

# 研究内容

## 行動の学習と獲得のための情報処理 (浅田グループ)

本研究グループでは，ロボットのタスク遂行のための学習を目指す，タスクに特化したアプローチではなく，脳型の情報処理として意味がある課題にアタックし，それにたいする手法を提案し，実験を通して検証することを目的としている．本節では，以下の三つの研究課題について実施内容及び成果を述べる．

1. 脚式移動ロボットの行動決定のための情報量規範に基づく観測戦略学習
2. 複数の学習器の階層的構築による行動獲得
3. 視覚情報に基づくヒューマノイドのリズム歩行パラメータの強化学習

### (1) 実施の内容

1. 脚式移動ロボットの行動決定のための情報量規範に基づく観測戦略学習 移動ロボットには視覚センサが搭載されることが多いが，視覚センサにより得られる情報は膨大であり，効率よく行動決定に必要な情報を抽出する必要がある．注視によりシーンの理解など視覚の問題が解きやすくなることは多く報告されており，タスクに応じた選択的注視法が求められている．

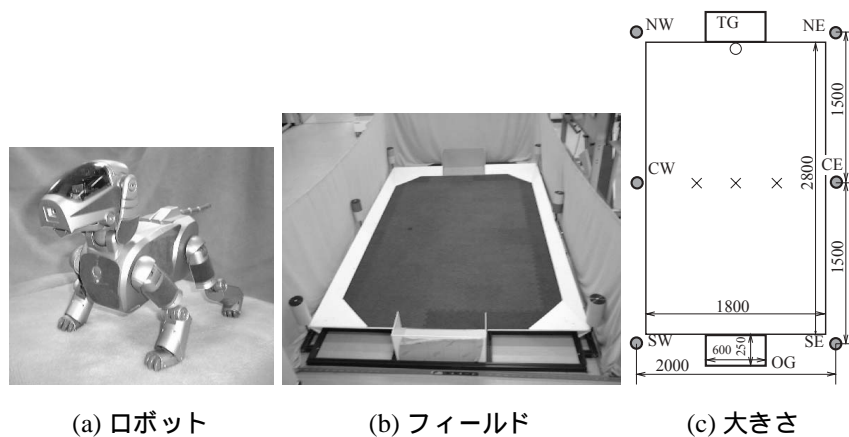


図 1: ロボットと実験フィールド (フィールドは 1999 年のロボカップ 4 足ロボットリーグのもの，×印はロボットの行動開始位置を，○印はボール位置を示す)

本研究の目的は，視覚センサを持つ脚式移動ロボットの行動決定のための効率のよい観測戦略を提案することである．観測戦略の効率は行動決定までの観測時間の長さであるとし，明示的な自己位置同定は行わない．観測の行動決定への寄与を情報量により比較し，寄与の大きい観測から実行することにより，静止観測時の効率的な観測戦略を実現する．さらに行動決定の効率の向上を図るため，移動中の観測を行動決定に利用できるような手法を拡張した．まず，観測対象(ランドマーク)選択

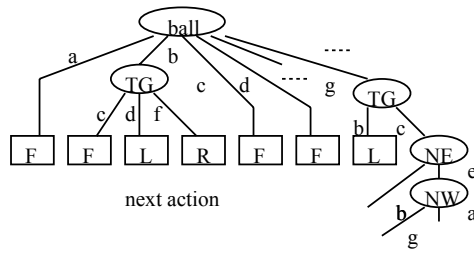


図 2: 生成された行動決定木の一部 . F, L, R はそれぞれ前進 , 左前進 , 右前進 .

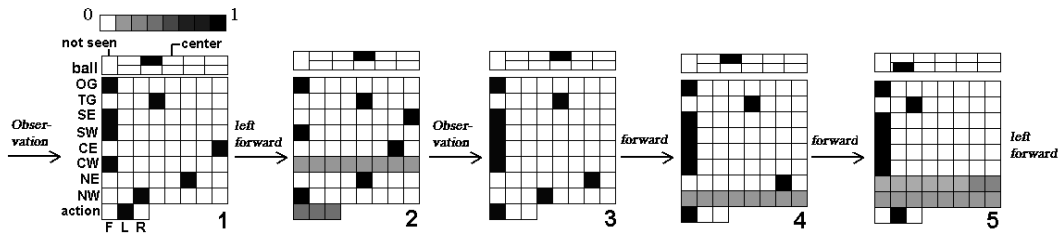


図 3: 観測行動確率の遷移例

の基準として情報量を用いることにより、より早く行動決定の曖昧さを減少することができることを示した [1] . トレーニングデータを圧縮して記憶するために、情報量を基準とした分類木の手法を用いて行動決定木を生成する . トレーニングデータは、ランドマーク群の観測される方角と、そのときにとるべき行動で与える . これに基づき各行動の生起確率より、シャノンの情報量を計算し、情報量の大きいランドマークから順に観測する行動決定木を生成する .

図 1 に示すロボット、フィールドを用いて、ゴール前(図の 印)にボールをおき、フィールドの中央 3 点(フィールド中央の 3 つの x 印)から開始してボールをゴールに入れるタスクに本手法を適用した実験結果を示す . このデータから生成した決定木の一部を Fig.2 に示す . ボールが左前方 (b の方向) に観測された場合には、回り込みのため敵ゴールの方角 (c,d,f) に応じてとるべき行動が変化することが分かる . 図 3 に教示再生を行った場合の予測と実際にとった行動例を示す . ロボットはフィールドの右側から行動を開始し、1) 左回転 2) ランドマーク確認 3) 前進 4) 前進 5) 左回転という行動をとった . 時刻 2 では行動確率が曖昧なため行動決定ができず、時刻 4, 5 では観測確率が曖昧なランドマークがあっても行動決定が行われている . すなわち、自己位置同定することなく、必要に応じた観測が実現されたことが分かる .

