学術・技術論文

階層型学習機構における状態行動空間の構成

高 橋 泰 岳* 浅 田 稔*

State-Action Space Construction for Multi-Layered Learning System

Yasutake Takahashi* *and* Minoru Asada*

This paper proposes a multi-layered reinforcement learning system that integrates lower learning modules and generates one of higher purposive behaviors based on which an autonomous robot learns from lower level behaviors to higher level ones through its life time. We decompose a large state space at the bottom level into several subspaces and merge those subspaces at the higher level. This allows the system to reuse the policies already learned and to learn the policy against the new features. As a result, curse of dimension is avoided. To show its validity, we apply the proposed method to a simple soccer situation in the context of RoboCup, and show the experimental results.

Key Words: State/Action Space Construction, Reinforcement Learning, Multi-layered Learning, RoboCup

1. はじめに

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（例えば文献 [1]〜[3]）．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．

異なる状態行動空間を扱った学習／制御器を効率的に利用することで，新しい環境下での適応や，さらに多様な状況に対する行為を獲得することが可能である．これまでにも一度獲得された政策を再利用する手法がいくつか提案されている（例えば文献 [4]〜[6]）．これらの研究は，以前に置かれていた環境下で得られた知識をもとに，新しい環境下において行動を調整することに主眼を置いたものである．これらの手法ではロボットが扱う状態変数は固定であり，また与えられるタスクも固定であるので，獲得した政策をもとに多様な状態変数で表現される新たな状態空間への対応は難しい．一方で，異なる状態空間で学習された結果をそれらの空間を直交化した状態空間に投影することで，より次元の高い状態空間における学習の効率化を目指した研究がある（例えば文献 [7] [8]）．しかし，これらの手法が目的としているのは政策が独立した二つのタスクをそれぞれ学習させ，これらの複合した新たなタスクに対して直交化された状態空間における学習効率を議論しているので，最終的に用意すべき状態空間は状態変数すべてを直交化して張り，これを密に離散化した巨大なものにならざるを得ない．

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

原稿受付
*大阪大学大学院工学研究科
*Graduate School of Engineering, Osaka University

が爆発するようなことは抑えられると考えられる．

鮫島ら [9] は環境の予測可能性に基づいて状況を分割する複数モジュール競合アーキテクチャMOSAIC を強化学習に拡張した手法を提案している．複数の線形予測モデルの競合を用いて非線型・非定常な環境を時空間的に区分し，それを状況として認識し，その状況に対応する制御則を強化学習により獲得する．この考え方は非常に興味深いが，基本的に一層における学習器の切り替えに重点をおいているので，抽象化された状況・行為を基により抽象度の高い行為を獲得するという過程は考慮されていない．

高橋，浅田 [10] は同一構造の学習器を複数用いて階層的に自律的に構築することによる行動獲得法を提案した．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，低レベルの行為を学習する．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．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．

しかしこの手法はすべての状態変数を含んだ状態空間が定義されたもとで状況・行為の抽象化が行われていることを仮定しているので，異なる部分空間を状態空間として持つ学習器を統合し，より多くの状態変数で表現される状況に再利用することを考慮していない．

そこで本論文では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手法を提案する．この手法は以下の利点がある．

① 学習済みの階層構造をそのまま利用して新しい学習器層を構成することが可能で，新たに増えた状態変数に対応した行動を獲得することができる．

② 下位の層において部分的な状態空間上で獲得された行動学習器群をより上位の層で組み合わせることによって，多くの状態変数を同時に扱う行為を獲得できる．

③ 全状態空間を部分空間に分割するので，結果として一つの学習器が対応する状態数が減り，必要な計算資源を抑えることができる．

提案する手法をロボカップ [11] に出場しているロボットに適用した結果を示す．

2. 階層型学習機構

階層型学習機構の基本的なアーキテクチャは [10] と同じである．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（**Fig.1**）．この学習器の階層構造はタスク分解の役割を担っていると見なすことができる．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 [10] を参照）．

この階層型学習機構の中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．**Fig.2** 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．

