挨拶した手を下ろすと同時に握手しだしたり，握手するように近づきながら包容したりする．これらの状況は容易に想像することができる．逆に，握手しようと手を差し伸べてきている相手に対して，手を挙げて挨拶したりはしない．つまり，コミュニケーションを行う時，何かしらの文脈に沿ったトップダウン的な行動の融合がなされていると考えることができる．

多目的行動調停は，目的毎に設定された基本行動に対して，それぞれに行動重み $W = (w_1, w_2, w_3, ...)$ を割り当て，基本行動の出力を重み付け平均することにより，実際の動作出力を決定する手法である．これを使用者との距離に基づく手渡し動作に適用する．図 5.7 に，多目的行動調停の概念図を示す．まず，内界センサ情報 $\theta(t)$ に対して，基本行動 NN により各関節の変化量が計算される．次に，各基本行動に対し，それぞれ行動重みを与え各出力に重み付けを行う．行動の総数を K すると，i 番目 $(i = 1, 2, 3, 4)$ の関節角は次式で計算される．

$$A_i(t) = A_i(t-1) + \frac{\sum_{k=1}^{K} w_k(t)\Delta\theta_{k,i}(t)}{\sum_{k=1}^{K} w_k(t)} \tag{5.12}$$

生成される軌道は，行動調停則に基づき，行動重みを逐次的に更新することで状況の変化に対応する．行動調停則は，外界センサ情報 $X(t)$ を入力とし行動重みを計算する

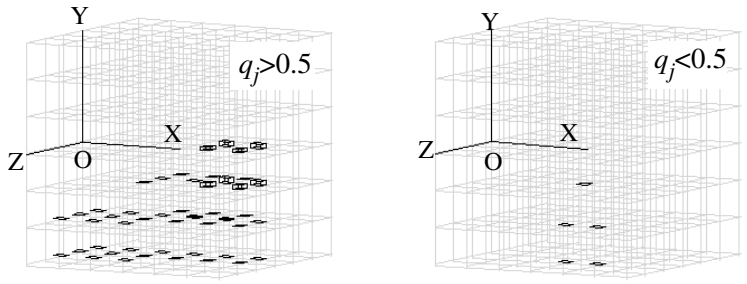


Fig 5.8 Inference landscape of the fuzzy state-value function.

が，本研究では，人間の位置を CCD カメラの画像情報と赤外線ラインセンサの距離情報で検出し，使用者の移動に対して更新する．

5.6 計算機シミュレーション

5.6.1 軌道生成と基本行動獲得

人型ロボット Hubot のアームの軌道生成を，計算機シミュレーションにより模擬し，提案手法の有効性を確認する．

手渡し動作のタスクを与えるにあたり，4 自由度のアームを想定する ($N = 4$)．手渡し動作は 6 つの中間姿勢で構成するものとする ($M = 6$)．中間姿勢は，1[s] 毎に角速度から計算しダイナミクスは考慮しない．ロボット前方に使用者が居るものとしボーナスゾーンを設定する．GA の個体数を 100 とし，100 回の評価毎に，使用者に軌道を提示するものとする．ただし，初期探索として 1000 回まで使用者への軌道の提示は行わない．評価関数の係数は，$\omega_1 = 0.8, \omega_2 = 0.19, \omega_3 = 0.01$ と設定する．提示された軌道は，1 から 10 の整数値で評価し，最も好ましい軌道を 10 とし，評価値 10 が得られた時を軌道生成の終了とする．ファジィ状態価値関数は，手先座標に対して，それぞれ 7 つのメンバシップ関数を用意し，後件部出力値をすべて 0.5 で初期化する．また，NN の学習係数と慣性係数は，0.05 と 0.01 に設定し，評価値が 0.65 を上回る 20 個体を常に保持しオンラインで学習する．さらに，軌道生成終了後に，個体群中上位 20 個体を用いて 1000 回追加学習を行う．

図 5.8 に軌道生成後のファジィ状態価値関数の後件部出力値を示す．図より，終端のボーナスゾーン付近と，その位置までの経路沿いの価値が高く (図 5.8(a))，ボーナスゾーンから外れた場所や腕を回り込ませるような領域の価値が低くなっている (図 5.8(b))．図 5.9 に GA の最良個体による軌道と，学習された NN による軌道を示す．NN

の時間ステップ幅は，0.5[s] としている．GA による経路と終端位置に準じた軌道が NN により生成されていることが確認でる．また GA の適応度の変化 (図 5.10) と NN の学習誤差の変化 (図 5.12) からも基本行動として学習できていることが解る．ここで，図 5.12 における 2400 回あたりの学習誤差の急増は，より良い軌道が評価されたために生じていると考えられる．

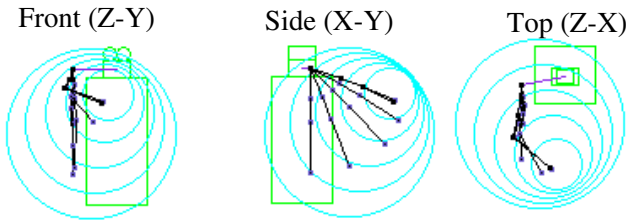


Fig 5.9 Trajectory by GA trajectory generation.

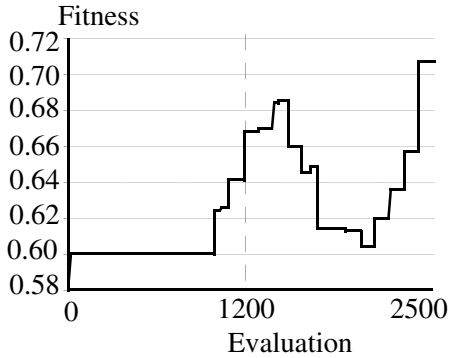


Fig 5.10 History of fitness by GA trajectory generation.

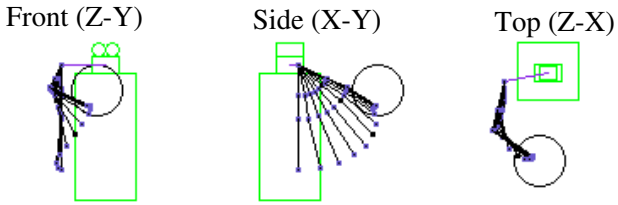


Fig 5.11 Trajecotry by learned NN behavior module.

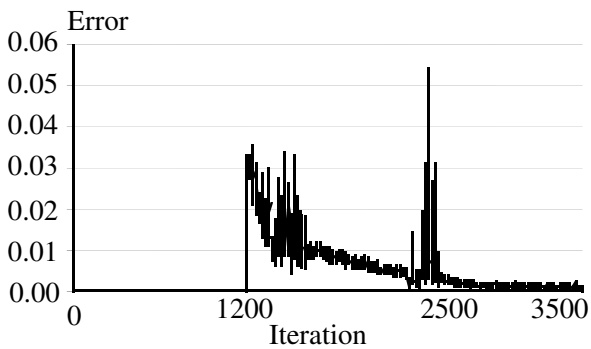


Fig 5.12 History of error in NN behavior learning phase.

5.6.2 使用者との距離に基づく多目的行動調停

多目的行動調停を行うにあたり，手渡し行動とは異なる基本行動を獲得する．エンドエフェクタが最も高い位置にあがるように評価関数を設定し，呼応行動（hand raising behavior）の軌道生成を行った．また，最も基本的な行動として，腕を下ろし初期位置で待機する待機行動（home positioning behavior）の獲得を行った．この行動の獲得は，手渡し行動と呼応行動の獲得時に同時に獲得できる．各軌道における終端姿勢に関わらず，生成された GA の解候補の関節角度情報を用いて，終端位置から初期位置に戻るように教師値を生成し，オンラインで基本行動 NN を学習する．これら 3 つの基本行動を用いて，近付いて来た使用者にモノを手渡しするタスクを想定する．このタスクにおいて，外界センサ情報により，人間の位置に基づき行動重みを図 5.13 のように変化させる．この行動重みの変化に基づいて調停した結果を図 5.14 に示す．呼応動作から待機行動，待機行動から手渡し行動，手渡し行動から待機行動へと動作が融合されていることが確認できる．

次に，人間の近づく速度が速い場合の実験結果を示す．この場合の，行動重みの変化を図 5.15 に示す．行動調停の結果生成された軌道を図 5.16 に示す．呼応行動から手渡し行動までの行動重みの大きな変化によって，待機位置まで移動する前に，呼応行動から手渡し行動へ一連の動作が接続されたことが確認できる．これらの実験から，人間との距離の変化に依存して，基本行動 NN の出力を融合することで，異なる状況で適切であろう動作が生成できることを示した．

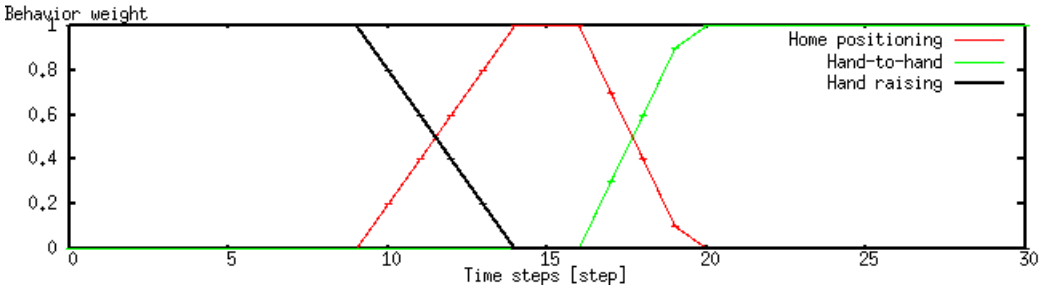


Fig 5.13 Simulation result of change in behavior weight in case 1.