次に、センサ値として離散値のみでなく連続値を扱うことができ、観測方向の決定に情報量を用いるよう手法の拡張した [2] . センサ値は連続した値であることが多いが、離散値とするために予めセンサ空間を分割していた . これには、離散値として扱うことにより数値の連続性 (大小関係) が損なわれる、必要なセンサの解像度を予め知る必要がある、といった問題がある . そこで自律的なセンサ空間の分割が重要となる . また、視野角の限られた視覚センサの場合では、視野角外の情報は当然

表 1: 行動決定木生成に用いたデータ数と平均観測方向数, 平均観測時間の比較

|       | データ数 | 平均観測方向数 | 平均観測時間 [s] |
|-------|------|---------|------------|
| 予め離散化 | 34   | 3.1     | 3.3        |
| 情報量   | 43   | 3.5     | 3.3        |
| 情報量速度 | 35   | 1.3     | 0.85       |

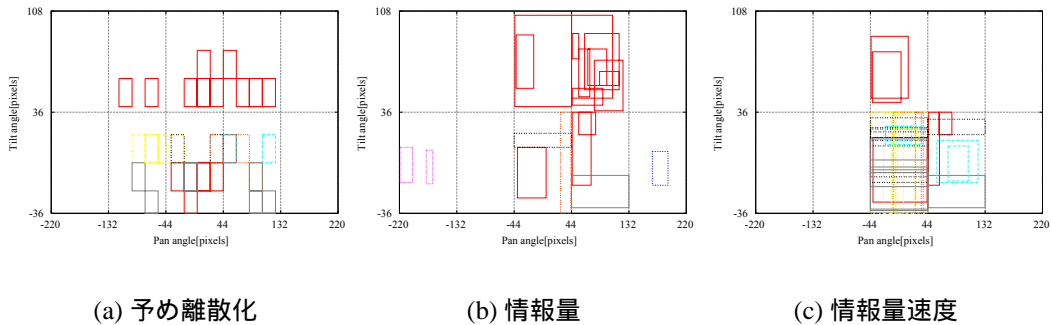


図 4: 生成された注視窓

ながら得ることができない。

離散値を扱う決定木生成法としては, C4.5 が知られているが, C4.5 で採用されている閾値による分割方法を用いると観測方向を一意に決定できない。それに対し, ランドマークが注視窓内に観測されるか否かを知るためには, その注視窓を含む方向の一度の観測で済む。そこで, 注視窓内に特定のランドマークが観測されるか否かを知ることによる情報量を比較し, 観測方向 (注視窓を含む方向) と観測対象を決定することにより, 自律的センサ空間の分割と行動決定木の生成を同時に手法を提案した。この行動決定木により, 観測対象と観測方向が決定される。また観測にかかる時間は, カメラの回転を伴うため観測方向によって異なる。そこで情報量を観測時間で割った情報量速度の利用も提案した。

ロボットによるサッカーを想定したナビゲーションタスクの実験を行った。表 1 に, 生成された木を使って実機で行動を決定した場合の, 平均観測方向数と平均観測時間を示す。予めセンサ値を離散化して注視窓とし情報量を用いて決定木を生成した場合, 情報量による離散化により注視窓と決定木を生成した場合, 情報量速度による離散化により注視窓と決定木を生成した場合, を比較すると, 明らかに平均観測時間が, 情報量速度の利用により減少していることが分かる。図 4 に, それぞれに生成された注視窓を示す。注視窓を予め用意する場合に対し, 必要に応じた大きさの注視窓が生成されることが分かる。また情報量速度を用いることにより観測窓が集中し, 観測時間を減少させていることが分かる。

次に, 歩行中の観測を利用するための拡張を行った [3]。より効率のよい観測を実現するためには, 静止状態での観測だけではなく, 移動中の観測も利用できる必要

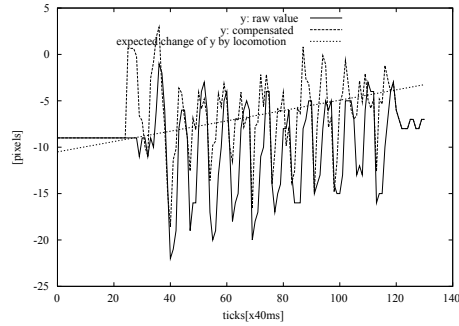


図 5: 提案手法による  $y$  軸方向の補正結果

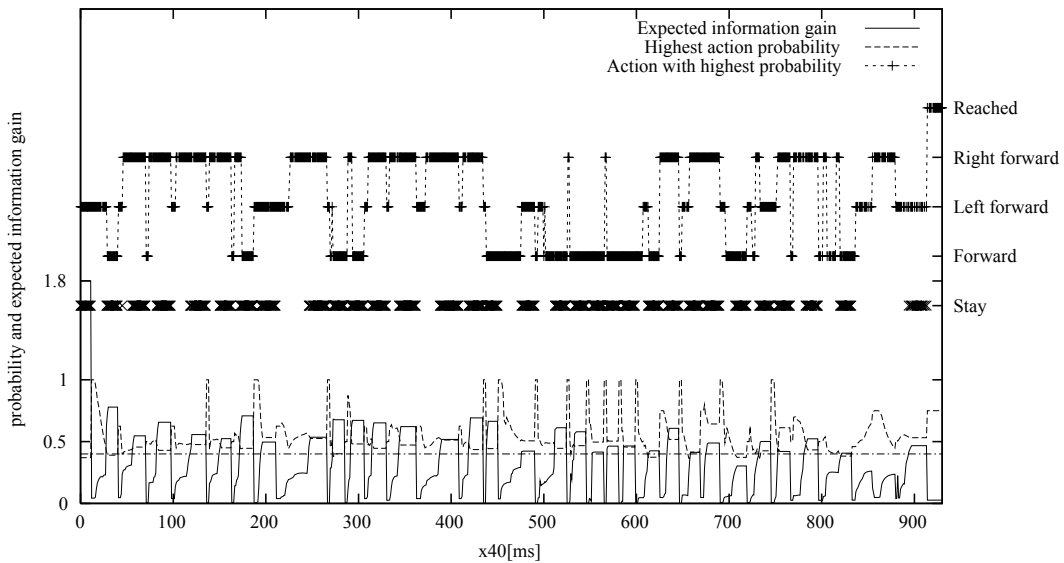


図 6: 1 試行中の、情報量期待値、最も高い行動確率、その行動と、実際にロボットがとった行動の変化

がある。

移動中の観測の利用と行動決定に関しては、

- 1) 静止時と比較してセンサに揺れなどの外乱が加わる、
- 2) 視覚情報の処理に比較的長い時間がかかり、処理結果が利用可能になったときには無視できないほど移動している場合や、複数方向の観測(首振り)の統合の際に観測間で移動が無視できない場合がある、
- 3) 観測の曖昧さから行動決定に曖昧さが残っている場合には、衝突などの問題を避けるため、移動速度を低下あるいは、静止観測の必要がある、