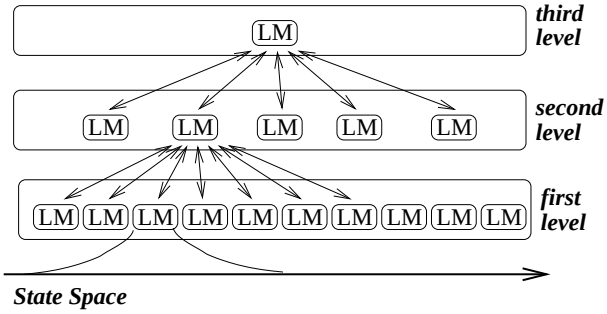


Fig. 1 An overview of a hierarchical learning architecture : LM stands for learning module

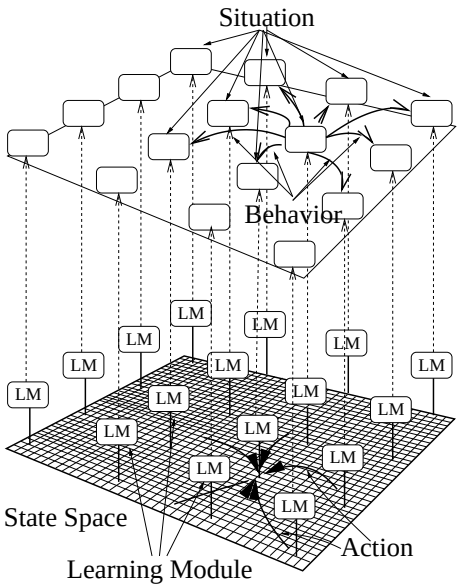


Fig. 2 The relationships of situations and behaviors inside a layer and between different levels

3. 上位の層における状態／行動空間の構成

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

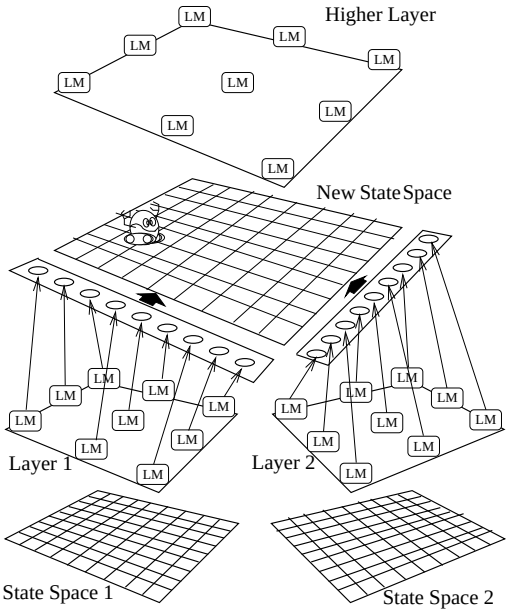


Fig. 3 State-action space construction based on multiplicative approach

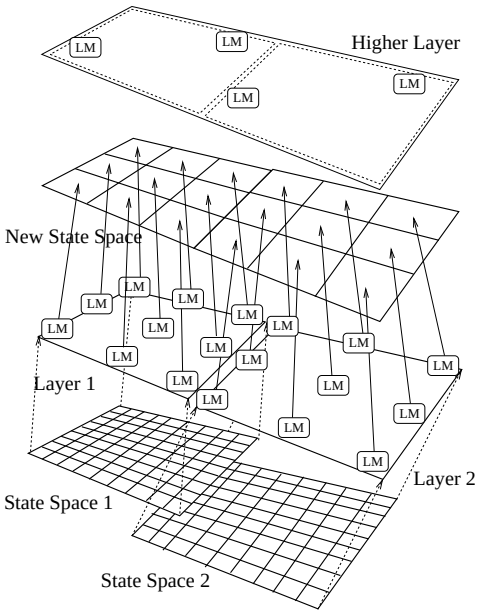


Fig. 4 State-action space construction based on complementary approach

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」を以下に提案する．

（a） 直積的手法

二つ以上の層を上位の層で統合する一つの方法を **Fig.3** に示す．状態空間は下位のそれぞれの層の状況の直積を取り構成する．行動空間は下位の層の学習器への指令として定義され，一つの行動につき一つの学習器が割り当てられる．例えば Layer 1 の学習器が n 個，Layer 2 の学習器が m 個だった場合，$n \times m$ の状態と $n+m$ の行動を作る．