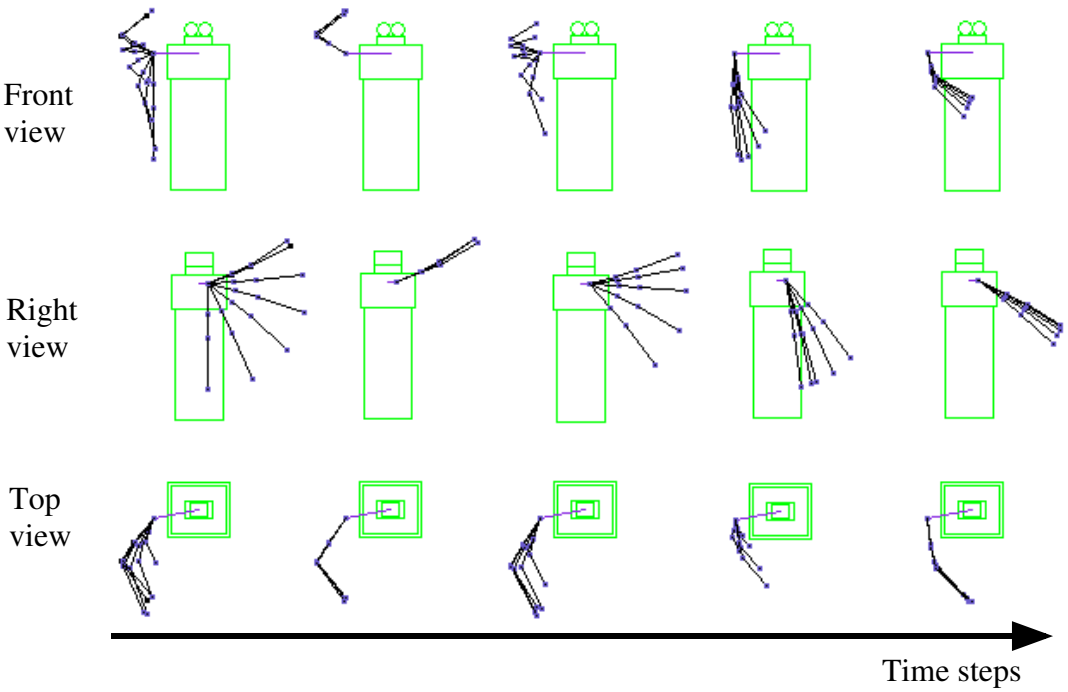


Fig 5.14 Trajectory of multi-objective action by coordinated behaviors in case 1.

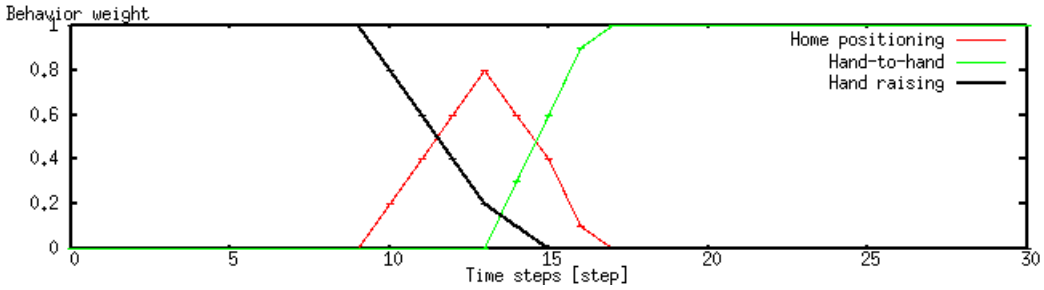


Fig 5.15 Simulation result of change in behavior weight in case 2.

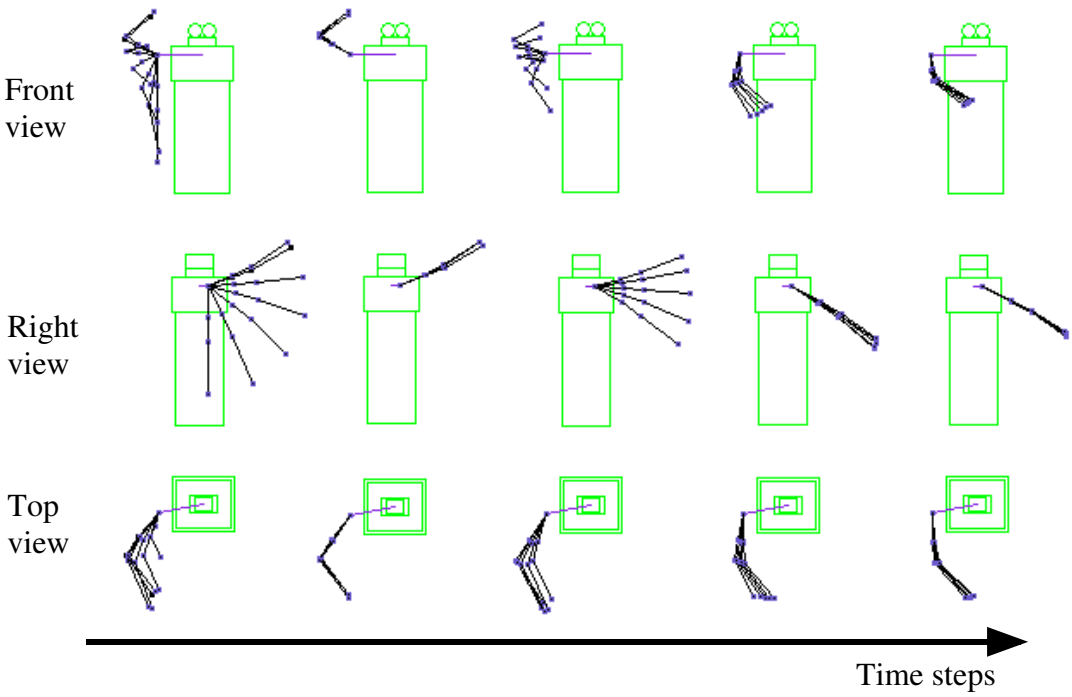


Fig 5.16 Trajectory of multi-objective action by coordinated behaviors in case 2.

5.7 実機による実験

5.7.1 手渡し動作の軌道生成

実機 Hubot を用いて，実験を行うにあたり，ボーナスゾーンの設定には，CCD カメラと胸部の赤外線ラインセンサを用いる．進化的計算で用いる個体数，内部評価回数はそれぞれ，100 個体と 100 回評価とする．ある程度，ボーナスゾーンに入るような軌道を最初に提示するために，初期評価を 2000 回行う．初期評価時には，(5.7) 式の係数 ω_1，ω_2，ω_3 を 0.0, 0.99, 0.01 に設定する．人間の評価（外部評価）が行われる度に，ω_2 を減じることで，徐々に手先位置の状態価値を反映させ，ボーナスポイントを無視するように設定する．

$$\omega_1 = 1.0 - 1.0/(g_2 + 1) \tag{5.13}$$

$$\omega_2 = 1.0/(g_2 + 1) - 0.01 \tag{5.14}$$

$$\omega_3 = 0.01 \tag{5.15}$$

ファジィ状態価値関数の後件部出力値 q は，全て 0.5 で初期化する．なお学習係数は 0.8 とする．状態の分割は，各軸に対して 5 つの三角型メンバシップ関数を用いる．

最初に，被験者に Hubot に近づいてもらう．被験者の検出結果を図 5.17 に示す．Hubot は，CCD カメラから取り込んだ画像データの時系列から，移動物体の方向を検出する．移動物体を人間とみなし，その方向に対する距離を胸部赤外線センサから計測し，ボーナスゾーンを設定する．図中のボーナスゾーンは一番内側のゾーンを表す．このボーナスゾーンを人間とロボットとの中間領域に設定することで，手渡ししやすい位置に軌道を生成するようになる．なお，高さ方向は検知できないため，ボーナスゾーンは，円筒として構成する．

図 5.18 に，被験者による評価（1 回目から 4 回目）と，そのとき提示した軌道を示す．被験者は，1 回目の提示された軌道に対しては，評価値 4 ポイントを与え，順に，8，7，8 ポイントを付けた．そして，5 回目に提示された軌道（図 5.19）に対して，10 ポイントを与え，手渡し動作の軌道生成は終了した．内部評価の履歴を図 5.20 に示す．最初に提示された軌道が，被験者の目の高さを通るものであり（図 5.18(a)），低い評価を与えた．そのため，2 回目の提示からは胸の高さあたりになり，2 回から 4 回目の評価によって，より詳細な軌道が探索されている．最大の評価を得た 5 回目の軌道を見ると，低い位置で被験者の前方まで手を回し（図 5.19(a) から (d)），最後に差し出すように前に出している（図 5.19(e)）．

軌道探索と同時に得られた状態価値関数の後件部出力値の分布を図 5.21 と 5.22 に示す．左図が評価の高い領域を表し，右図が評価の低い領域を表す．また図中のボックス

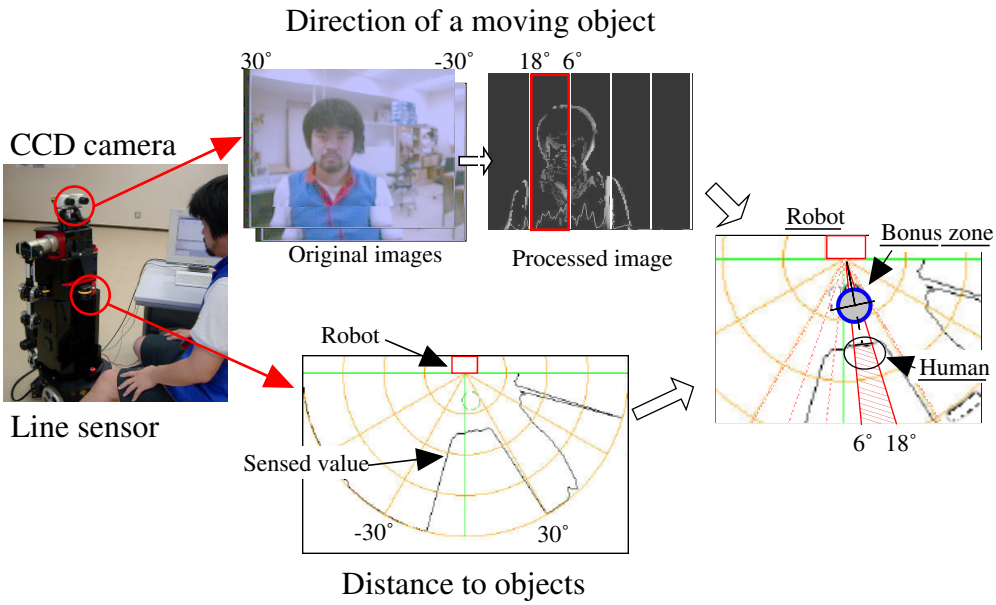
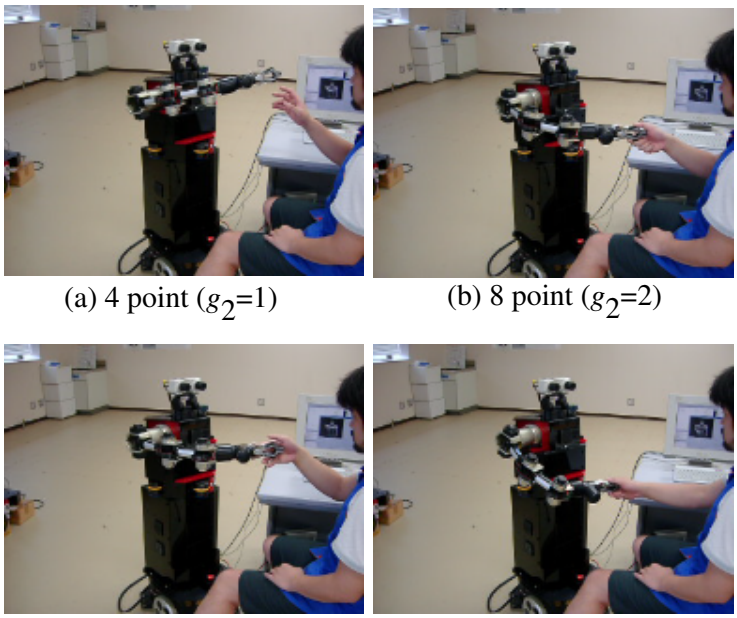


Fig 5.17 Human detection and setting of bonus zone.