等の問題がある。情報量を基に生成された行動決定木を用いた行動決定においては、1) 歩行による揺れの補正、2) 移動中の複数観測の統合、3) 適切な行動決定の指標、の解決が必要となる。

そこで、行動決定の曖昧さの指標として、行動確率と共に、観測により将来得られると期待される情報量を曖昧さの指標として用い、行動確率あるいは情報量期待値が閾値を満足しない場合には、移動せずに観測する戦略を用いて上記3)の問題に対処する。また、歩行による揺れや、視点移動の影響を軽減する画像補正メカニズムを提案し、上記1), 2)の問題に対処する。画像列のみから補正値を計算し、それにより視覚センサによる観測から、静止時にある観測が得られる確率を計算し、複数観測の統合と揺れの補正を実現した。図5に歩行による画像の揺れとその補正結果を示す。図6にナビゲーションタスクにおける提案手法による観測行動決定の様子を示す。

2. 複数の学習器の階層的構築による行動獲得 強化学習や遺伝的アルゴリズム等の手法を用いたロボットによる自律的な行動の獲得に関する研究がこれまで多くなされてきた。これらの手法は先見的な知識をほとんど必要とせず、適応的で即応的な行動を獲得できる利点がよく知られている。ロボットにこれらの手法を適用することで、生涯にわたり環境との相互作用を通して自分自身の行為を発達させていくことができるようになった。しかしながら、従来の研究の多くは設計者が目標状態、状態行動空間、評価関数等を定義し、合理的な学習時間内に目的の行為をロボットに獲得させることを目的としてきた。したがってロボットはある一つの与えられたタスク以外のことを学習することは非常に困難である。

実環境の中で様々な行為を自律的に学習していくロボットを実現するためには、状況に応じて必要な状態行動変数を組み替え、探索範囲を無駄に広げることなく、比較的小さな状態行動空間をもとに学習/制御することが必要である。なぜならば、ロボットが搭載しているセンサ情報およびモータコマンドすべてを考慮した状態行動空間を作り、これをもとに与えられたタスクに対する政策を獲得することは、計算資源の面から事実上不可能であり、また学習に要する時間は非現実的になりやすいからである。

これらの問題を回避するための一手法として、全体としての学習/制御システムに階層構造を導入することが考えられる。つまり、

1. 全状態行動空間の中の部分空間を扱う学習/制御器を複数用意し、
2. 獲得された学習/制御器に対応する状況および行為を抽象化し、
3. 抽象化された状況・行為をもとに複数の部分空間にわたる新たな状態行動空間を作り、拡大された空間の中で新しい行為を獲得する

ことである。最終的には上位の層で多数の状態変数からなる状態空間を扱うことになるが、それらはすでに下位の層で状態・行動ともに抽象化されているので、従来の手法のように状態数が爆発するようなことは抑えられると考えられる。

本研究では、同一構造の学習器を複数用いて階層的に構築することによる行動獲得法を提案し、ロボカップに出場しているロボットに適用した結果を示す。下位の層の学習器はそれぞれ異なったサブゴールを担当し、低レベルな行為を学習する。同時に上位の学習器は、下位の学習器を利用し、より高いレベルの行為を学習する。

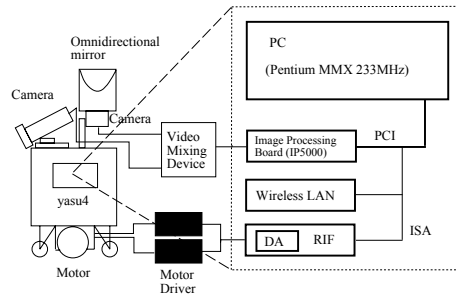


図 7: 実験に利用した移動ロボット, ボールとゴール (左図), およびロボットのシステム構成 (右図)

従来の手法とは異なり, それぞれの学習器が担当するサブゴール, タスクは自律的に決定され, また階層も自律的に構成される. また, 階層型学習機構において異なる部分空間を状態空間として持つ下位の層の学習器を統合し, より多くの状態変数によって表現される状況における行為を上位の層で獲得することが可能であることを示す.

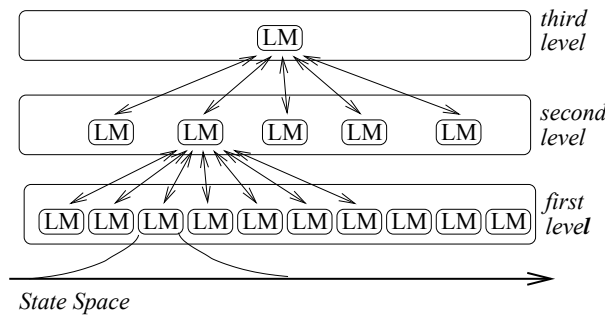


図 8: 階層型学習機構の概要図: LM は学習器

図 7 に使用する提案する手法を検証するためのロボットとボール, ゴール, またロボットの簡単なシステムを示す. ロボットはセンサとして広角レンズを装着した CCD カメラと全方位ミラーを装着したカメラを持ち, 専用の画像処理ボードを使って実時間でボールやゴールの重心を抽出する. カメラの搭載位置のため, 広角レンズを装着したカメラはロボット前方を, 全方位ミラーを装着したカメラはロボットの側方と後方を観測することになる. また移動機構は左右独立駆動機構である. これらのカメラやシステムのキネマティックなパラメータ等はロボットには未知であり, 環境との相互作用の中でセンサ情報からモータコマンドへのマッピングを提案する手法を用いて獲得することになる. 環境はボール一つ, 敵および味方ゴール, そして移動ロボットで構成される.

