自己の価値に基づく他者行為理解

Behavior Understanding based on Shared Value

高橋 泰岳 †， 河又 輝泰 †， 浅田 稔 ††

† 大阪大学大学院工学研究科　†† 大阪大学大学院工学研究科/科学技術振興機構 ERATO 浅田共創知能システムプロジェクト

大阪府吹田市山田丘 2-1

Yasutake TAKAHASHI †　Teruyasu KAWAMATA †　and　Minoru ASADA ††

† Graduate School of Engineering, Osaka University　†† Graduate School of Engineering, Osaka University/JST ERATO Asada Synergistic Intelligence Project

Yamadaoka 2-1, Suita, Osaka, Japan

要約

神経生理学の分野において，ある行動を自身が再現する際に発火するだけでなく，同じ行動を他者が再現した際にも発火するミラーニューロンの存在が示された．このミラーニューロン・システムのコンセプトは非常に興味深く，行為の獲得と他者行為に認識は密接に関係があることを示している．つまり，行為学習器は行動獲得や実現のみならず，他者行為の認識・理解・推定にも利用できる可能性がある．

そこで，複数の行為を学習するだけではなく，他者行為の認識・理解を観察者の学習済みの行為の強化学習における状態価値を用いることにより可能にする手法を提案する．ここで状態価値とは将来に渡って得られるであろう報酬の減衰和であり，目標状態のみで正の報酬を得られる場合，任意の意図に従って行動する際には，報酬を得られる目標状態に向かうため，この状態価値が向上する．つまり，同じ目的をもった行為を実行している限り，動作系列が異なっても状態価値の値は増加していく傾向にあるので，この傾向から他者行為を認識できる．また，状態価値の変化の傾向は，観測状態の相対的な変化から比較的容易に得ることが可能であるので，観察者と行為実行者の視点の差異を吸収可能であると期待できる．本論文ではロボットのサッカーを例題とし，提案手法の有効性を検証する．

キーワード ： 価値システム，行為理解，模倣，強化学習

Abstract

Neurophysiology has revealed the existence of mirror neurons in brains of macaque monkeys and they shows similar activities during executing and observation of goal directed movements performed by self and other. The concept of the mirror neurons/systems is very interesting and suggests that behavior acquisition and recognition of observed behavior are closely related to each other. That is, the behavior learning modules might be used not only for behavior acquisition/execution but also for the understanding of the behavior/intention of other.

We propose a novel method not only to learn and execute a variety of behavior but also to understand behavior of others supposing that the observer has already acquired the utilities (state values in reinforcement learning scheme) of all kinds of behavior the observed agent can do. The method does not need a precise world model or coordination transformation system to deal with view difference caused by different viewpoints. This paper shows that an observer can understand/recognize a behavior of other not by precise object trajectory in allocentric/egocentric coordinate space but by estimated utility transition during the observed behavior.

Key words ： Value system, Behavior understanding, Emulation, Reinforcement Learning

1. はじめに

人間がどのように他者の行動を捉えているかを示唆した知見として，ミラーニューロンが挙げられる．ミラーニューロンはある行動を自身が再現する際に発火するだけでなく，同じ行動を他者が再現した際にも発火するニューロンである[2, 8]．また，Jeannerod[5]は行動を再現する際に発火するニューロンが，行動を想起し，仮想的にシミュレートするだけでも発火することを実験によって示している．これらの知見から，人間が他者の行動を観察した際には，その行動に対応した自身の表象に照らしあわせてその行動を捉え，また自身の内部でその行動をシミュレートすることが可能であると考えられる．