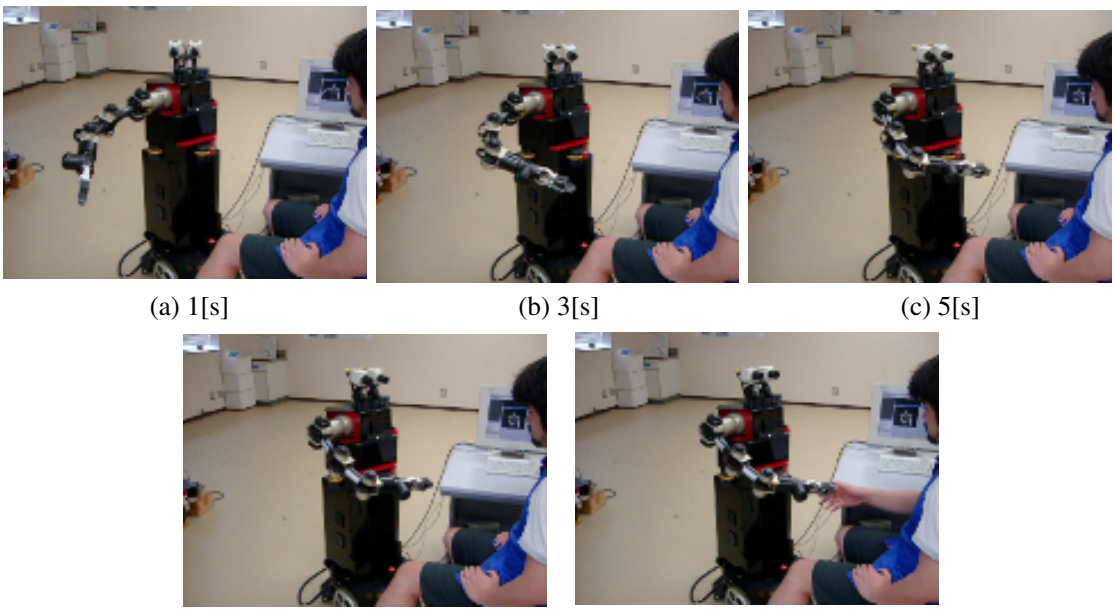
の高さで，評価値 s_j の大きさを表す．最初の軌道提示において，目の高さを通る軌道（図 5.18(a)）に対し低い評価を与えた．これにより，図 5.21(a) に示すように状態価値関数が学習され，次の内部評価から目の高さを通る軌道が淘汰されやすくなる．2 度目の外部評価によって，肩の高さを通る軌道が強化される（図 5.18(b)，図 5.21(b)）．つづく 3 回目の外部評価で，2 回目と同じような軌道が提示され，2 回目よりも低い評価値により平滑化され（図 5.22(c)），同時にボルツマン選択の温度係数が高くなり，大域的な探索に切り替わる．これにより，4 回目の軌道提示で，いままでよりも低い位置を通る軌道が提示され，高い評価値により状態価値関数が強化され，同時に温度係数が低下することで内部ループで局所探索を行い，5 回目の外部評価で最大評価値の軌道が提示された．最終的に目の高さの領域を通る軌道が低い評価になり，胸の辺りの領域で且つ，被験者との中間領域を通る軌道が高い評価になるように評価モデルが同定されていることが確認できる．



(a) 4 point (g_2=1) (b) 8 point (g_2=2)

(c) 7 point (g_2=3) (d) 8 point (g_2=4)

Fig 5.18 The snapshot of final configuration and human evaluation value at 1st-4th outer evaluations.



(a) 1[s] (b) 3[s] (c) 5[s]

(d) 7[s] (e) 10[s]

Fig 5.19 The snapshot of final configuration at 5th outer evaluation.

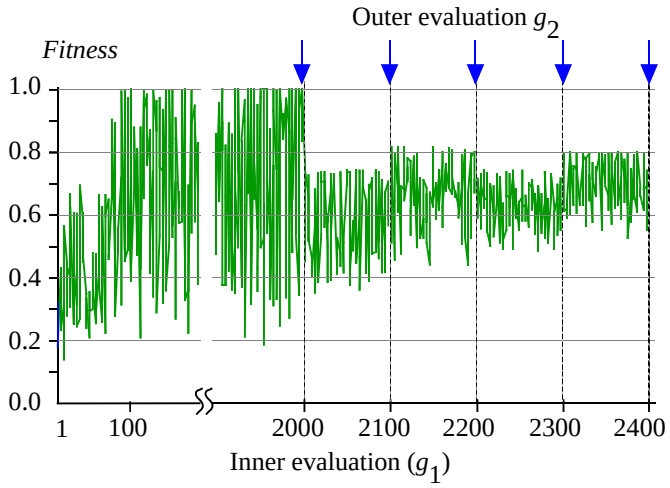
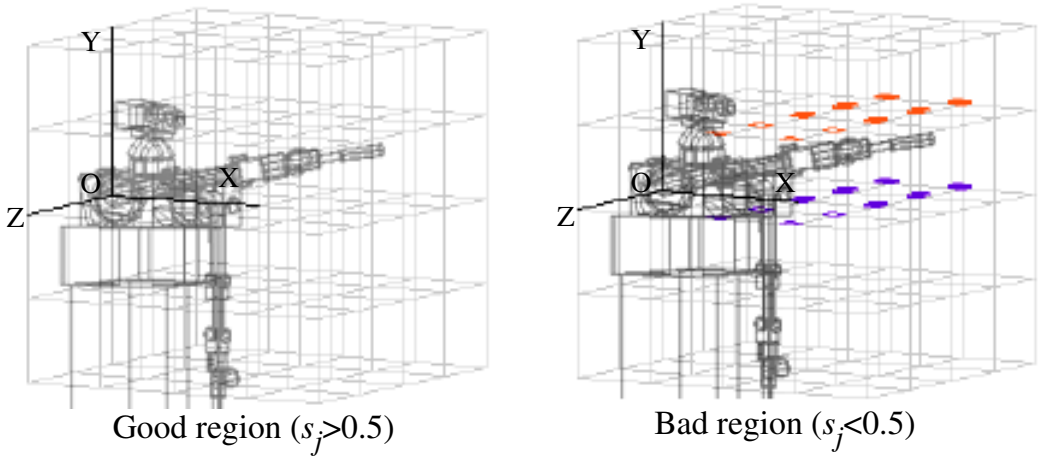
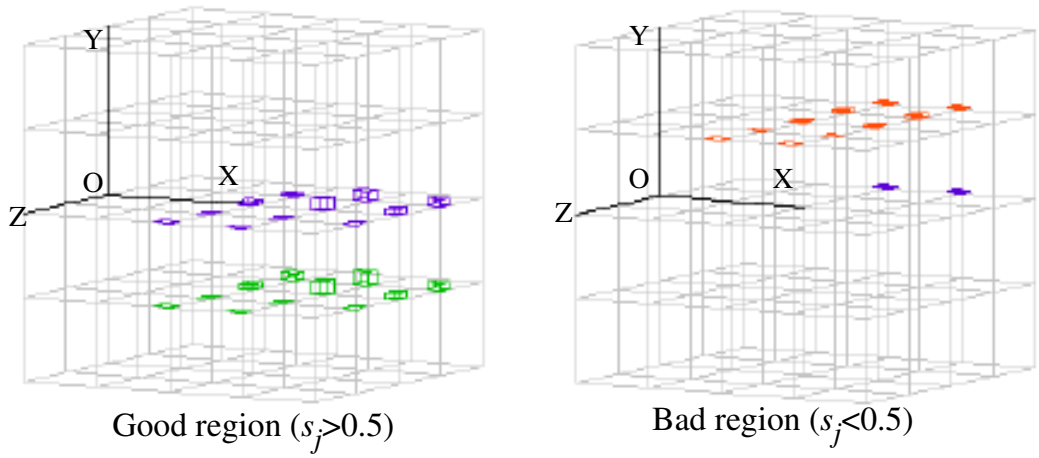


Fig 5.20 The change of fitness in inner evaluation.

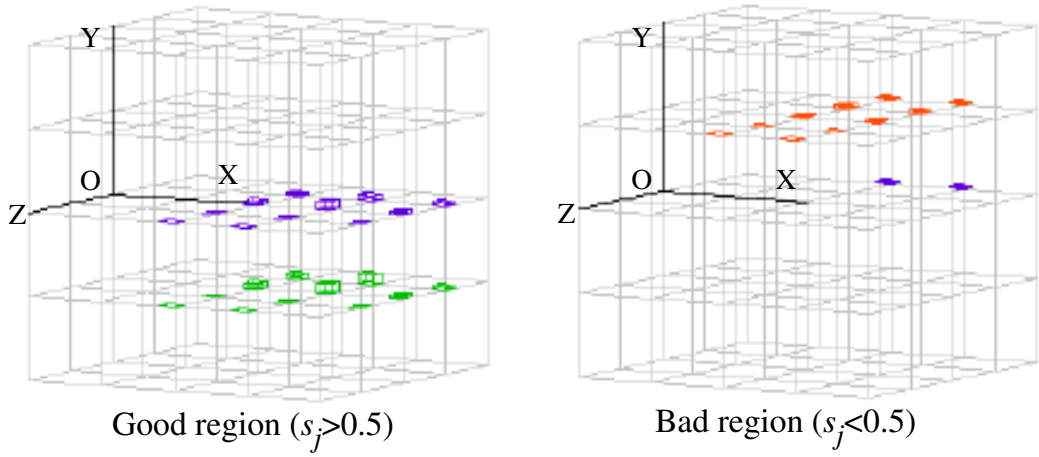


(a) First human evaluation (4 point)