（b） 相補的手法

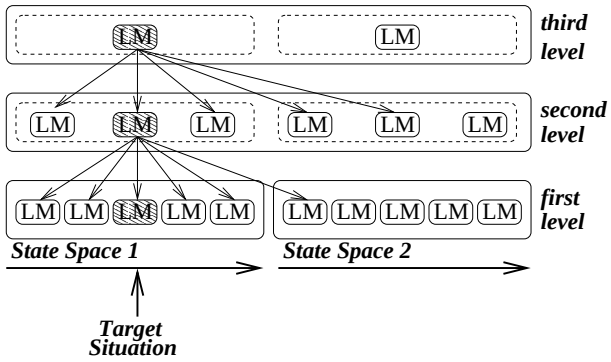
もう一つの方法を **Fig.4** に示す．一つ上の層の学習器で，下位の二つの層を統合する．統合する際，下位の層の学習器によって定義される状況／行為が上の層の学習器のそれぞれ状態／行動に 1 対 1 に対応する．具体的には下位の一つの層に n 個の学習器，もう一つの層に m 個の学習器がある場合，$n+m$ 次元の状態ベクトルと行動ベクトルを構成する．この方法は直積的手法に比べて計算資源の使用量を抑えることができる．下位の層（Fig.4 の Layer 1）の学習器が担当する範囲（State Space 1）を他の層（Layer 2）の学習器が担当する範囲（State Space 2）で補間する（あるいはその逆）ことで，上位層（Higher Layer）の学習器がより広範囲の状況をシステム全体として把握することをこの手法は可能にする．

4. タスク遂行時の階層型学習機構内での行動戦略

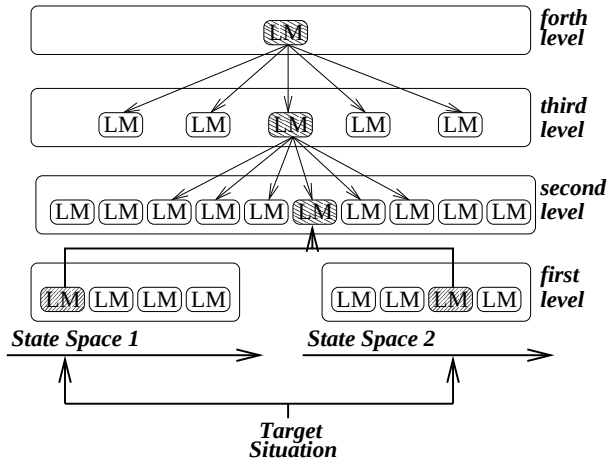
システムに与えられる目標状態は最下位の層の状態空間の中で与えられる．**Fig.5**（a） にこの状況を示す．もし目標状態が一つの状態空間上で与えられれば，文献 [10] で示した方法で行動をとる．つまり，システムは最下位の層の中で与えられた目標状態に一番近いゴール状態を持つ学習器を探す．この学習器で現在の状況から目標状態に到達できるなら，この学習器を起動させて目標状態へ移動する行動をとらせる．そうでないときは，一つ上の層で目標状態に近い学習器を探す．これを繰り返すことになる．以下にアルゴリズムを示す．

① 目標状態が指定される．
② 最下位の層の中で与えられた目標状態に一番近いゴール状態を持つ学習器を探す．
③ この学習器が現在の状況に対応しているか?
YES :この学習器を起動させて目標状態へ移動する行動をとらせる．
NO :一つ上の層で目標状態に近い学習器を探し，③へ．

もし目標状態が複数の状態空間にまたがって指定されたとき，システムは目標状態が一つの状態空間で表せられる最下位の層を探す．Fig.5（b）にこの状況を示す．目標状態が最下位の層の異なる複数の状態空間で指定されている．システムは下位の二つの層を統合した上位の層の状態空間上で目標状態を内部で規定する．次にこの層に含まれる学習器のなかで一番目標状態に近いものを選び，この学習器で現在の状況から目標状態に到達できるなら，この学習器を起動させる．そうでないときは，さらに一つ上の層で目標状態を探す．以下にアルゴリズムを示す．



(a) One target situation at one state space



(b) Another target situation at two state spaces

Fig. 5 Strategy in the multi-layered control structure

① 目標状態が指定される.
② 指定された空間をすべて含む空間を状態空間として持つ層を全てリストアップする.
③ そのリストの中で最下位の層を探す.
④ その層の中で与えられた目標状態に一番近いゴール状態を持つ学習器を探す.
⑤ この学習器が現在の状況に対応しているか?
YES :この学習器を起動させて目標状態へ移動する行動をとらせる.
NO :リストの中で次に低い層を探し，④へ.