このロボットに階層型学習機構を導入する. 同種の行動学習器を複数並べて層を作り, これを階層的に構築する. この学習器の階層構造はタスク分解の役割を担っ

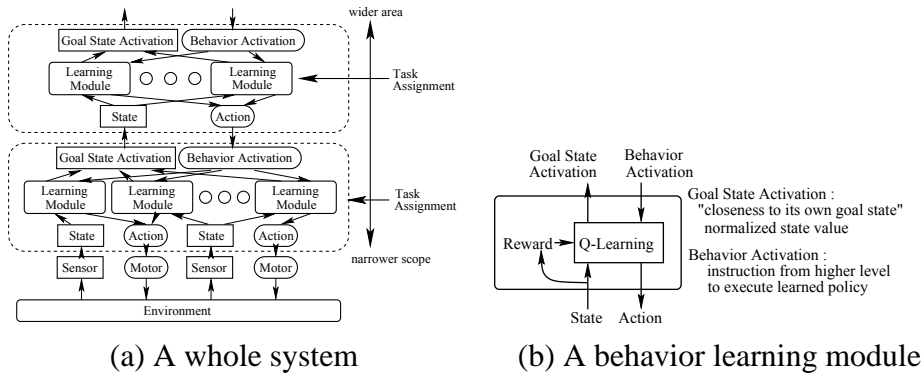


図 9: 階層型学習機構

ていると見なすことができる(図 8)。下位の学習器は、与えられた環境下で狭い範囲を探索し、より低レベルで基本的な行動を獲得する。一方、上位の学習器は下位の学習器を利用することにより、より広い範囲を探索し、より高レベルで抽象化された行動を獲得する(詳細は文献 [4] を参照)。

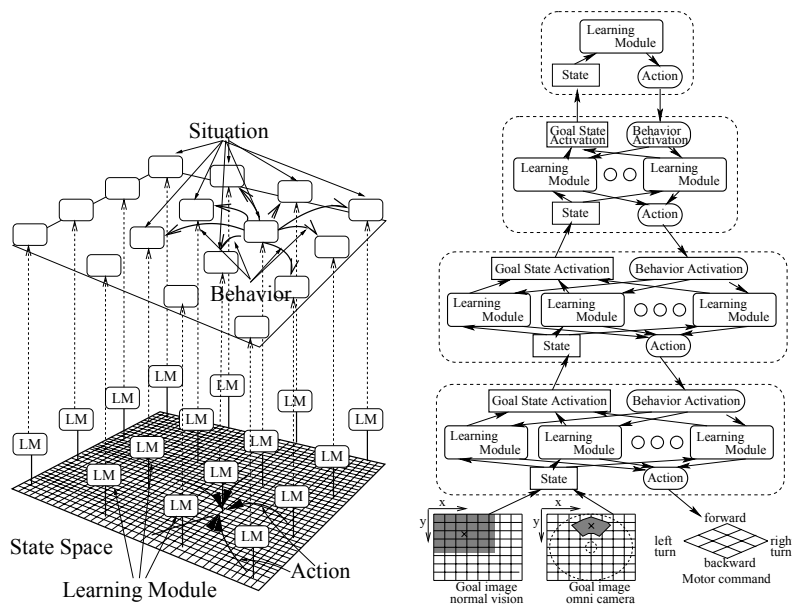


図 10: 獲得された学習器に基づく状況と行為の定義(左図)と今回構築された階層型学習機構

図 9 に提案する階層型学習機構を示す。図 9(a) は階層構造を示しており、図 9(b) は行動学習器を示している。それぞれの学習器は状態と行動を認識し、それぞれのゴール状態により報酬を発生させ、これをもとに  $Q$  学習を用いて行動を学習する。最下位の層の学習器の状態、行動はロボットの持つセンサ情報やモータコマンドを使って構成する。

上位層では獲得された学習器に基づいて下位の層の状態・行動空間から状況・行

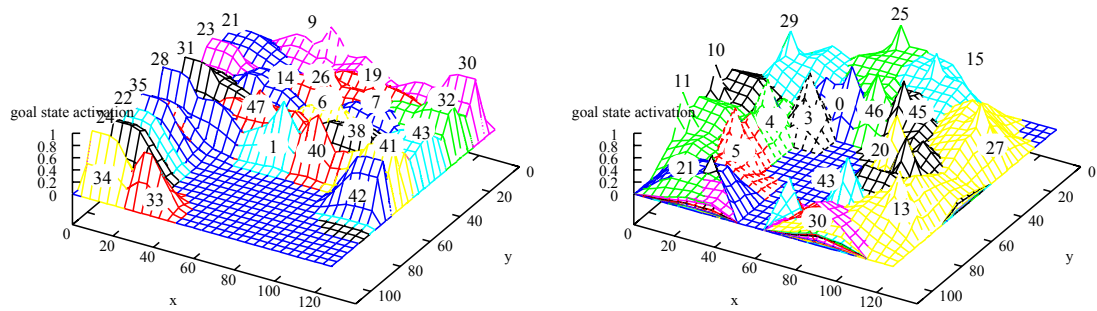


図 11: 前方位視覚(左)と全方位視覚(右)上のゴールの位置を状態空間として持った学習層における学習器の自律的分配の様子

為を定義し、これを上位の層の状態・行動空間として採用し、より抽象化された行為を学習していく。図 10 左に下位の層の学習結果を用いて状況・行為を把握する様子を示す。ある層において状態空間上に学習器を振り分け、それぞれの学習器はその状態空間上で自分に割り当てられた状態へ遷移するための行動を学習する。その上の層では、下位のそれぞれの学習器が担当している領域を状況と判断し、隣接する状況への移動を行為と認識する。