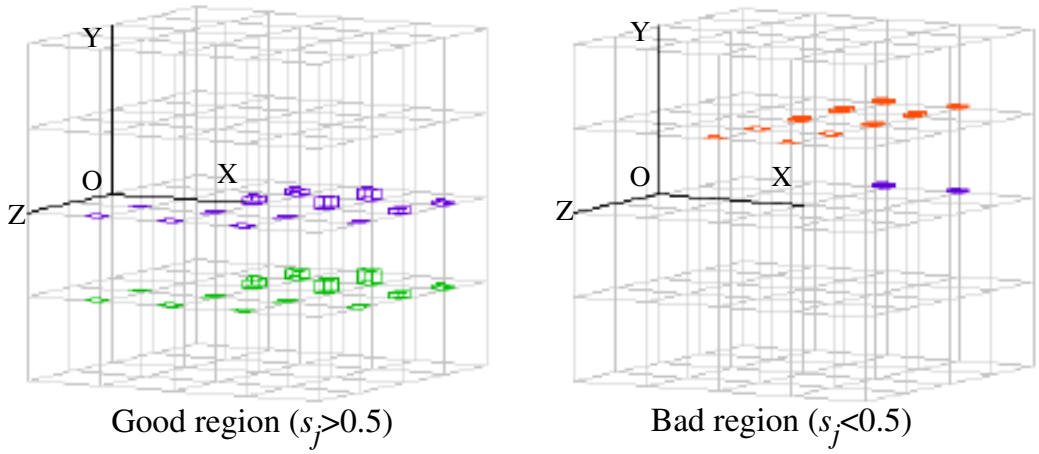


(b) Second human evaluation (8 point)

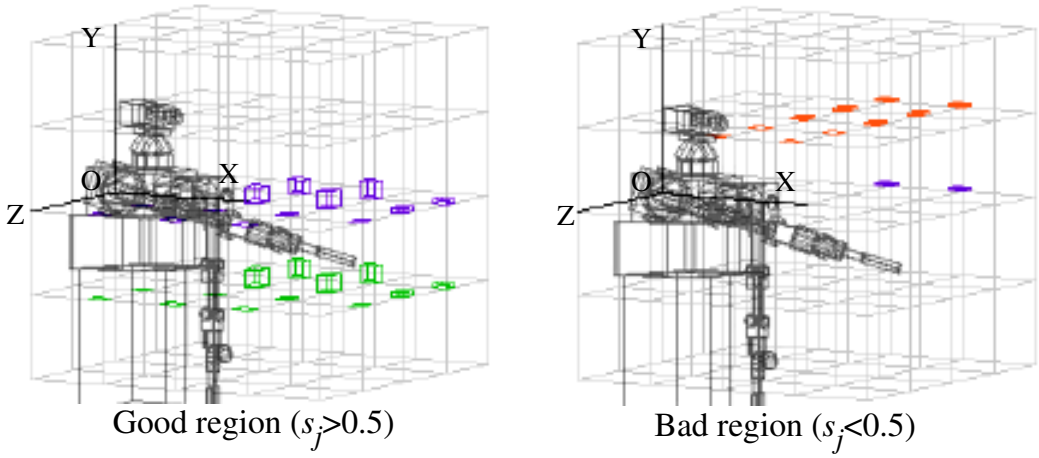
Fig 5.21 The learning result of consequent parts of the fuzzy state-value function.



(c) Third human evaluation (7 point)



(d) Fourth human evaluation (8 point)



(e) Fifth human evaluation (10 point)

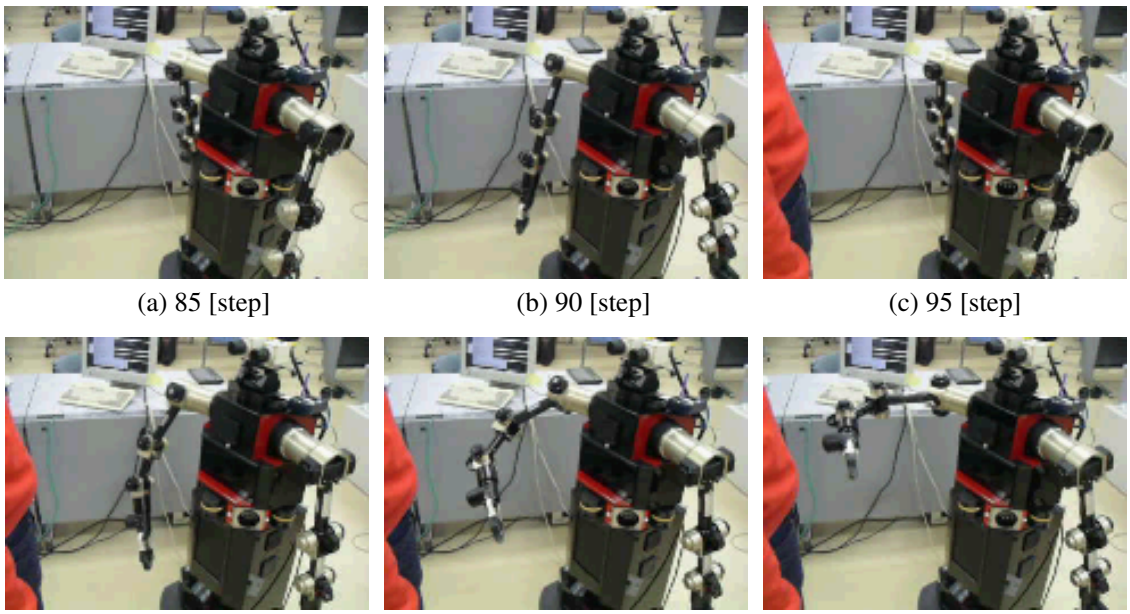
Fig 5.22 The learning result of consequent parts of the fuzzy state-value function.

5.7.2　多目的行動調停の適用

NN で学習した基本行動を多目的行動調停に導入し，二つの実験を行った．一つは，腕振り行動（walking behavior）している Hubot に，人間が近づいてくる状況を考える．このとき，近づいてくる人間に対して，手を差し出してモノを渡す（或は，握手をする）手渡し行動（hand-to-hand behavior）を行うものとする．待機位置にアームを戻す待機行動（home positioning behavior）と融合することで，滑らかな主行動の遷移による動作生成を行う．もう一つの実験は，台車を用いた人間探索と手渡し動作の基礎実験である．Hubot が動きまわり人間を見つけ手渡しを行うものとする．この実験では，探索行動における停止位置の学習も行い，人間が手渡しし易い距離を獲得する．これら二つの実験を通して，動作の滑らかさと多目的性について検証する．

腕振り—手渡し動作

図 5.23 と図 5.24 に，実験のスナップショットと行動重みの変化をそれぞれ示す．人間の感知は，胸部赤外線ラインセンサにより行い，一定距離（1.2m）以内に近づくことで手渡し行動に対する行動重みが増加するように調停則を設計している．また，人間がアーム先端を握ることで，もとの位置に戻る待機行動の行動重みが増加するようになっている．93[step] まで人間を感知せず，腕振り行動を主行動とする動作を行う．93[step] 以降，人間の接近とともに徐々に手渡し行動の行動重みが増加し，腕振り動作から手渡し動作へと滑らかに変化していく（図 5.23(c)(d)(e)）．腕振り動作から手渡し動作へと変化していく過程で，動作を停止あるいは，アームを初期位置に戻すことなく，次の動作へと遷移できることを示した．



(a) 85 [step] (b) 90 [step] (c) 95 [step]

(d) 100 [step] (e) 105 [step] (f) 110 [step]

Fig 5.23 The snapshots of hand-to-hand behavior while swinging arms.

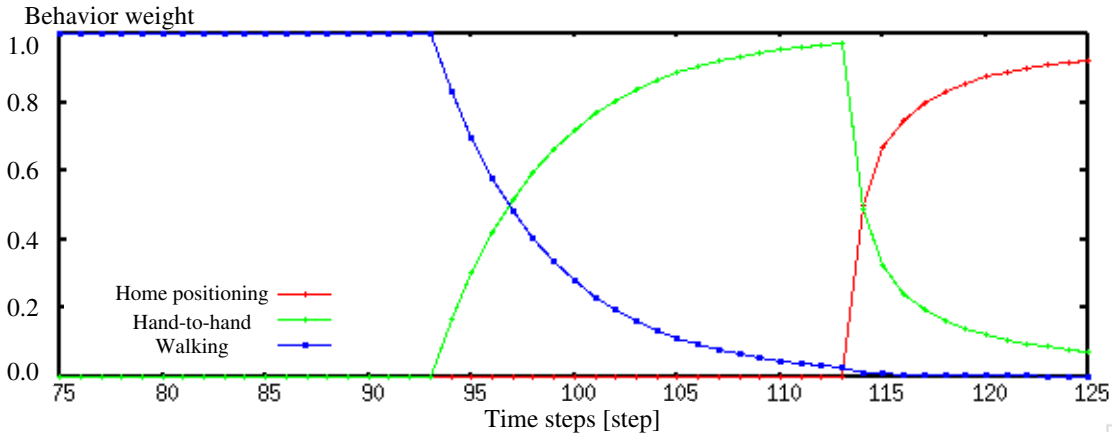


Fig 5.24 Change in behavior weights.