一方で，近年ではロボットに関する研究や開発が多くなされており，今後は人間の生活に近いシーンでもロボットが活躍することが期待されている．そのため，ロボットにも人間と協調して行動できる能力が望まれており，その基礎研究として複数ロボットにおける協調行動などを目指した研究がさかんに行なわれている[4][6][12]．他者の行動予測に関して研究としてNagayuki et al.[7]は観察者が他者の動作系列を観察して記憶し，他者の行為のモデルとして獲得する手法を提案し，観察者自身の行動の決定の際に，他者の行為モデルを用いて予測することで，自身のタスク達成に最適な行動を選択している．Tohyama et al.[9]は，他者の行為が単一ではなく複数ある場合に対応して，観察者が他者の行為モデルを複数持つことを提案している．その複数のモデルの中から，観測された他者の状態遷移に対して尤度の高い他者の行動モデルを他者の意図とみなし，予測に用いている．また，Inamura et al.[3]は隠れマルコフモデル (HMM) を用いてヒューマノイドにおける運動パターンの認識および生成を行なっている．行為認識においては，事前の学習によって，複数の行動を HMM を用いたシンボルとして表現しておき，実際に観測によって得られた他者の関節角に基づく行動要素系列に対して，尤度の高い HMM を他者の行為として認識をしている．これらの研究では他者の状態や行動を観察者が完全に知ることを仮定しているが，この仮定は現実的ではない．

これに対し，鮫島ら[11]は大脳基底核の計算論モデルである強化学習を応用した MOSAIC によって，他者のセンサデータを直接必要とせずに，観察によって他者の行動の行為認識を実現している．鮫島らは意図を”「大きな目標」を達成するための「動的な小さな目標の連鎖」“として捉えており，他者の選択している学習器を推定することを行為認識としている．この手法では，観察者は複数の学習器を持っており，それぞれの学習器は予測器と制御器によって構成されている．他者の行動の観察により他者の状態を得，次状態を各学習器の予測器毎に予測させる．そして，次の時刻において実際に観測された他者の状態と，予測された状態のユークリッド距離を誤差として比較し，最も誤差の小さい予測をした学習器を他者の行為と見なす手法である．しかし，同じ行為でも複数の状態遷移系列が多く存在する場合，予測器の予測する状態の遷移系列と実際の状態遷移の系列を比較する鮫島らの手法では，それぞれの状態遷移系列が同一の行為を示していると認識できない．これは相手の行為を自身の唯一の行為としての状態遷移系列によって捉えててしまうことの欠点である．

一方，他者の状態を観察によって獲得する場合，観察者の視点によって，状態識別に大きな誤差が生じる．鮫島らや従来の手法のように，他者の行為を状態遷移に基づいて推定する場合，観察者の視点の差による状態の誤識別が，行為認識に大きく影響を及ぼす可能性がある．この問題に対し，観察によって得た画像情報を三次元再構成等の手法を用いて座標変換することも考えられるが，これらの手法は環境やタスクに関する多くの情報を事前に必要とするため望ましくない．

そこで，本研究では強化学習における状態価値を用いることで，他者の意図を推定する手法を提案する．状態価値とは将来に渡って得られるであろう報酬の減衰和であり，目標状態でのみ正の報酬を得られる場合，合目的的な方策に従って行動する際には，状態価値は上昇する．状態価値のこの性質から，同一の目標に到達するための行為であれば状態価値が高くなる傾向にあるので，動作系列が異なる場合もそれらを同一の行為として認識することが可能である．

2. 状態価値による他者行為の認識

本節では強化学習，状態価値，状態価値に基づく行為認識を実現する提案手法について述べる．なお，本節以降では行為を認識する主体を観察者 (observer)，対象となる行為を実行する主体を実演者 (performer) と呼ぶ．

2．1 強化学習に基づく行為学習と状態価値

Fig.1 に強化学習の概念図を示す．ロボットは環境から状態 $s \in \mathcal{S}$($\mathcal{S}$ は可能な状態の集合) を観測する．環境はマルコフ過程に従うと仮定し，ロボットは現状態 s_t においてある方策 π に基づき行動選択し，次状態に遷移し，報酬 r_{t+1} を受け取る．状態価値 V^π は方策 π に従って行動選択しているときに将来にわたって受け取るだろう報酬の減衰和であり，以下の式で表