5. 実　　験

5.1 設定

提案する手法を検証するために，移動ロボットによる簡単なホーミング行動とシュート行動の実験を行った．**Fig.6** に使用するロボットとボール，ゴールを示す．**Fig.7** に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，2 枚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．目標状態は，実際に移動ロボットをシュート行動が完了した状態まで移動させ，センサ情報を読み込ませることで与える．具体的にはロボットはボールと敵ゴールを前方のカメラで中央下にとらえた状態になる．

Fig. 6 A mobile robot, a ball, and a goal

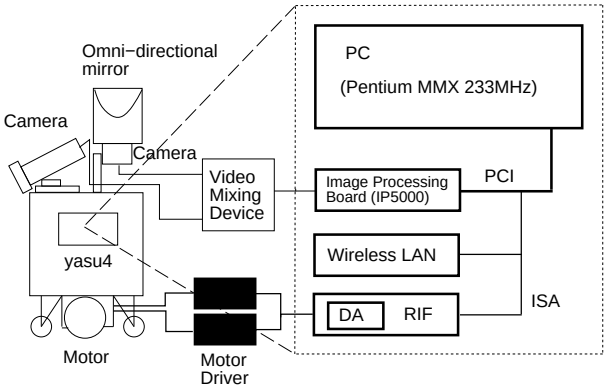


Fig. 7 An overview of the robot system

最下位の層の状態空間として **Fig.8** にあるように，前カメラ画像上と全方位画像上それぞれにおいてボールとゴールのそれぞれの位置を離散化した空間を用意した．前カメラ画像上では 11×21，全方位画像上では 15×15 にそれぞれ離散化した．また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，これも 5×5 で離散化した．

階層構造は今回 Fig.8 のように設定した．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.*×*goal pers.*” 層と第 3 レベルの “*ball*×*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

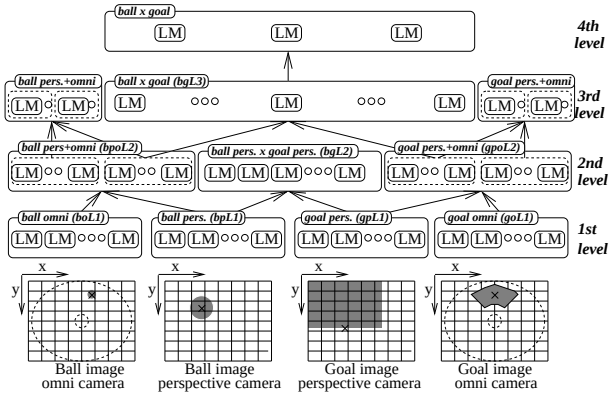


Fig. 8 A hierarchy architecture of learning modules

ら行動活性度への矢印はここでは省略した．第 4 レベル以上の層は文献 [10] で示される方法で自律的に構成させた．

5. 2 実験結果

実験は学習段階と学習結果を用いたタスク遂行の段階の二つの段階からなる．まず，ロボットは約 2 時間，環境の中をランダムに移動し，センサ・モータコマンド系列のデータをとり，これ使って上記の階層構造学習機構を学習した．

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．ホーミング行動を取る場合はボールもしくはゴールに関する情報のみを必要とするので，例えばゴール前に移動するホーミング行動の場合，Fig.8 に示す右側の層（第 1 レベルの *goal pers.* 層および *goal omni* 層，第 2 レベルの *goal pers.+omni* 層，第 3 レベルの *goal pers.*）の学習器を利用し，タスクが遂行される．詳しい動作については文献 [10] とほぼ同じなので割愛する．シュート行動を実行するときロボットがタスクを遂行するのに必要な情報はボールとゴール両方であるので，Fig.8 のほとんどの層（第 3 レベルの *ball pers+omni* 層および *goal pers.+omni* 層以外）を活用する．ホーミング行動とシュート行動という具合に，タスクによって必要とする方法が異なる場合でも同じ階層型学習機構を使い，内部で活用する学習器を切替えることで対応できることが分かる．