人間探索ー手渡し動作

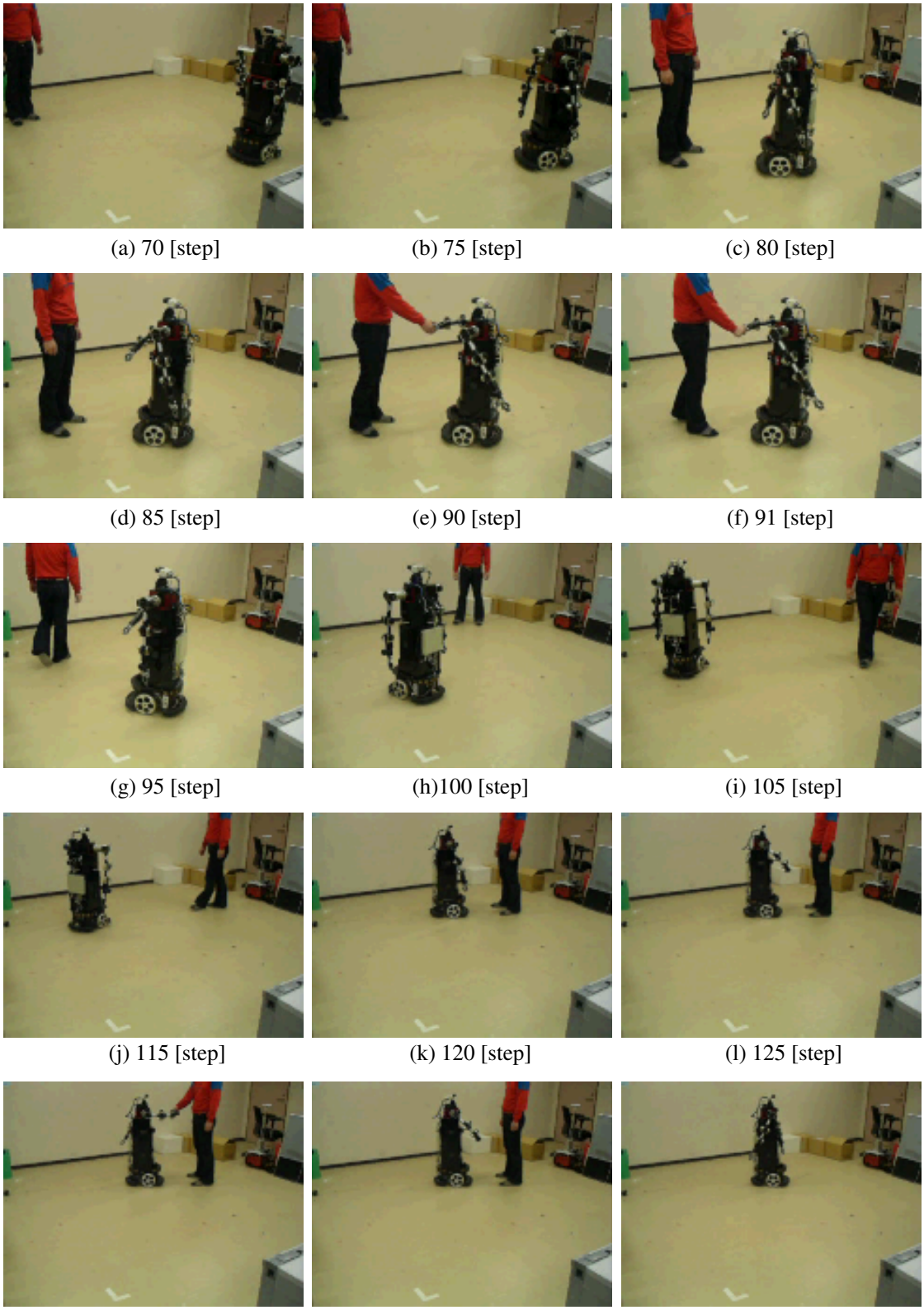
カメラ画像と台車を用いて，人間の探索と手渡し動作を行った．適用した基本行動は，手渡し行動，腕振り行動，待機行動，障害物回避行動，目標追従行動の 5 つである．台車とアームとでは制御次元が異なるため，障害物回避行動と目標追従行動の二つは，別の調停則により重み付けし動作出力を計算する．障害物回避行動は，台車の超音波センサの距離情報を入力とし，簡略化ファジィ推論により両輪の速度を決定する．目標追従行動は，カメラ画像から人間の服の色の多い方向を検出し，その色の方向に向かうように両輪の速度を決定する．行動調停則は，カメラ画像中の服の色を占める割合が一定以上になったとき，目標追従行動の行動重みを増加するように設定する．

目標追従行動は，人間に向かうだけでなく，手渡しし易い位置で止まる必要があり，停止距離を学習を行う．この手渡しし易い距離というのは，ロボットが近づきながら手を差し出すという一連の動作を，人間が実際に見ること（経験すること）によって感じるものである．また，実際に停止する位置は，他の行動との兼ね合いからも変わりうる．そこで，実際に停止した後の，手渡し前後での人間との距離を計測し，次式の簡単な学習則により停止距離を学習する．

$$d_{stop} \leftarrow d_{stop} + \gamma(d_{before} - d_{after}) \tag{5.16}$$

d_{stop}，d_{before}，d_{after}，γ は，それぞれ，両輪の速度を 0 にする距離，ロボットが停止した時の人間との距離，人間がアーム先端を触れた時の人間との距離，学習係数 $(0 < \gamma \leq 1)$ とする．これらの距離は全て胸部赤外線ラインセンサの前方 30 度（15 点の距離情報）の最短距離とする．実験では，d_{stop} の初期値を 100cm としている．学習係数は 0.5 とする．また，120cm 以内に近づくことで手渡し行動の行動重みが増加するものとする．

図 5.25 に，実験のスナップショットを示す．図 5.26 と図 5.27 に台車の行動重みの変化とアームの行動重みの変化を示す．実験結果から，腕を振りながら走り回り，人間を見つけると近づき手渡し動作をしているのがわかる．一連の動作は滑らかに接続されているのが確認できる．図 5.28 に目標追従における停止距離の履歴を示す．最初の手渡しの時，手渡し時に人間が少し前に乗り出している．これにより，手渡し前後で互いの距離に差異が生じ，停止距離を学習している（図 5.25(e)(f)）．この後，手渡し時に人間の移動がほとんど計測されず，ほぼ同じ停止距離になっている．これらの結果から，一連の動作の滑らかさと，その動作の中で，人間との手渡し距離を学習できることを示した．



(a) 70 [step] (b) 75 [step] (c) 80 [step]

(d) 85 [step] (e) 90 [step] (f) 91 [step]

(g) 95 [step] (h)100 [step] (i) 105 [step]

(j) 115 [step] (k) 120 [step] (l) 125 [step]

(m) 129 [step] (n) 134 [step] (o) 145 [step]

Fig 5.25 The snapshots of hand-to-hand behavior while searching and tracing user.

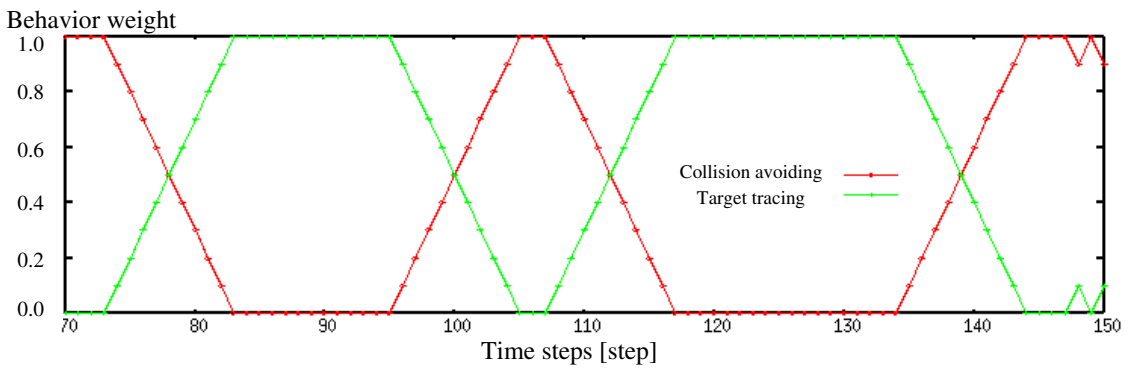


Fig 5.26 Change in behavior weights of the mobile robot.

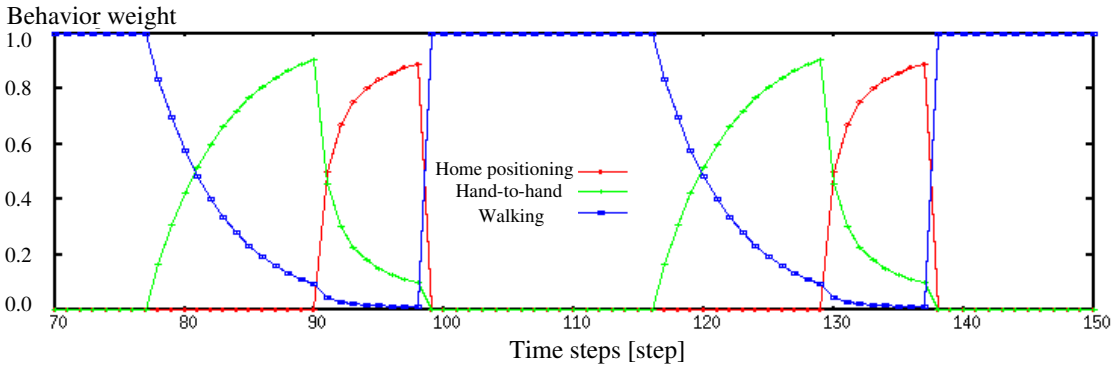


Fig 5.27 Change in behavior weights of the arms.

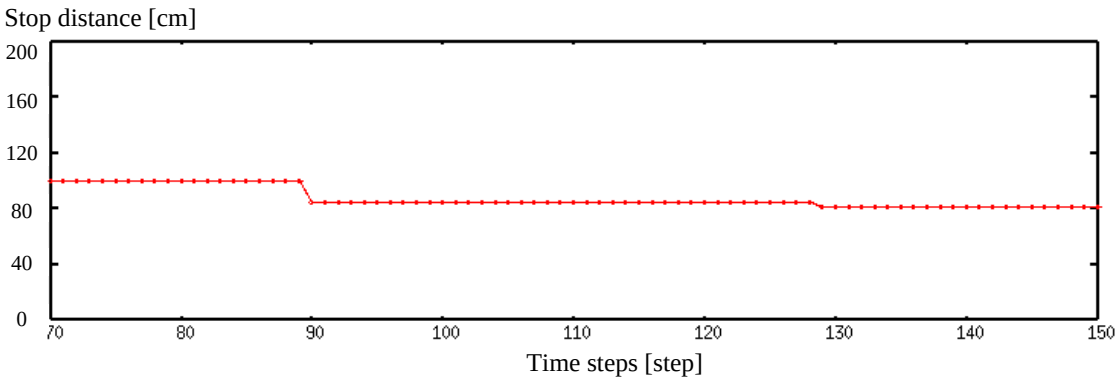


Fig 5.28 Change in the stop distance for target tracing.