本手法を適用した様子を図 10 右に示す。最下位層の学習器の状態空間は二つのカメラから得られた画像上のゴールの座標で構成され、行動空間は左右の車輪のモータコマンドに与える指令値で構成される。前方を担当するカメラに物体が写っているときは、全方位ミラーつきカメラからの情報を無視し、前方担当のカメラからのセンサ情報を優先した。それより上の層の状態や行動は、自分より一つ下の層の学習器のゴール状態活性度と行動活性度によって構成される。それは自律的に割り当てられる学習器によって最終的に決定される。

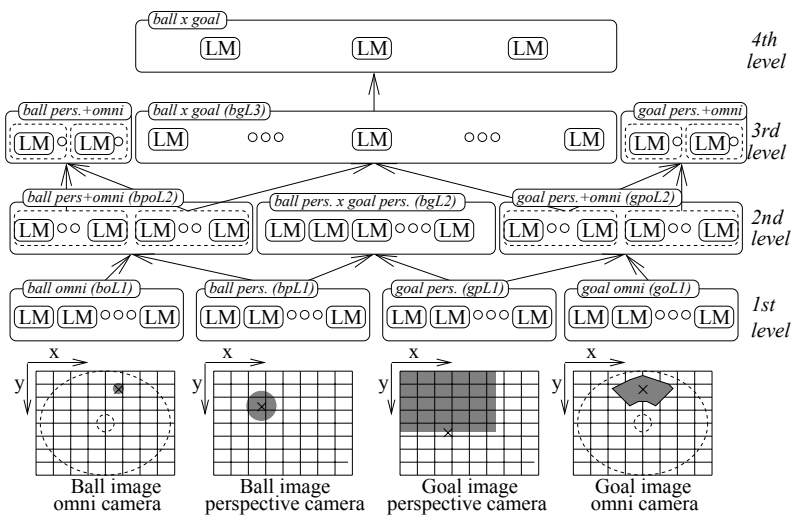


図 12: 上位層で部分空間を統合する階層型学習機構



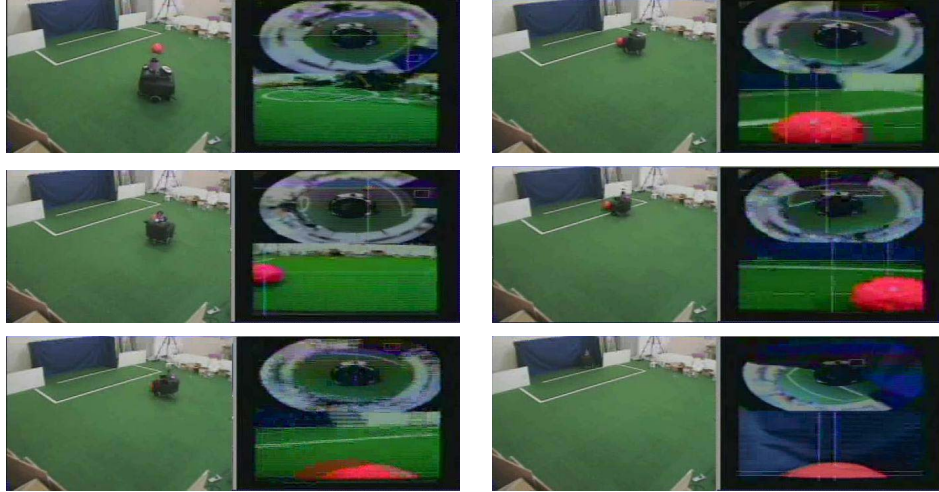


図 13: シュート行動をとらせた時の様子：右は両カメラからの画像

自律的に学習器を状態空間に配置するために「ある層の全状態空間をその層に割り当てられた学習器で一様に振り分ける」ようにアルゴリズムを組み、実装した。このとき状態間の距離をあらかじめ知ることができないため、学習器のゴール状態活性度をお互いの学習器間の距離関数として用いた。ゴール状態に到達したときのみ正の報酬が与えられ、それ以外はゼロの報酬が割り当てられる場合、学習の結果得られたある状態の  $Q$  値の最大値 (状態価値) はその状態からゴールまでの距離の近さを表現している」と考えられるからである。図 11 にそれぞれゴールが前方担当カメラ、全方位ミラー付きカメラの画像上に写ったときの最下位層のそれぞれの学習器のゴール状態活性度の分布を示す。それぞれ  $x, y$  座標は画像上のゴールの中心座標に対応している。

同一レベルの複数の層において獲得された状況・行為を基に上位の層の状態・行動空間を構成する場合、下位の複数の層の状態空間が互いに独立であるか強い相関関係があるかによって適切な構成方法を選ぶべきである。下位のそれぞれの層が異なる対象を認識し、それぞれに対して行為を学習しているならば、基本的にそれぞれの層の空間は独立であり、一方で同じ対象を認識し、行為を学習しているなら強い相関関係があると考えられる。例えば片方の層はナビゲーション行動を獲得しており、もう一方は物体の操作動作を獲得している場合は、それぞれの層の学習器は独立していると考えられる。また、同じ対象を異なるセンサによって認識している場合は、お互いの層は基本的に強い相関関係を持っているであろう。この場合、上位の層の学習器は、状況が下位のある一つの層の学習器の担当範囲外で使えないとき、下位の他の層の学習器を利用することで状況を広く認識すると期待される。そこで、ここでは前者に対し「直積的手法」を、後者に対し「相補的手法」を提案し (詳細は文献 [5] を参照)、これを実装した (図 12)。最下位では四つの学習層があり、それぞれがそれぞれの論理センサ (前カメラ画像上および全方位画像上でのボールの位置およびゴールの位置) を担当した。第 2 レベルの “ball pers.  $\times$  goal pers.” 層と第 3 レベルの “ball  $\times$  goal” 層は直積的手法で構成し、それ以外の第 2, 第 3 レベルの