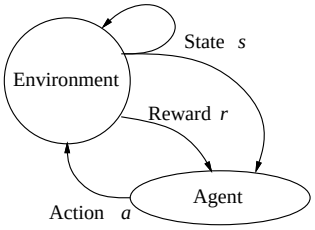


Fig. 1　Agent-environment interaction

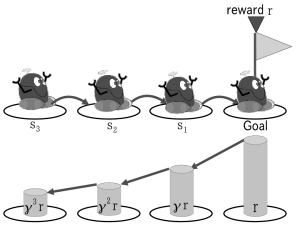


Fig. 2　Sketch of state value propagation

される．

$$V^{\pi} = \sum_{t=1}^{\infty} \gamma^t r_t \tag{1}$$

Fig.2 にロボットがゴール状態に止まったときに正の報酬を，それ以外の状態では 0 の報酬を受け取った場合の状態価値の模式図を示す．状態価値はロボットが正の報酬を受け取る場所で最も高い値を持ち，それより遠い状態には減衰された値が伝搬される．

合目的的な行為を行う方策 π に従うことで状態価値が上がる傾向にある．強化学習ではこの状態価値がより高くなるように方策を修正し，修正した方策に対応する状態価値を推定し直す手続きを繰り返し踏むことで，方策を改善する．本論文では獲得された方策 π に従う行動の系列を行為と呼ぶ．より詳しい説明は Sutton and Barto の著書 [10] や学習ロボットのサーベイ [1] に詳しい．

2．2　状態価値に基づく行為認識

状態価値とは式 (1) が示すように，適切な行動を取った時の報酬の減衰和であり，その直観的な意味は目標状態に対する現在の状態の良さである．観察者が状態価値関数を学習する際に獲得した方策はその行為の目的を実現する無数の方策の中の一つであるが，得られた状態価値関数はその行為のゴール状態とそれに対する他の状態の良さを表現している．ここで，他の方策に基づき行動した際の状態遷移系列を獲得された状態価値関数によって状態価値を写像すると，値が上昇する傾向にある．例えば，ロボットがボールに近付くというタスクを行なう場合を考える (Fig.3 参照)．ここで，状態 s はロボットの位置座標 (s_1, s_2) で構成されるとする．観察者が獲得した最適な方策による状態遷移系列，すなわち位置座標の系列は path1 のようなものであるのに対し，実演者が示す行為は path2 であったとする．path1 と path2 は状態遷移として比較すると大きく異なるが，Fig.4 で示すように，状態価値関数 $V(s)$ によって状態価値に写像し，さらにその時間変化を見ると，path1 も path2 も状態価値が上昇するという傾向では同じであり，同じ行為として認識が可能になる．

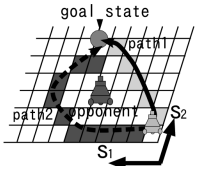


Fig. 3　Sketch of different behavior for one intention

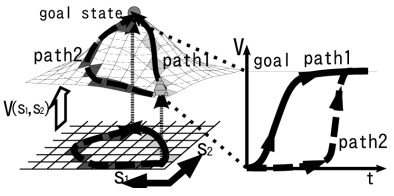


Fig. 4　Observed behavior recognition based on estimation of sequence of state value

2．3　モジュール型行動学習システム

複数の行為を観察・認識・学習・実行するためには，複数の行為の目標・報酬に対応しなくてはならない．ここではモジュール型行動学習システムを導入し (Fig.5)，複数の行為の学習・認識に対応する．提案する手法では観察者は実演者が取り得る行為に対する状態価値関数を知っていなければならない．そのため他者の行為認識を行なう前にあらかじめ各行為の状態価値関数を強化学習によって獲得する．このシステムは，複数の行為モジュール (Behavior Module; BM) を持っており，一つの行為モジュールが一つの行為に対応する．行為モジュールは環境から状態 s を入力として受け取ると，強化学習によって得られた適切な行動を決定し出力する．各行為モジュールが出力した行動は，学習している行為に応じてゲートによって適切に選択された後，ロボットの最終的な行動として出力される．