5.8 結言

人間–ロボット環境における行動獲得として，アームの手渡し行動を対話型進化的計算手法を提案した．事前に設計できない使用者の評価をファジィ状態価値関数でモデル化し，ロボットが提示する軌道に対し人間が評価を行い，その評価に基づき状態価値関数を更新する．この人間の評価による状態価値関数の更新と，内部での軌道探索を繰り返すごとに，より適した手渡し軌道を探索できることを，計算機シミュレーション及び実機人型ロボットを用いた実験により示した．

また，個々の基本行動の獲得時において，評価の高い解候補を教師値として，現在の関節角に対する関節角の変化量を NN で出力できるよう学習する手法を提案した．学習した基本行動 NN は，他の行動と融合し多目的な動作を生成することができることを示した．さらに，実機人型ロボットへ適用することで，この融合された多目的な動作が滑らかに基本行動を接続できることを示した．ただし，その融合された一連の動作によってうける印象は，実際に人間が直面してみないと解らず，逐次人間との手渡し距離を学習する手法を提案し，適した動作を適した距離で実行できることを示した．

一般に，このような人型ロボットでは，基本行動を切り替えて一連の動作が作られている．また，各基本行動も事前に設計されたものである．パートナーロボットとして実際に人間の環境に入り共生するには，使用者にとって好ましい動き方や距離の取り方を学習するべきであり，本提案手法はそのための一つの方法論として用いることができると考えられる．

第6章　結論

6.1　本研究のまとめ

与えられたタスクを遂行するために，複数の基本行動を用意した行動に基づくロボットが盛んに研究されているが，ほとんどの研究では動作の滑らかさや行動の再利用性について，あまり議論していない．本研究では，様々な状況に対応するための多目的かつ滑らかな動作生成と，タスクを達成するために必要となる基本行動と行動調停則の学習を，時系列観測情報を考慮した多目的行動調停に基づくロボットを用いて行った．多目的行動調停は，どの基本行動をどのくらい用いるかをヒューリスティックに決定し，重み付け平均で動作を生成する．ここで，行動調停則はトップダウン的な役割を果たすが，実際の動作は，個々の基本行動の出力に依存する．この意味で，動作はボトムアップ的であり，さらに，この動作によって，観測情報が変化し，行動調停則の行動重みを変更する．つまり，トップダウン的な行動調停則とボトムアップ的な動作により，ロボットは構成され，さらに行動調停則と基本行動が，互いに機能を限定し合う相互依存の入れ子構造を成す．このようなの入れ子構造を持つ多目的行動調停に基づくロボットの学習に関して，環境条件に合わせた構造的な学習手法を示し，動作の滑らかさや基本行動の再利用性などの議論を行った．

第三章では，多目的行動調停に基づくロボットシステムの静的未知環境における基本行動の獲得を，観測情報の時系列を考慮した行動調停則のもとで行った．行動調停則により役割が割り当てられた基本行動の学習を通して，観測情報に対して滑らかな主行動の遷移が行われる動作が獲得された．また，様々な静的未知環境下での評価を通して，基本行動が短い世代で獲得できることを示した．さらに，未学習の未知環境下においても，学習環境と同様の動作を行うことができることを示した．一方，時系列の観測情報を考慮しない反射的な行動切り替え型のロボットが，学習環境に依存して個々の基本行動が特化され，未学習の環境で上手く動作できないことも示した．多目的行動調停を用いたロボットの行動獲得は，学習環境（一時刻の観測情報）に特化されるのではなく，観測情報に依存してトップダウン的に与えられる役割（観測情報の変化）に対して特化される．またその役割の度合い（行動重み）は，時系列観測情報により変化するため，単に切り替えるよりも役割が多様となる．そのため，未学習の

静的未知環境に対しても，環境の障害物の配置や疎密さなどが大きく変わらないような環境であれば，問題なく対応できることが明らかになった．つまり，行動調停則の設計の段階でのヒューリスティクスから，環境が大きく異ならない限りにおいて，学習環境と同様の性能が得られ，基本行動の再利用性が高いことが明らかになった．

第四章では，静的未知な環境から，移動障害物を含む動的環境へと問題を拡張し静的未知環境で学習したロボットの適応性に関して議論を行った．動的環境下では，ロボットが意思決定を行っている間に環境が変化する．そのため，静的環境下で用いてきた行動調停則のヒューリスティクスとは異なり，移動障害物に対する行動重みが，適切に計算されず衝突することがある．これに対し，新たに移動障害物に対応する回避行動を基本行動として加え，新たな行動調停則を付加することで，移動障害物に対応することもできるが，どのような時に，どのように行動重みを変化させ，回避行動を用いれば良いかは，実際に移動障害物に遭遇してみなければ解らず，適用は難しい．そこで，新たな行動や行動調停則を導入するのではなく，静的未知環境で獲得した行動調停則を，部分的に調整することで，移動障害物に対応する手法を提案した．移動障害物との遭遇による一連の動作を局所エピソードとして保持し，衝突時の緊急回避的な行動重みを局所エピソードに遡って学習することで，移動障害物との同じような遭遇に対し，衝突を回避できることを示した．この結果から，基本行動を変えることなく，トップダウン的な行動調停則を少し変えるだけで，移動障害物を含む動的環境に適応できることが明らかになった．提案手法は，少ない学習回数で移動障害物に対応することができ，動的環境下における適応学習として，基本行動の再利用の観点からも有効な手法であるといえる．

第五章では，人間と共生するパートナーロボットの構築へ向けて，基本行動の獲得と多目的行動調停の適用に関して議論を行った．使用者となる人間の評価基準は，事前に特定することができず，人間とロボットとの距離や，ロボットの動作によって時々刻々と変化しうる．最初に，実際の対話により生成される人間の評価モデルの同定と，同時に評価の高い手渡し動作を探索する対話的学習手法を提案した．提案手法により，対話を通して人間の評価モデルが迅速に学習され，少ない評価回数で手渡し動作が獲得できることがわかった．また，NN に現在の関節角度に対する次の変化量を学習させることで，手渡し動作を基本行動として獲得できることを示した．この手渡し行動と他の基本行動の出力を多目的行動調停により融合することで，状況に合わせた滑らかな主行動の遷移が可能であることを示した．さらに，融合された一連の動作において，手渡し動作のための距離を学習することで，適した動作を適した距離でロボットが実行できることを示した．パートナーロボットを構築するにあたり，実際の使用者に合わせた動き方や距離の取り方を学習する手法として有効な手法と考えられる．また，対話

により獲得した基本行動とその行動出力を簡単な調停則により融合した多目的な動作は，従来の人型ロボットにおける事前作り込みを軽減することができると考えられる．

本論文で示した基本行動（行為）と多目的行動調停則（知覚）の入れ子構造を観測情報の時系列的変化のもとで構成することで，静的未知環境や動的環境下で再利用可能な行動が獲得可能であることを示した．また，人間の評価を考慮して獲得した基本行動を，多目的動作へと再利用可能であることを示した．このように，入れ子構造による知覚と行為の構造化は，知能ロボットの適応可能性を高め，様々な環境やタスクに適用できると期待できる．

6.2 今後の課題

本論文では，トップダウン的な行動調停則とボトムアップ的な動作の両者の獲得に関して議論してきたが，基本行動か行動調停則か，どちらを，いつ，どのくらい学習させるかに関して，更なる議論が必要である．

知覚と行為の依存関係における循環的なプロセスを考えると，どちらかが固定されることによる最適化は，固定され続けられることにより，モジュールの特化が生じる可能性がある．これは局所解に他ならず，よりよい行動を獲得するためには，徐々に依存関係を改善するような学習が必要である．この問題は，人間とロボットとの共生を考えたとき，タスクや環境の変化や，互いの価値の変化により大きな問題となりうる．この問題に関して，先行研究として学習し続ける知覚と行為のデュアルラーニングを提案 [106] しており，さらに議論していく予定である．

謝辞

本研究を行うにあたり，研究の継続ならびに研究に関する多くの機会と助言を賜りました神戸大学大学院自然科学研究科機械・システム科学専攻の小島史男教授に深く感謝いたします．また，本論文をまとめるにあたり貴重な御教示を賜わりました神戸大学大学院自然科学研究科機械・システム科学専攻の田浦俊春教授ならびに情報・電子科学専攻の上原邦昭教授に深く感謝いたします．

本研究をすすめるにあたり，福井大学知能システム工学科の久保田直行助教授には，多大なる御指導ならびに公私にわたる多くの助言を賜わり大変感謝いたします．また，神戸大学大学院自然科学研究科機械・システム科学専攻の小林太助手とシステム機能科学専攻の橋本節雄氏，情報知能工学専攻の北川幸宏氏には，考えのまとまらない段階で多くの議論に付合っていただき感謝しております．このほか，小島研究室学生諸氏には，有意義な議論の場と生活の場を作っていただき感謝しております．最後に，長い学生生活を精神的にも経済的にも支えてくれた家族に深く感謝いたします．

参考文献

[1] S.J.Russell and P.Norvig, *Artificial Intelligence*, Prentice-Hall, Inc., (1995)

[2] 白井，人工知能の理論，コロナ社，(1992)

[3] M.Minsky, 心の社会，産業図書, (1990)

[4] L.A.Zadeh, The Birth and Evolution of Fuzzy Logic, 日本ファジィ学会誌 Vol.2, No.1, pp.2-11, (1990)

[5] 福田編，インテリジェントシステム－適応・学習・進化システムと計算機知能－，昭晃堂，(2000)

[6] 坂和, 馬野, 大里 編，ソフトコンピューティング用語集, 朝倉書店, (1996).

[7] 市橋, 渡辺, 簡略ファジィ推論を用いたファジィモデルによる学習型制御, 日本ファジィ学会誌, Vol.2, No.3, pp.157-165, (1990)

[8] 林, 古橋編, ファジィ・ニューラルネットワーク, 朝倉書店, (1996)

[9] 馬場，小島，小澤，ニューラルネットの基礎と応用，共立出版，(1994)

[10] 川人，脳の計算理論，産業図書，(1996)

[11] R.S.Sutton and A.G.Barto, *Reinforcement Learning*, The MIT Press, (1998)

[12] E.Mizutani, Learning from Reinforcement, Chapter 10 in *Neuro-Fuzzy and Soft Computing*, New Jersey' Prentice-Hall.Inc., pp.258-300, (1997)

[13] 木村，宮崎，小林，強化学習システムの設計指針，計測と制御，Vol.38, No.10, pp.618-623, (1999)

[14] 高玉，マルチエージェント学習–相互作用の謎に迫る–，コロナ社，(2003)

[15] J.J.Grefenstette, Credit Assignment in Rule Discovery Systems Based on Genetic Algorithms, *Machine Learning 3*, pp.225-245, (1988)

[16] 堀内, 藤野, 片井, 椹木, 連続値入出力を扱うファジィ内挿型 Q-Learning の提案, 計測自動制御学会論文集, Vol.35, No.2, pp.271-279 , (1999)

[17] D.E.Goldberg, *Genetic Algorithms - in Search, Optimization & Machine Learning* -, Addition- Wesley, (1989)

[18] 三宮，喜多，玉置，岩本，遺伝アルゴリズムと最適化，朝倉書店，(1998)

[19] T.Back, D.B.Fogel and T.Michalewicz (edit.), *Evolutionary Computation 1 - Basic Algorithms and Operators* -, IOP, (2000)

[20] H.Kitano, Continuous Generation Genetic Algorithms, 計測と制御, Vol.32, No.1, pp.31-38, (1992)