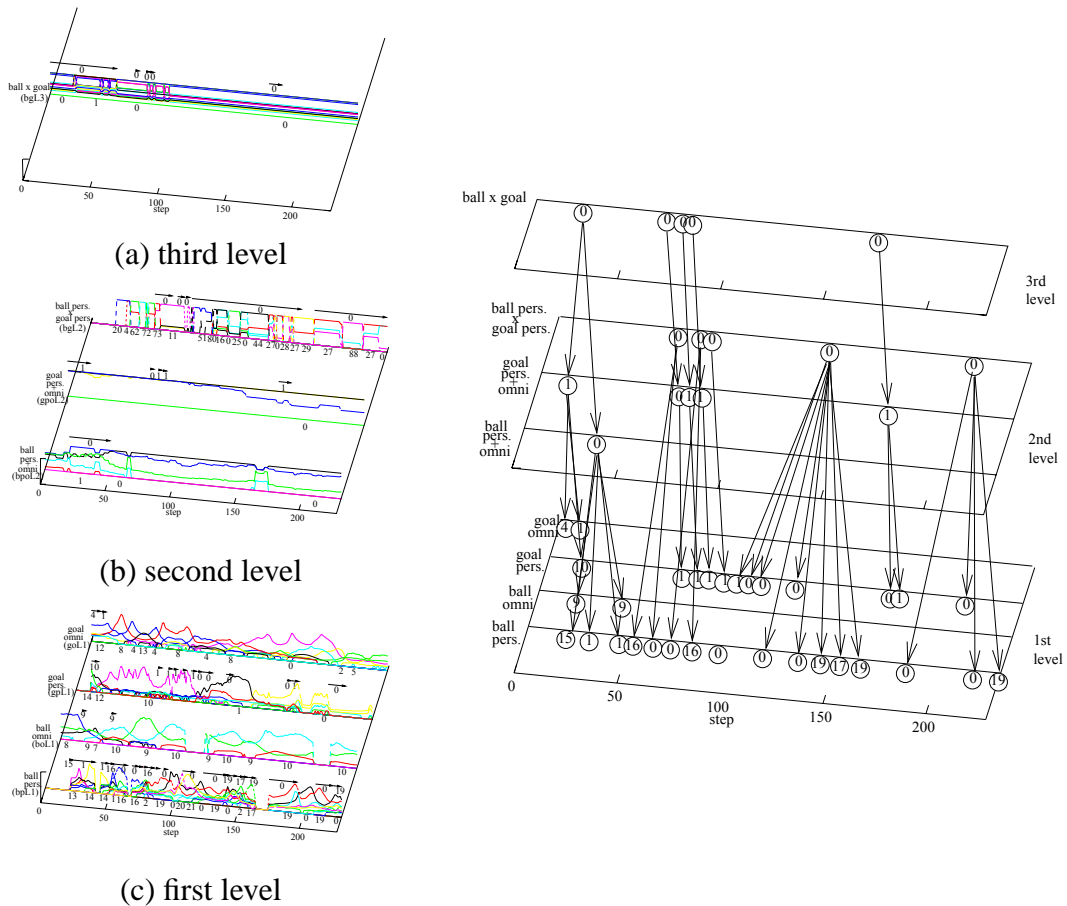


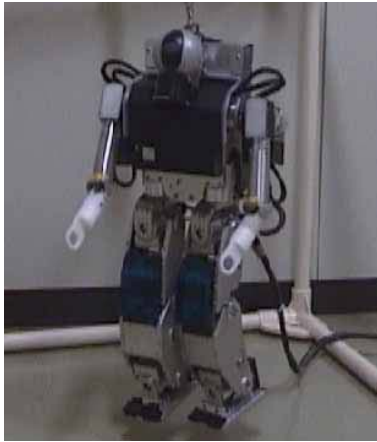
図 14: それぞれの層における学習器のゴール状態活性度と行動活性度の遷移 (左図) , および上位層から下位層への指令の系列 (右図)

層は相補的手法で構成した．図中の矢印はゴール状態活性度から状態ベクトルへの流れを示す．行動ベクトルから行動活性度への矢印はここでは省略した．

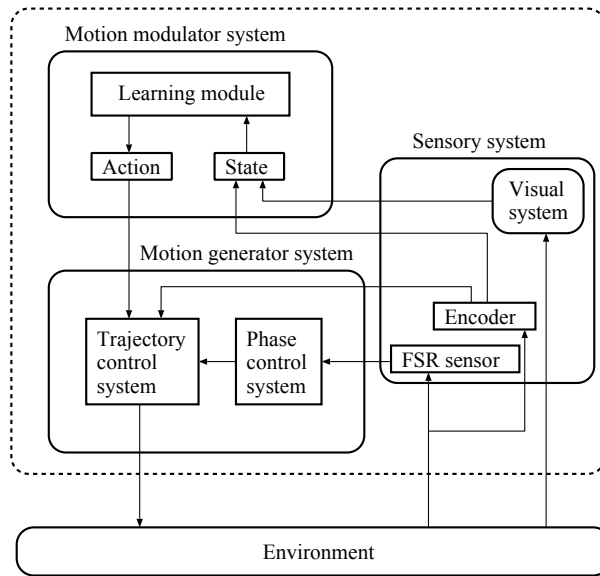
この階層型学習機構を使ってホーミング行動とシュート行動を実行させた．シュート行動ではロボットをゴールから離して反対方向を向かせて置き，ボールをそれらの間に置いた．目標状態はボールとゴールが前カメラで下の中央に見える状態を指定した．図 13 にそのときのロボットの動きと搭載されたカメラの画像を示す．

図 14 左にシュート行動をさせた時の階層型学習機構におけるそれぞれのレベル，それぞれの層のゴール状態活性度と行動活性度の遷移を示す．行動活性度はゴール状態活性度の上部に矢印として表示している．図 14 右に活性化された学習器の遷移と上位から下位への学習器への指令などの概略図を示す．下への矢印が上の層の学習器から下位の層のどの学習器を活性化させたかを示している．