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．**Fig.9** 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．**Fig.10** にそれぞれのレベル，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．**Fig.11** 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．

このシステムはまず目標状態（ボールとゴールが前カメラの画面中央下に見える状態）を担当している第 2 レベルの “*ball pers.×goal pers.*” 層の番号 0 の学習器（*bgL2-0*）で目標状態への到達を試みるが，前カメラによって対象を認識しておらず，制御できない．そこで第 3 レベルの “*ball×goal*” 層の番号 0 の学習器（*bgL3-0*）に制御を任せる．この学習器は第 2 レベルの

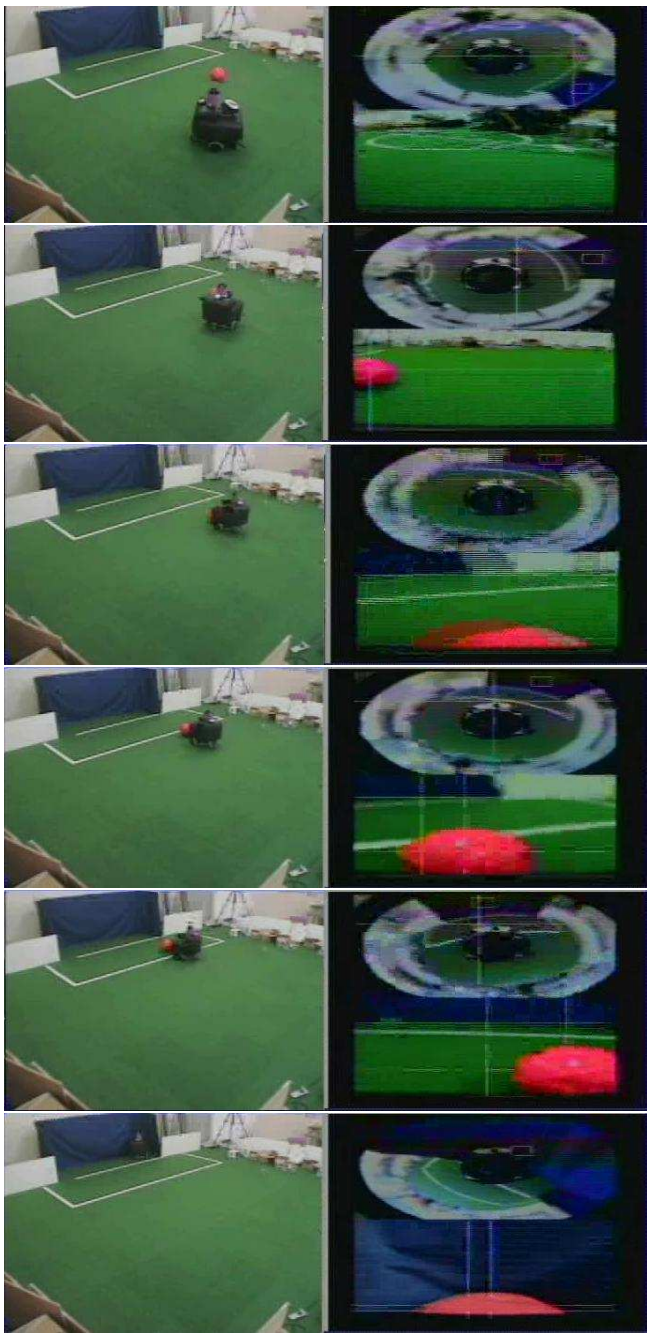
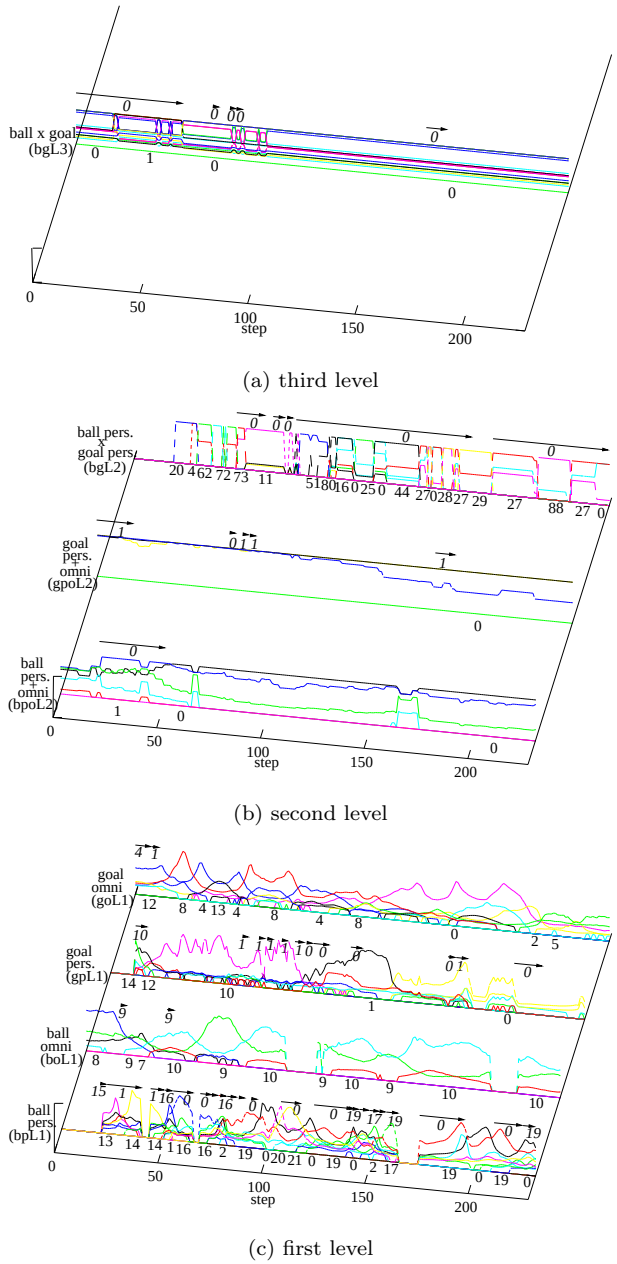