[21] K-A.DeJong and J.Sarma, Generation Gap Revisited, *Foundations of Genetic Algorithms 2* ed. L.D.Whitley, Morgan Kaufmann, pp.19-28, (1993)

[22] G.Syswerda, A Study of Reproduction in Generational and Steady-State Genetic Algorithms, *Foundation of Genetic Algorithms*, Morgan Kanfmann, (1991)

[23] 高木，畝見，寺野，対話型進化計算法の研究動向，人工知能学会誌，Vol.13, No.5, 1-13, (1998)

[24] H.Takagi, Interactive Evolutionary Computation, Fusion of the Capacities of EC Optimization and Human Evaluation, *Proceedings of the IEEE*, Vol.89, No.9, pp.1275-1296, (2001)

[25] T. Unemi, A Design of Genetic Encoding for Breeding Short Musical Pieces, *E.Bilotta, D. Gross, et al eds. ALife VIII Workshops, ALMMA II*, pp.25-29, (2002)

[26] T.Unemi, SBEAT3, http://www.intlab.soka.ac.jp:80/∼unemi/sbeat/

[27] 片上，山田，対話的進化ロボティクスの観察に基づく教示の設計，システム制御情報学会誌，Vol.16, No.6, pp.279-286，(2003)

[28] 松生，古橋，配管が接続される機器同士のまとまりを考慮した機器配置に関する一考察，第 19 回ファジィシステムシンポジウム講演論文集，pp.91-94, (2003)

[29] 中野，高木，インタラクティブ EC を用いた医療画像強調処理，第 19 回ファジィシステムシンポジウム講演論文集，pp.111-112, (2003)

[30] 上野，古橋，対話型看護師勤務表作成支援システムのインタラクティブ性に関する一考察，第19回ファジィシステムシンポジウム講演論文集，pp.125-128, (2003)

[31] 竹内，遺伝的アルゴリズムによる機械学習，計測と制御，Vol.32, No.1, pp.24-30, (1993)

[32] 玉置，遺伝的機械学習アルゴリズム，電気学会誌C編，Vol.119, No.8/9, pp.1-6, (1999)

[33] J.H.Holland, *Hidden Order*, Addison-Wesley Publishing Company, Inc., (1995)

[34] H.Ishibuchi, T.Nakashima, T.Murata, A Fuzzy Classifier System that Generates Fuzzy If-then Rules for Pattern Classification Problems, *Proceedings of 2nd IEEE International Conference on Evolutionary Computation*, pp.759-764, (1995)

[35] 塩瀬, 岡田, 椹木, 片井，双参照モデルにおける社会性の発現機構：目玉ジャクシの原初的サッカーにおける社会的秩序について, 電子情報通信学会, 信学技報, AI97-66, pp.79-86, (1998)

[36] H.Inoue, K.Takadama, K.Shimohara, and O.Katai, Acquisition of a Specialty in Multi-Agent Learning Systems - Approach from Learning Classifier System with Index -, *Proceedings of Computational Intelligence in Robotics and Automation*, pp.1090-1095, (2003)

[37] G.Dudek and M.Jenkin, *Computational Principles of Mobile Robotics*, Cambridge University Press, (2000)

[38] A.Saffiotti, *The Use of Fuzzy Logic in Autonomous Robot Navigation*, Soft Computing, Vol.1, pp.180-197, (1997)

[39] O.Khatib, Real-Time Obstacle Avoidance for Manipulators and Mobile Robots, *The International Journal of Robotics Research*, Vol.5, No.1, pp.90-98, (1986)

[40] R.Brooks, A Robust Layered Control System for a Mobile Robot, *IEEE Journal of Robotics and Automation*, Vol.RA-2, No.1, pp.14-23, (1986)

[41] R.Brooks, *Cambrian Intelligence; The Early History of the New AI*, The MIT Press, (1999)

[42] S.Nolfi and D.Floreano, *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*, The MIT Press, (2000)

[43] R.Pfeifer and C.Scheier, Sensory-motor Coordination: The Metaphor and Beyond, *Robotics and Automation System*, Vol.20, pp.157-178, (1997)

[44] M.I.Jordan and R.A.Jacobs, Hierarchical Mixtures of Experts and the EM Algorithm, *Neural Computation*, Vol.6, pp.181-214, (1994)

[45] T.Caelli, L.Guan, and W.Wen, Modularity in Neural Computing, *Proceedings of the IEEE*, Vol.87, No.9, pp.1497-1518, (1999)

[46] 藤田，平田，スイッチング制御，計測と制御，Vol.38, No.3, pp.176-188, (1999)

[47] 石黒，浅田，國吉他，特集，認知ロボティクス，日本ロボット学会誌，Vol.17, No.1, (1999)

[48] J.R.Koza, Evolution of Subsumption Using Genetic Programming, *Proceedings of the First European Conference on Artificial Life (ECAL-91)*, pp.110-119, (1992)

[49] 近藤，石黒，内川, 生体内免疫系を参考にした自律調停機構の創発的生成に関する一手法, 計測自動制御学会論文集, Vol.33, No.1, pp.1-9, (1997)

[50] 近藤，石黒，内川，P.Eggenberger，進化ロボティクスにおける制御器の頑健性の実現，計測自動制御学会論文集，Vol.35, No.11, pp.1-8, (1999)

[51] 渡辺, 泉, ロボットのためのインテリジェント制御, 第 14 回ファジィシステムシンポジウム講演論文集, pp.557-578, (1998)

[52] 古橋, 知的制御のための進化的アルゴリズム, 第 14 回ファジィシステムシンポジウム講演論文集, pp.573-576, (1998)

[53] 渡辺, 泉, あいまい行動型制御（第 1 報，制御系実現の提案）, 日本機械学会論文集 C 編，Vol.64, No.620, pp.1278-1286, (1998)

[54] 福田, 長谷川, 複数のコントローラの学習方法, 日本機械学会論文集 C 編, Vol.63, No.610, C, pp.2043-2051, (1997)

[55] F.Hoffmann, An Overview on Soft Computing in Behavior Based Robotics, *Proceedings of Joint 10th IFSA World Congress*, 2003, CD-ROM

[56] P.Pirjanian, Multiple Objective Behavior-Based Control, *Robotics and Autonomous Systems* Vol.31, Issue 1-2, pp.53-60, (2000)

[57] A.Bonarini, G.Invernizzi, T.H.Labella, and M.Matteucci, An Architecture to Coordinate Fuzzy Behaviors to Control an Autonomous Robot, *Fuzzy Sets and Systems* 134, pp.101-115, (2003)

[58] J.Zhang and A.Knoll, Integrating Deliberative and Reactive Strategies via Fuzzy Modular Control, in: A.Saffiotti, D.Driankov (Eds.), *Fuzzy Logic Techniques for Autonomous Vehicle Navigation*, Springer, Berlin, (2000), pp. 367-387 (Chapter 15)

[59] T.Balch and R.C.Arkin, Behavior-based Formation Control for Multi-Robot Teams, *IEEE Transactions on Robotics and Automation*, Vol.14, No.6, pp.926-939, (1999)

[60] J.Kosecka and R.Bajcy, Discrete Event Systems for Autonomous Mobile Agents, *Proceedings of Intelligent Robotic Systems '93 Zakopane*, pp.21-31, (1993)

[61] P.Althaus, H.I.Cristensen, and F.Hoffman, Using the Dynamical System Approach to Navigate in Realistic Real-World Environments, *Proceedings of the IEEE/RSJ International Conference on Intelligent and Systems*, pp.1023-1029, (2001)

[62] D.M.Wolpert, R.C.Miall, and M.Kawato, Internal Models in the Cerebellum, *Trends in Cognitive Sciences*, Vol.2, No.9, pp.338-347, (1998)

[63] D.M.Wolpert and M.Kawato, Multiple Paired Forward and Inverse Models for Motor Control, *Neural Networks*, Vol.1, pp.1317-1329, (1998)

[64] 小池，銅谷，マルチステップ状態予測を用いた強化学習によるドライバモデル，電子情報通信学会論文誌，D-II, Vol.J84-D-II, No.2, pp.370-379, (2001)

[65] K.Doya, K.Samejima, K.Katagiri, and M.Kawato, Multiple Model-Based Reinforcement Learning, *Neural Computation*, Vol.14, pp.1347-1369, (2002)

[66] D.Cliff, I.Harvey, and P.Husband, Explorations in Evolutionary Robotics, *Adaptive Behavior*, Vol.2, pp.73-110, (1993)

[67] 中村，石黒，内川，R.Pfeifer，環境との相互作用を用いた自律移動ロボットの識別能力の実現，日本ロボット学会誌，Vol.18, No.7, pp.963-971, (2000)

[68] A.Thompson, Notes on Design Through Artificial Evolution: Opportunities and Algorithms. *Adaptive Computing in Design and Manufacture V*, (2002)

[69] D.Floreano, J-C.Zufferey, and C.Mattiussi, Evolving Spiking Neurons from Wheels to Wings, *Proceedings of the Third International Symposium on Human and Artificial Intelligence Systems*, pp.57-62, (2002)

[70] T.Higuchi, H.Iba, and B.Manderick, Evolvable Hardware, *Massively Parallel Artificial Intelligence*, MIT Press, pp.398-421, (1994)

[71] 久保田，進化的ロボティクスと適応，システム制御情報学会論文誌，Vol.47, No.12, pp.565-570, (2003)

[72] N.Kubota, T.Morioka, F.Kojima, and T.Fukuda, Learning of Mobile Robots Using Perception-based Genetic Algorithm, *Measurement*, No.29, pp.237-248, (2001)

[73] K.Fujimura and H.Samet, A Hierarchical Strategy for Path Planning among Moving Obstacles, *IEEE Transactions on Robotics and Automation*, Vol.5, No.1, pp.61-69, (1989)

[74] O.Brock and O.Khatib, Elastic Strips: A Framework for Integrated Planning and Execution, *Proceedings of the 1999 International Symposium on Experimental Robotics*, pp.245-254, (1999)

[75] 杉山，秋下，流体力学的ポテンシャルを用いた自律移動ロボットの経路計画〜十字路における障害物回避〜，日本ロボット学会誌，Vol.16, No.16, pp.839-844, (1998)

[76] 藤澤，早川，青木，鈴木，大熊，自律移動ロボットにおける実時間行動探索，日本ロボット学会誌，Vol.17, No.4, pp.503-512, (1999)

[77] 井上，井上，大川，複数移動障害物の行動予測に基づく自律移動ロボットのオンライン回避行動計画，日本ロボット学会誌，Vol.15, No.2, pp.249-260, (1997)