3. 視覚情報に基づくヒューマノイドのリズム歩行パラメータの強化学習 二足歩行ロボットの実現手法は，ZMP 規範に代表されるモデルベースアプローチと，CPG に代表されるリズム歩行アプローチの二つの大別でき，後者は前者ほど詳細なパラ



(a) 実験で利用した富士通 HOAP-1



(b) システム概要

図 15: 実験で用いたヒューマノイドと全体システム

メータを必要とせず，環境への頑強性が期待されている．環境への適応性をより向上させるために，視覚情報などの外界センサーを用いた手法も提案されているが，環境モデルの精密な情報を必要とするものが多い．本研究は，ヒューマノイドの多様な行動生成手法のひとつとして，視覚情報に基づくヒューマノイドのリズム歩行パラメータの強化学習手法を扱う [6]．手法は二段階の学習過程からなる．最初は，後段の強化学習のための行動空間の生成である．歩行パラメータは前進ステップ幅と旋回角度であり，実現可能な組み合わせを直前の姿勢との組み合わせから，経験を通じて求める．このようにして求められた行動空間と視覚情報並びに姿勢情報から構成させる状態空間を用いて，歩行パラメータを強化学習する．タスクとしてロボカップ・ヒューマノイドリーグにおける状況を想定し，ゴール前に置かれたボールにアプローチしキックする動作を対象とする．

図 15 に実験で用いたヒューマノイド 富士通 HOAP-1(伸長 48cm) と全体システムの概要を示す．運動生成システムに埋め込まれているリズム歩行の実現に関しては，本研究プロジェクトの京都大学土屋・辻田研究グループが開発した関節角度の位相制御による歩行安定化手法を用いた．これは，軌道追従制御と位相シフト制御からなり，脚底に設置された FSR センサーからの信号により位相シフトが起動する．

これの上段に位置する強化学習モジュールでは，視覚情報に基づいて，リズム歩行パラメータである旋回角  $\alpha$ ，ステップ幅  $\beta$ (図 16(a) 参照) を決定する．この際，リズム歩行パラメータはヒューマノイドの姿勢に応じて，選択可能である場合と不可能の場合に分かれる．後者の意味は，転倒などの不安定姿勢に遷移することを表し，実機のロボットを使った学習実験では，ダメージを最小化する意味でも避けなけれ

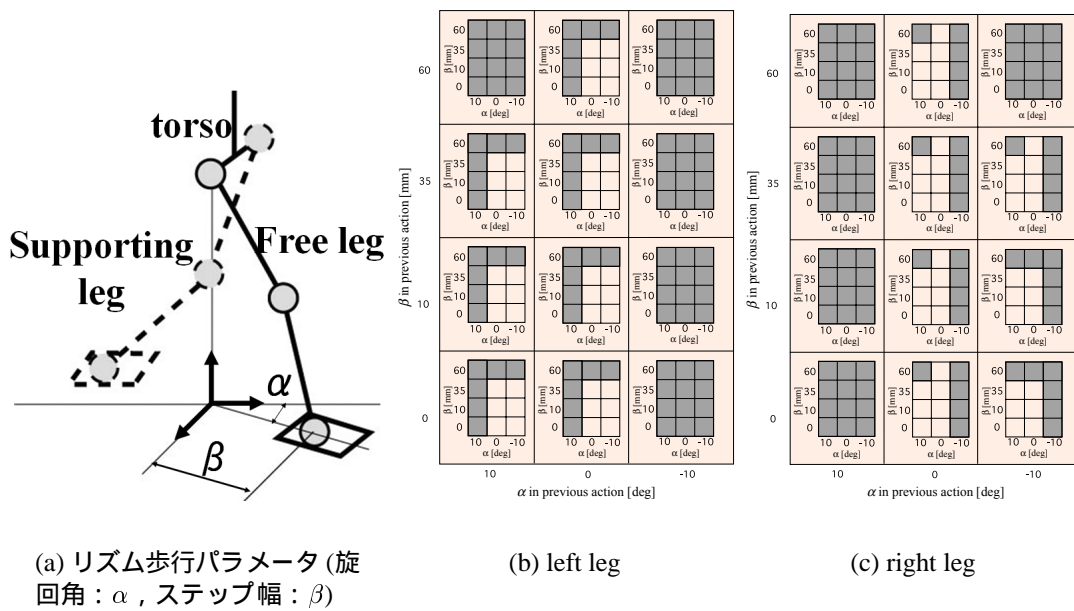


図 16: リズム歩行パラメータと選択可能な歩行パラメータ

ばならない．そこでまず最初に，選択可能なリズム歩行パラメータを求めた．旋回角  $\alpha$  は -10, 0, 10[度] の 3 種類，ステップ幅  $\beta$  は 0, 10, 35, 60[mm] の 4 種類，合計 12 種類の選択肢があるが，直前に選択されたパラメータにより，これらが転倒に至るかどうかを試行錯誤により識別した．図 16(b) ならびに (c) にその結果を示す．縦横の各軸が直前に実行された歩行パラメータのステップ幅及び旋回角をそれぞれ表す．白い部分が選択可能なパラメータであり，灰色の部分は選択禁止パラメータである．理想的には左右対称になるべきであるが，差異は実験で用いたロボット個体のハードウェアの属性の違いによる．