Fig. 9 A sequence of a shooting behavior and its camera images

“*goal pers.+omni*” 層の番号 1 の学習器（*gpoL2-1*），および “*ball pers.+omni*” 層の番号 0 の学習器（*bpoL2-0*）を起動する．これらの学習器は第 1 レベルの適切な学習器を起動することによって，所定の行動を実行している．第 2 レベルの “*ball pers.×goal pers.*” 層の番号 0 の学習器 *bgL2-0* で，目標状態への到達が可能と判断されれば，ロボットの制御をこの学習器に任せている．たまに，ボールとぶつかることでボールが横にこぼれるなどしてこの学習器の制御可能範囲外に状況が変化すれば，また第 3 レベルの “*ball×goal*” 層の番号 0 の学習器（*bgL3-0*）に制御を任せ，目標状態への遷移を試みている．最終的には第



(a) third level

(b) second level

(c) first level

Fig. 10 A sequence of the goal state activation and behavior activation of learning modules

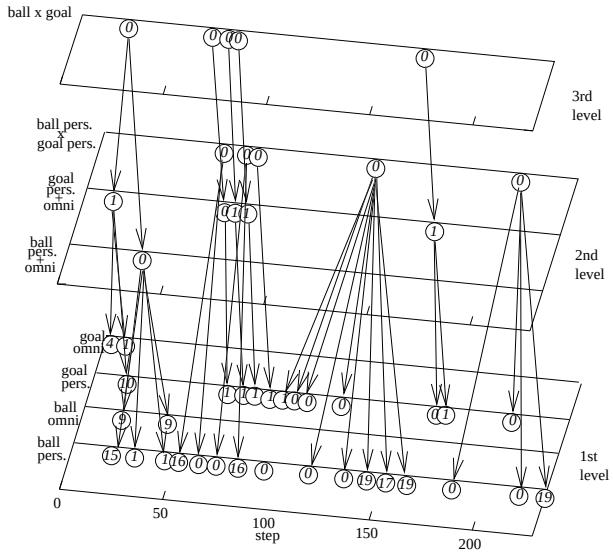


Fig. 11 A sequence of the behavior activation of learning modules and the commands to the lower layer modules

2 レベルの “*ball pers.*×*goal pers.*” 層の学習器 *bgL2-0* が担当している目標状態へ状況が遷移することで，タスクを遂行している．