[78] 新井，藤井，浅間，鈴木，嘉悦，遠藤，群ロボット環境における局所的通信に基づく衝突回避，日本ロボット学会誌，Vol.19, No.1, pp.45-58, (2001)

[79] P.Fiorini and Z.Shiller, Robot Motion Planning in Dynamic Environments, *G.Girald and G.Hirzinger, editors, International Symposium of Robotic Research*, pp.237-248, (1995)

[80] C.C.Chang and K-T.Song, Environment Prediction for a Mobile Robot in a Dynamic Environment, *IEEE Transaction on Robotics and Automation*, Vol.13, No.6, pp.862-872, (1997)

[81] K-Team SA, *Khepera User MANUAL*, Version 5.02, (1999)

[82] T.Fukuda and N.Kubota, An Intelligent Robotic System Based on a Fuzzy Approach, *Proceedings of the IEEE*, Vol.87, No.9, pp.1448-1470, (1999)

[83] N.Kubota, S.Yamaji, F.Kojima, and T.Fukuda, Behavior Learning of Human-Friendly Robots by Symbolic Teaching, *Machine Intelligence and Robotic Control, Cyber Scientific*, Vol.l.1, No.2, pp.79-86, (1999)

[84] 神田，今井，小野，石黒，人－ロボット相互作用における身体動作の数値解析，情報処理学会論文誌，Vol.41, No.6, pp.1000-1010, (2000)

[85] T.Kanda, H.Ishiguro, M.Imai, T.Ono, and K.Mase, A Constructive Approach for Developing Interactive Humanoid Robots, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002)*, pp.1265-1270, (2002)

[86] 祖山，深瀬，石井，黒岩，食事支援ロボット「マイスプーン」の操作インターフェース，第 3 回 SICE システムインテグレーション部門講演会，講演論文集，pp.433-434, (2002)

[87] O.Khatib, K.Yokoi, O.Brock, K-S.Chang, and A.Casal, Robots in Human Environments, *In Archives of Control Sciences*, Special Issue on Granular Computing, Vol.11 (XLVII), 2001, No. 3-4, pp.123-128 (2002)

[88] N.Kubota, D.Hisajima, F.Kojima, and T.Fukuda, Fuzzy and Neural Computing for Communication of a Partner Robot, *Journal of Multi-Valued Logic and Soft Computing*, Vol.9, pp.221-239 (2003)

[89] 久保田，渡邊，小島，人間共存型ロボットの軌道生成のための対話型遺伝的アルゴリズム，日本機械学会論文集 C 編，66 巻, 647 号, pp.2274-2279 (2000)

[90] N.Kubota, A.S.Indra, and F.Kojima, Interactive Genetic Algorithm for Trajectory Generation of a Robot Manipulator, *Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution and Learning*, pp.146-150 (2002)

[91] N.Kubota, T.Arakawa, and T.Fukuda, Trajectory Planning and Learning of a Redundant Manipulator with Structured Intelligence, *Journal of the Brazilian Computer Society*, Vol.4, No.3, pp.14-26, (1998)

[92] Y.Davidor, A Genetic Algorithm Applied to Robot Trajectory Generation. *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, pp.144-165 (1991)

[93] R.A. サッチマン，佐伯監訳，プランと状況的行為：人間–機械コミュニケーションの可能性，産業出版，(1999)

[94] J.J.Gibson 著，古崎ら共訳，生態学的視覚論–ヒトの知覚世界を探る–，サイエンス社，(1985)

[95] D.C.Dennett, 土屋訳，心はどこにあるのか，草思社，(1997)

論文リスト

＜著　書＞

[96] Y.Nojima, F.Kojima, and N.Kubota, Local Episode-based Learning of a Mobile Robot in a Dynamic Environment, *Dynamic Systems Approach for Embodiment and Sociality –From Ecological Psychology to Robotics–*, International Series on Advanced Intelligence Vol.6, pp.318-322, (2003)

＜投稿論文＞

[97] 能島，小島，久保田，動的環境における多目的行動調停に基づく進化ロボットのための局所エピソード学習，神戸大学自然科学研究科紀要第 22 号，(2003)（掲載予定）

[98] 能島，小島，久保田，多目的行動調停に基づく移動ロボットの行動獲得，日本機械学会論文集 C 編，68 巻 671 号，pp.2067-2073, (2002)

[99] N.Kubota, Y.Nojima, F.Kojima, T.Fukuda and S.Shibata, Path Planning and Control for a Flexible Transfer System, *Journal of Robotics and Mechatronics*, Vol.12, No.2, pp.103-109, (2000)

[100] 久保田，能島，小島，福田，柴田，学習機構を持つ自在搬送システムの最適化，日本機械学会論文集 C 編，66 巻 652 号，pp.3970-3976, (2000).

＜国際会議＞

[101] N.Kubota, Y.Nojima, Indra Adji S., and F.Kojima, Interactive Trajectory Generation using Evolutionary Programming for a Partner Robot, *Proceeding of the 12th IEEE Workshop Robot and Human Interactive Communication RO-MAN 2003*, pp.335-340, (2003)

[102] Y.Nojima, F.Kojima, and N.Kubota, Trajectory Generation for Human-Friendly Behavior of Partner Robot using Fuzzy Evaluating Interactive Genetic Algorithm, *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA2003)*, pp.306-311 in CD-ROM, (2003)

[103] Y.Nojima, F.Kojima, and N.Kubota, Local Episode-based Learning of Multi-Objective Behavior Coordination for a Mobile Robot in Dynamic Environ-

ments, *Proceedings of the IEEE International Conference on Fuzzy Systems (FuzzIEEE2003)*, pp.307-312 in CD-ROM, (2003)

[104] Y.Nojima, F.Kojima, and N.Kubota, Local Episode-based Learning of a Mobile Robot in a Dynamic Environment, *Proceedings of the Third International Symposium on Human and Artificial Intelligence Systems*, pp.384-388, (2002)

[105] Y.Nojima, F.Kojima, and N.Kubota, Perception and Behavior of Pet Robots based on Emotional Model, *Proceedings of Knowledge Based Intelligent Information Engineering System & Allied Technologies (KES2001)*, pp.859-863, (2001)

[106] N.Kubota, Y.Nojima, F.Kojima, and T.Fukuda, Dual Learning for Perception and Behavior of Mobile Robots, *Proceedings of Joint 9th IFSA World Congress and 20th NAFIPS International Conference (IFSA/NAFIPS2001)*, pp.1401-1406, (2001)

[107] N.Kubota, Y.Nojima, F.Kojima, and T.Fukuda, Multi-Objective Behavior Coordinate for a Mobile Robot with Fuzzy Neural Networks, *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN2000)*, CD-ROM Proc., (2000)

[108] N.Kubota, Y.Nojima, N.Baba, F.Kojima, and T.Fukuda, Evolving Pet Robot with Emotional Model, *Proceedings of Congress on Evolutionary Computation 2000 (CEC2000)*, CD-ROM Proc. pp.1231-1237, (2000)

[109] Y.Nojima, N.Kubota, F.Kojima, and T.Fukuda, Control of Behavior Dimension for Mobile Robots, *Proceedings of the Forth Asian Fuzzy Systems Symposium (AFFS2000)*, pp.652-657, (2000)

[110] N.Kubota, Y.Nojima, F.Kojima, T.Fukuda, and S.Shibata, Intelligent Control of Self-organizing Manufacturing System with Local Learning Mechanism, *Proceedings of the 25th Annual Conference of the IEEE Industrial Electronics Society (IECON'99)*, CD-ROM Proc., (1999)

＜国内会議＞

[111] 能島，小島，久保田，パートナーロボットのための対話型動作計画と多目的行動調停，第 19 回ファジィシステムシンポジウム講演論文集，pp.95-98, (2003)

[112] 能島，小島，久保田，多目的行動調停に基づく移動ロボットの局所エピソード記憶による動的環境適応，第 46 回システム制御情報学会研究発表講演会講演論文集，pp.467-470，(2003)

[113] 北川，能島，小島，久保田，多目的行動調停と環境予測に基づく移動ロボットの動的環境下における行動制御，第 46 回システム制御情報学会研究発表講演会講演論文集，pp.547-548, (2002)

[114] 能島，小島，久保田，多目的行動調停則の獲得に基づく移動ロボットの状況と行動，第 11 回日本ファジィ学会北信越支部シンポジウム，CDROM，(2002)

[115] 能島，予期による知覚と行動多様性，*Proceedings of Be Ambitious Conference 2001 (BAC2001)*, CDROM, (2001)

[116] 能島，小島，久保田，知能ロボットのための情動モデルによる割り込み機能，第 11 回インテリジェント・システム・シンポジウム講演論文集，pp.189-192, (2001)

[117] 能島，小島，久保田，情動モデルに基づくペットロボットの制御と学習，第 45 回システム制御情報学会研究発表講演会講演論文集，pp.257-258, (2001)

[118] 能島，久保田，小島，移動ロボットにおける直接知覚と行動，日本ファジィ学会ファジィ・コンピューティング研究部会ワークショップ「第 11 回言いたい放題の合宿研究会」，(2000)

[119] 能島，久保田，小島，構造化知能における知覚系と行動系の相互学習機構，第 44 回システム制御情報学会研究発表講演会，pp.183-184，(2000)

[120] 能島，知能ロボットに基づくデュアルラーニング，*Proceedings of Be Ambitious Conference 2000 (BAC 2000)*, pp.75-76, (2000)

[121] 能島, 久保田, 小島, 福田, 構造化知能を持つ移動ロボットの行動獲得, 第 9 回インテリジェントシステムシンポジウム, pp.780-783, (1999)

[122] 久保田，能島，小島，福田，構造化知能を持つロボットシステムの行動制御の次元について，ポトラックシンポジウム'99，pp.291-292，(1999)

[123] 久保田，能島，小島，福田，構造化知能を持つ移動ロボットのルール抽出，第5回創発システムシンポジウム，(1999)

[124] 久保田，能島，小島，福田，ファジィ生産スケジューリングのための遺伝的アルゴリズム，ロボティクス・メカトロニクス講演会'99，CD-ROM Proc.，(1999)

[125] 能島，久保田，小島，福田，柴田，稲田，山本，自在搬送システムの経路計画および制御，ロボティクス・メカトロニクス講演会'99，CD-ROM Proc.，(1999)

[126] 久保田，小島，能島，福田，遺伝的アルゴリズムと局所意思決定を用いた知的搬送システム，第8回インテリジェント・システム・シンポジウム，pp.433-436，(1998)

[127] 久保田，小島，能島，福田，遺伝的アルゴリズムによる知的搬送システムの最適化，第7回日本ファジィ学会北信越支部ファジィシンポジウム，pp.25-28，(1998)