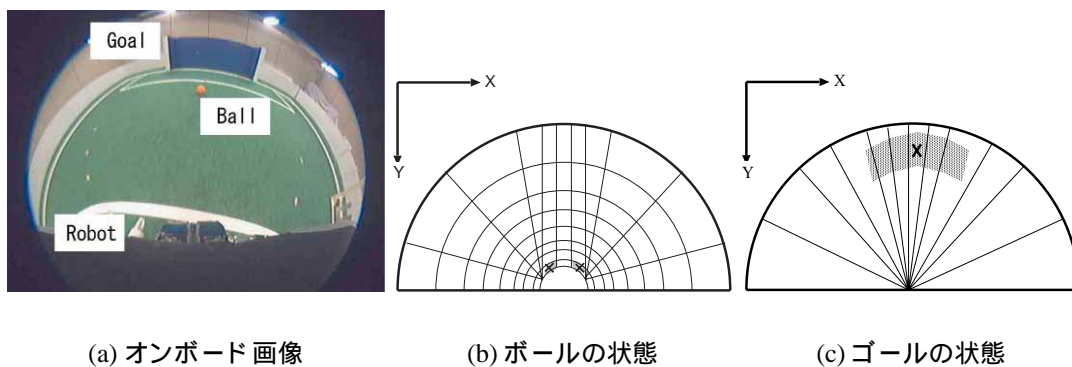


図 17: オンボード画像とボール，ゴールの状態空間

視覚情報に基づいた歩行パラメータ学習のタスクとして，ロボカップ・ヒューマノイドリーグにおける状況を想定し，ゴール前に置いたボールにアプローチしキッ

クすることを考える．視覚情報からボールやゴールの状態を決定し，それらに加えてヒューマノイドの姿勢(ここでは直前に実行した歩行パラメータに対応)を加えた状態空間を構成した．図 17(a)はヒューマノイドのオンボード画像であり，広角レンズを用いて視野を広め，色情報を用いてボール，ゴールを検出する．図 17(b)，(c)にボールとゴールの状態空間を示す．ボールについてはアプローチするので距離と方向で詳細な状態を，ボールに関しては方向のみからなる状態空間を構成する．ボールに関しては最近傍の半円の次の円弧状の領域のうち中央上部の左右の隣の領域(図中の×印)に報酬 1，それ以外は報酬ゼロである．これは左右どちらかの脚でキックするためである．ゴールに関しては，中央部分(図中の斜線部)で報酬 1，それ以外は報酬ゼロである．



図 18: 実験結果

歩行パラメータは 2 次元であるが，状態空間のサイズは 25632 であり，かなり探索に時間を要する．そこで，探索空間を絞るために，最初の 1 時間はインストラクターが歩行パラメータを教示し，そのあと約 10 時間の実機での学習により，ゴールにアプローチする行動が獲得された．図 18 にその様子を示す．同図 (a) では，斜め直進の様子が示されている．やや見難いが，アプローチするにつれ歩幅を調整している様子が伺える．同図 (b) では，ボールとゴールの見え方が報酬が得られる姿勢から程遠く，探索が必要であるが，ほぼ正しく歩行パラメータが求められている．同図 (c) では，最初，ボールもゴールも見失った状態から出発し，旋回し，発見してボールにアプローチしている．

## (2) 得られた研究成果の状況及び今後期待される効果

1. 脚式移動ロボットの行動決定のための情報量規範に基づく観測戦略学習 移動ロボットのための効率のよい観測を実現するための，情報量に基づく観測戦略を提案し，検証することを目的とした．情報量に基づく観測対象選択を行う観測戦略，情報量によるセンサ空間の自律的分割と注視窓による観測対象と観測方向の選択を行う観測戦略，移動中の観測を利用しより効率のよい行動決定のための観測戦略を提案した．それぞれについて、脚式移動ロボットを用いた実験により観測戦略が有効

に働き，適切な行動決定が行われることを検証した．

2. 複数の学習器の階層的構築による行動獲得 同一構造の学習器を複数用いて階層的に構築することによる行動獲得法を提案し，提案する手法をロボカップに出場しているロボットに適用した結果を示した．また，階層型学習機構において低レベルの大きな状態空間をいくつかの部分空間に分け，より高いレベルでそれらを統合する手法を提案し，同じく実機にて検証を行なった．今後は階層をより多層化し，より複雑なタスクに適用しすることで本手法の有効性を確かめたい．また複数のタスクを学習するとき，以前の学習結果を再利用する事による有効性を検証予定である．現段階では自律的な階層構造の構築のための機構は，単純な階層についてのみ提案しているが，設計者の介在なしに学習器の層を自律的に追加するように将来拡張する予定である．また，本研究では主に状態空間の階層性に注目してきたが，行動空間の階層性についても本手法と同様のアプローチが可能か，現在検討中である．

3. 視覚情報に基づくヒューマノイドのリズム歩行パラメータの強化学習 CPGに代表されるリズム歩行を用いて，外界からの情報を基に歩行パラメータを学習する手法を提案し，実機を用いて検証した．タスクとしてロボカップ・ヒューマノイドリーグの状況を想定し，ゴール前に置いたボールにアプローチする歩行を実現した．ヒューマノイドの多様な行動を外界情報から学習する手法として，今後，様々なタスクへの応用が期待できる．今後の課題として，状態空間の自律的構成などが挙げられる．

## 参考文献

- [1] 光永, 浅田：移動体の意思決定のための情報量基準に基づく観測対象選択戦略, 日本ロボット学会誌, vol.19, no. 6, pp.793–800, 2001.
- [2] 光永, 浅田：情報量に基づく移動ロボットの注視制御のためのセンサ空間構成, 日本ロボット学会誌, vol.20, no. 7, pp.751–758, 2002.
- [3] 光永, 浅田：歩行中の観測を考慮した情報量に基づく注視制御, 日本ロボット学会誌, vol.21, no. 7, 2003 (掲載予定).
- [4] 高橋 泰岳, 浅田 稔：複数の学習器の階層的構築による行動獲得, 日本ロボット学会誌, Vol.18, Num.7, pp.1040–1046, 2000
- [5] 高橋 泰岳, 浅田 稔：階層型学習機構における状態行動空間の構成, 日本ロボット学会誌, Vol.21, Num.2, pp.38–45, 2003
- [6] M. Ogino, Y. Katoh, M. Aono, M. Asada and K. Hosoda : Reinforcement learning of humanoid rhythmic walking parameters based on visual information, Proc. of the

2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003  
(to appear).