5.3 考察

ここでは 1 章で述べた三つの利点を検討する．

① 学習済みの階層構造をそのまま利用して新しい学習器層を構成することが可能で，新たに増えた状態変数に対応した行動を獲得することができる

今回の実験では Fig.8 に示す階層構造にて行為の学習を行わせた．もしロボットが置かれた環境にボールしかなかった場合，もしくは他のものがあっても知覚されなかった場合は，システムとして Fig.8 の左半分，つまり第 1 レベルの “*ball omni*” 層および “*ball pers.*” 層，第 2 レベルの”*ball pers.+omni*” 層，第 3 レベルの “*ball pers.+omni*” 層の学習器に経験を積ませることで，ボールへのアプローチという行為を獲得することができる．次にシステムがゴールを知覚したとき，Fig.8 に示す右半分，つまり第 1 レベルの “*goal pers.*” 層および “*goal omni*” 層，第 2 レベルの”*goal pers.+omni*” 層，第 3 レベルの “*goal pers.+omni*” 層の学習器を追加することで，ゴールに対する簡単なナビゲーション行動を獲得するといった，新しい状態変数に対応した行動を獲得することが可能である．

② 下位の層において部分的な状態空間上で獲得された行動学習器群をより上位の層で組み合わせることによって，多くの状態変数を同時に扱う行為を獲得できる．

第 1 レベルではボールとゴールを同時に把握する学習器はなかったが，第 2 レベルに “*ball pers.*×*goal pers*” 層，第 3 レベルに “*ball*×*goal*” 層，第 4 レベルに “*ball*×*goal*” 層を設けることにより，ボールとゴールを同時に把握する行動，シュート行動を獲得することができることを実験を通して示した．また，ナビゲーションの経験を積んだシステムが，獲得した学習結果をもとにシュート行動をより効率良く獲得する結果をシミュレーションを元にすでに得ている [12].

③全状態空間を部分空間に分割するので，結果として一つの学習器が対応する状態数が減り，必要な計算資源を抑えることができる．

最後に本手法が従来の手法と比べて計算機の必要とされるメモリ資源が大幅に緩和されることを示す．Table 1 に本手法で実際に構築されたそれぞれの層の状態数，行動数，Q テーブルのサイズ，配置された学習器数，層全体で必要とされたメモリ数を示す．比較として Table 2 に従来手法を適用したときの予想されるデータを示す．現実問題としてこれを実装することは非常に困難であるので，概数として示す．第一レベルにおいて，提案する手法の方は状態空間が分割されているので状態数は 10^2 のオーダですみ，すべての学習器に必要なメモリを足しても 10^5

Table 1 Required memory size for the proposed layered learning system

level	layer	# of states A	# of actions B	Q table size $C = A \times B$	# of modules D	memory size $C \times D$
1	ball pers.	231	25	5,775	25	144,375
	ball omni.	225	25	5,625	18	101,250
	goal pers.	231	25	5,775	24	138,600
	ball omni.	225	25	5,625	23	129,375
2	ball pers.+omni.	43	43	1,849	9	16,641
	goal pers.+omni.	47	47	2,209	9	19,881
	ball pers. × goal pers.	600	49	29,400	65	1,911,000
3	ball pers.+omni.	9	9	81	1	81
	goal pers.+omni.	9	9	81	1	81
	ball × goal	81	18	1,458	15	21,870
4	ball × goal	15	15	225	8	1,800
				sum total	193	2,465,073

Table 2 Required memory size for the layered learning system with monolithic state and action spaces

level	layer	# of states A	# of actions B	Q table size $C = A \times B$	# of modules D	memory size $C \times D$
1	full state-action	$207,936$	25	$5,198,400$	10^4	10^{11}
2	full state-action	10^4	10^4	10^8	10^3	10^{11}
3	full state-action	10^3	10^3	10^6	10^2	10^8
···	full state-action	···	···	···	···	···
				total	10^4	10^{11}

のオーダですむ．一方で従来手法ではすべての状態変数を考慮して状態空間を構成するため，状態数は 10^5 のオーダになり，一つの学習器に必要なメモリ容量は 10^6 のオーダにもなる．実際はこれにより多くの学習器が配分されるので，その学習器を掛けた分のメモリ容量が必要となる．ここでは用意した状態空間の状態数の 10 分の 1 の学習器が各層において割り当てられると仮定した．どちらのシステムにおいても，第 2 レベルにおいて計算資源がより多く必要とされることが分かる．最終的なオーダとしては状態空間を分割して後で統合する提案手法が 10^6，分割せずに巨大な状態空間を元に階層化をする手法が 10^{11} である．必要とされる計算資源は大幅に緩和されたと考えられる．