2．4　行為認識システム

観察者が実演者の行為を認識するシステムを Fig.6 に示す．基本的なシステムの構造は学習時のシステムと同じである．まず観察者は実演者を含んだ環境から

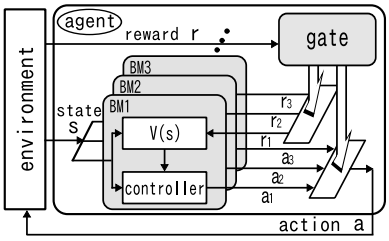


Fig. 5　A modular learning system

状態 s を入力として受け取る．しかし，実際に行動を行なっている実演者とそれを第三者として観測している観察者では視点が異なるため，観測された状態 s から次節で述べる方法で適切に状態を推定する．推定された状態を受け取った行為モジュールは，状態価値関数によって状態を状態価値に写像して出力し，推定された状態価値に基づく行為認識信頼度を出力する．

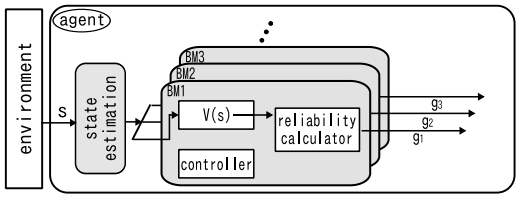


Fig. 6　A system for observed behavior recognition

2．4．1　状態推定器における状態変数の仮定

観察者の持つ状態価値関数を用いて他者の意図を推定する場合，状態価値関数 $V(s)$ が状態 s の関数であることから，本来，実演者の知覚している状態である s^p を得て，自身の状態価値関数に当てはめなくてはならない．しかしながら，観察によって推定した状態 $\hat{s}^p$ を真の状態 s^p の代わりとして用いるが，観察者と実演者の視点の差などからの問題から，s^p と $\hat{s}^p$ を完全に一致させることは困難である．従って，従来手法のように状態遷移で他者の意図を推定する手法では，観測から推定された状態 $\hat{s}^p$ と真の状態 s^p の誤差の影響で，大きく推定を誤る可能性がある．s^p と $\hat{s}^p$ は正確に一致はしないが，状態変数の値の大小関係は保存されていると仮定する．これは，Fig.7 で示すように，推定された状態 $\hat{s}^p$ のある状態変数の値 $\hat{s}^p_1, \hat{s}^p_2, \cdots$ と，真の状態 s^p の対応する状態変数の値 $s^p_1, s^p_2, \cdots$ はその大小関係が維持されるという仮定である．この仮定が成り立つ時，状態価値関数によってそれぞれの状態を状態価値に写像すると，節 2．2 で議論したように状態価値の変化としては同じ傾向となる．つまり，真の観測状態 s^p を写像した状態価値の変化量と，観察から推定された状態 $\hat{s}^p$ を写像した状態価値の変化量が大きくは異ならない．他者行為を推定する際には状態価値の変化が分かればよいので，観察者は，適切な状態変数 $\hat{s}$ を選ぶことで，正確な座標変換等を用いずに他者行為を認識できる．

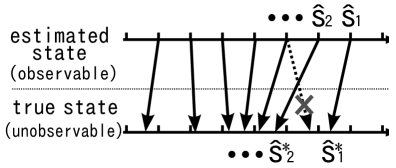


Fig. 7　Order relationship between the estimated states and the true ones

2．4．2　行為認識信頼度

行為認識システムの各行為モジュールは状態価値の勾配から算出した行為認識の信頼度 (reliability) を出力する．行為モジュール i の行為認識信頼度 g_i は

$$g_i = \begin{cases} g_i + \beta & \text{if } V_i(s_t) - V_i(s_{t-1}) > 0 \text{ and } g_i < 1 \\ g_i & \text{if } V(s_t) - V(s_{t-1}) = 0 \\ g_i - \beta & \text{if } V_i(s_t) - V_i(s_{t-1}) < 0 \text{ and } g_i > 0 \end{cases} \quad (2)$$