6. お わ り に

本論文では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た．提案する手法をロボカップ [11] に出場しているロボットに適用した結果を示した．

現段階では自律的に階層構造を構築する機構が入っていないが，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．

謝辞 本研究は科学技術振興事業団の戦略的基礎研究推進事業「脳を創る」中村プロジェクトの援助を受けた．

参 考 文 献

[1] M. Asada, S. Noda, S. Tawaratumida, and K. Hosoda. Purposive behavior acquisition for a real robot by vision-based reinforcement learning. *Machine Learning*, Vol. 23, pp. 279–303, 1996.

[2] Eiji Uchibe, Masateru Nakamura, and Minoru Asada. Cooperative behavior acquisition in a multiple mobile robot environment by co-evolution. In Minoru Asada, editor, *RoboCup-98: Robot Soccer World Cup II, Proc. of the second RoboCup Workshop*, pp. 237–250, 1998.

[3] Noriaki Mitsunaga and Minoru Asada. Observation strategy for decision making based on information criterion. In *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1038–1043, 2000.

[4] Sebastian Thrun. A lifelong learning perspective for mobile robot control. In *In Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems*, Vol. 1, pp. 23–30, 1994.

[5] Fumihide Tanaka and Masayuki Yamamura. An approach to lifelong reinforcement learning through multiple environments. In *6th European Workshop on Learning Robots*, pp. 93–99, 1997.

[6] 港, 浅田. 環境の変化に適応する移動ロボットの行動獲得. 日本ロボット学会誌, Vol. 18, No. 5, pp. 706–712, 2000.

[7] Eiji Uchibe, Minoru Asada, and Koh Hosoda. Behavior coordination for a mobile robot using modular reinforcement learning. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems 1996 (IROS '96)*, pp. 1329–1336, 1996.

[8] 榎田修一, 河野宗一, 大橋健, 江島俊朗. センサ空間の拡大を用いたeq-学習. 第 19 回日本ロボット学会学術講演会, pp. 85–86, 2001.

[9] 鮫島和行, 銅谷賢治, 川人光男. 強化学習 mosaic: 予測性によるシンボル化と見まね学習. 日本ロボット学会誌, Vol. 19, No. 5, pp. 551–556, 2001.

[10] 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, No. 7, pp. 1040–1046, 2000.

[11] M. Asada, H. Kitano, I. Noda, and M. Veloso. Robocup: Today and tomorrow – what we have learned. *Artificial Intelligence*, Vol. 110, pp. 193–214, 1999.

[12] Yasutake Takahashi and Minoru Asada. Behavior acquisition by multi-layered reinforcement learning. In *Proceeding of the*

1999 IEEE International Conference on Systems, Man, and Cybernetics, pp. 716–721, 1999.

高橋　泰岳（Yasutake Takahashi）

1972 年 12 月 13 日生．1994 年大阪大学大学院学研究科博士前期課程修了．2000 年同大学博士後期課程中退，同年同大学大学院工学研究科助手．知能ロボットの行動獲得に関する研究に従事．博士（工学）．（日本ロボット学会正会員）

浅田　稔（Minoru Asada）

1953 年 10 月 1 日生．1982 年大阪大学大学院基礎工学研究科後期課程修了．同年，大阪大学基礎工学部助手．1989 年大阪大学工学部助教授．1995 年同教授．1997 年大阪大学大学院工学研究科知能・機能創成工学専攻教授となり現在に至る．この間，1986 年から 1 年間米国メリーランド大学客員研究員．知能ロボットの研究に従事．1989 年，情報処理学会研究賞，1992 年，IEEE/RSJ IROS'92 Best Paper Award 受賞．1996 年日本ロボット学会論文賞受賞．博士（工学）．電子情報通信学会，情報処理学会，人工知能学会，日本機械学会，計測自動制御学会，システム制御情報学会，IEEE R&A, CS, SMC societies などの会員．

（日本ロボット学会正会員）