とする．ここで $V_i(s_t)$ は状態 s_t における行為モジュール i の状態価値を表している．また，β は更新度であり，本論文中の実験では 0.1 としている．行為モジュールの状態価値が増えるほど，信頼度は大きな値を持つ．

3．　サッカーロボットによる他者行動理解

提案した手法をサッカーロボットを用いたタスクにおいて実装する．ロボットが獲得可能な環境からの情報は，ロボットに取り付けられたカメラからの画像情報のみである．想定するタスクの概略図を Fig.8 に示す．環境中には観察者 (observer)，実演者 (performer) とチームメートが 2 体 (teammate1, teammate2)，計 4 体のロボットが存在する．このうち，チームメートは実際には移動せず，パスを受けるためにその場に固定されている．観察者も移動はしないが，実演者の行動を正面のカメラで観察できるようにその場で回転運動のみを行なう．実演者は与えられたタスクに従って行動する．実演者の行為は，

- 青ゴールへの移動 (GoToBlueGoal),
- 黄色ゴールへの移動 (GoToYellowGoal),
- ボールへのアプローチ (GoToBall),
- 青ゴールへのシュート (ShootBlueGoal),
- 黄色ゴールへのシュート (ShootYellowGoal),
- チームメート 1 へのパス (PassToTeammate1),

- チームメート 2 へのパス (PassToTeammate2),

の計 7 種類である．これらの行為のいずれか一つを実演者が行動するのを観察者が観察し，どの行為であるかを識別する．ただし，ある行為を実現するための行動のパターンは観察者と実演者で異なる場合がある．観察者が行動パターンの差を越えて実行者の行為を理解できれば提案手法の有効性が示される．

3．1　実験環境

提案手法の有効性の検証とその分析を行うため，実験はまずコンピュータシミュレーションによって行なった (Fig.9 参照)．このシミュレータはロボカップの中型リーグで使われているロボットを想定したものであり，ロボットは移動機構として全方位移動機構，視覚センサとしてロボット上部に全方位カメラ，正面方向に通常のカメラを備えている．本実験ではロボットの視覚情報源は正面カメラのみとしている．

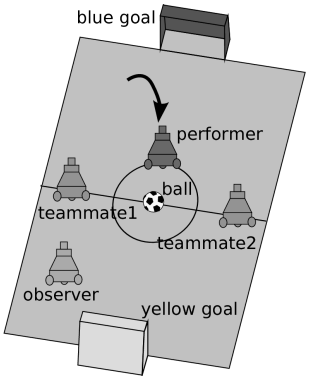


Fig. 8　Sketch of a task

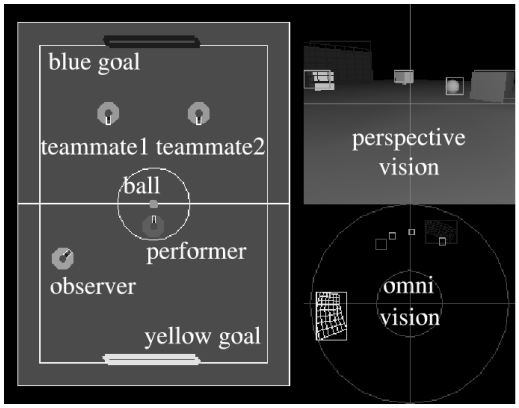


Fig. 9　A simulation environment

3．2　状態価値関数の学習

実演者は 7 種類の行為を持っており，観察者はそれに対応するためにあらかじめ 7 種類の行為モジュールの状態価値関数を獲得しておく．各行為モジュールの学習において用いた状態変数は主に画像上での距離を表現するものと位置関係を表現するものである．正面透視カメラ画像から物体の距離に関した状態を得るために，カメラ画像上での物体の y 座標値を用いた．ただし，ボール，ロボットとゴールにおいては y 座標値の基準とする位置が異なる．すなわち，Fig.10 で示すように，ボールは画像イメージでの上端の座標 (y_u) を代表値とし，ロボットとゴールにおいては画像イメージの下端の座標 (y_l) を代表値とする．この y 座標を適切な分割数により離散化し，状態変数とする．さらに，物体が正面透視カメラの視野内に見えない場合は特例として “lost” 状態と見なす．カメラ画像上で物体が画面右端から消失しても，左端から消失してもここでは同じく”lost” 状態と見なす．

シュート行動 (パス行動) では，自身とボールとゴール (チームメート) の位置関係を状態変数とする．この状態変数は Fig.11 で示すように，ボールを中心とした角度 θ で表される．図で示した点 A は，正面透視カメラ画像の下端中央の点で，観察者自身を表すものである．また，どちらか一方の物体がカメラ画像上から消失した場合，Fig.12 で示す点 B にあると仮定して θ を求める．点 B は，高さは画像の中央，横方向は画像幅の半分外側の点である．ただし，物体が画面の左端から消失した場合は逆の地点にあると考える．また，両方の物体が画像上で見えない場合は，“lost” 状態と見なす．これらの距離に関する状態変数と位置関係に関する状態変数を組み合わせ，各行為モジュールの状態変数とした．**Table 1** にその詳細をまとめる．

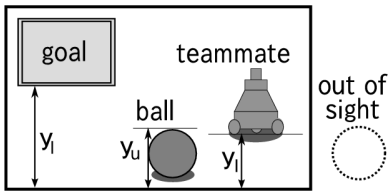


Fig. 10　State variables representing distances to the objects

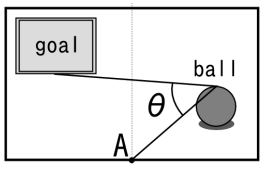


Fig. 11　A state variable θ representing the positional relationship between the objects

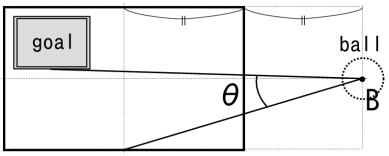


Fig. 12　A state variable θ when one of the objects is out of sight

Table 1　State variables of each behavior module

Behaivor Module	State variables
GoToBall	ボールの距離
GoToBlueGoal	青ゴールの距離
GoToYellowGoal	黄色ゴールの距離
ShootBlueGoal	ボールの距離 青ゴールの距離 ボール，青ゴールの位置関係
ShootYellowGoal	ボールの距離 黄色ゴールの距離 ボール，黄色ゴールの位置関係
PassToTeammate1	ボールの距離 チームメート 1 の距離 ボール，チームメート 1 の位置関係
PassToTeammate2	ボールの距離 チームメート 2 の距離 ボール，チームメート 2 の位置関係

3．3　他者行為の認識

本手法の有効性を示すため，まず実演者の動作系列が観察者の動作系列と全く同じ場合での行為認識性能を検証し，次に動作系列が異なる場合において性能を検証した．

観察者が実演者の行動を観察することで状態変数を得なければならないが，ここで用いる状態変数は，2. 4. 1 節で議論したように，学習の際に用いた状態変数と値の大小関係が一致するように選択する．Fig.13 に，観察者の正面カメラ画像から，物体と実演者の距離に関する状態を得る方法を示す．これは，正面カメラ画像上での物体と実演者の距離を状態変数とするものである．この距離を表す変数は，Fig. 10 で示した自分とオブジェクトまでの距離を表す変数と同じように，近ければ小さな値を，遠ければ大きな値を返す．透視投影カメラの特性上，カメラ画像での距離と実際の距離は完全に一致しないが，ロボットがオブジェクトがある方向に前後する場合，学習で用いた状態変数と推定で使う状態変数の値の大小関係は保たれている．Fig.13 中の d_1 は，実演者とゴールの距離を表現した状態変数，d_2 は，実演者とボールの距離を表現した状態変数である．図には示していないが，チームメートへの距離の取り方も同様である．

Fig.14 に，観察者の正面カメラ画像から，物体の位置関係に関する状態を得る方法を示した．これは正面カメラ画像上でボールを中心としてゴールと実演者の成す角度を状態変数とするものである．この状態変数も，ロボットの移動を実時間で観測している範囲では，学習時の位置関係を表現した状態変数と位相関係を十分に保っている．

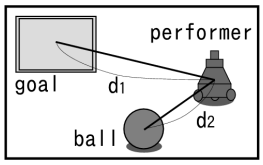


Fig. 13　Estimated state variables representing distances

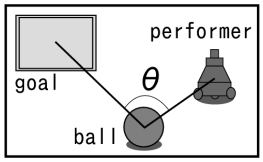
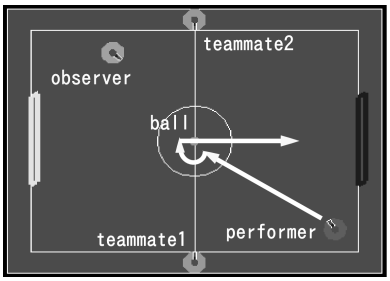


Fig. 14　An estimated state variable θ representing the position relation among objects

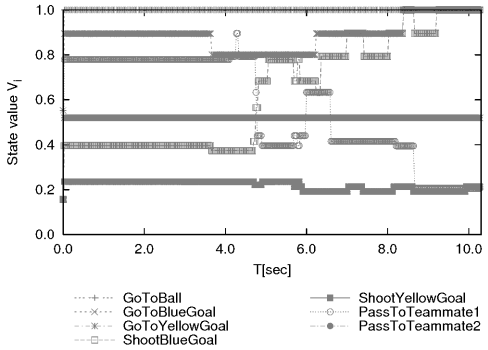
3．3．1　自身と同じ行動系列をとる行為認識実験

7 つの行動パターンに対して行為認識の実験を行なった．Fig.15 と Fig.16 はそれぞれ行為 ShootBlueGoal と PasstoTeammate1 を実演者が行なった際に，提案手法によって行為認識した結果である．ただし，実演者が見せた行為は観察者が行動学習し獲得した動作系列と全く同じものである．

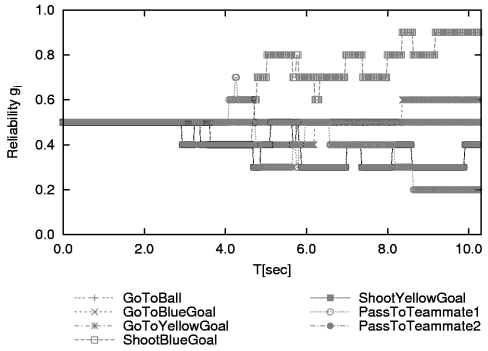
Fig.15(b) は，実演者が青ゴールにシュートする行動を行なった際に，観察者が自身の行為モジュールを用いて推定した状態価値の変化を示している．また，Fig.15(c) はタスクにおける信頼度の変化を示している．(a) を見ると実演者は実際にはボールに近付いているにも関わらず，(b) での GoToBall の状態価値が上がらないのは，観察者の視点からはすでに実演者とボールが最も近い状態と見えていて，ゴール状態と判定され，頭打ちになっているためである．また，約 6.0sec に PassToTeammate1 の状態価値が上がるが，これは観察者の視界に入っていなかったチームメート 2 が突如視野内に入ってきたためであるが，そのあとは状態価値が下がっていく．結果として，(c) において ShootBlueGoal に該当する緑のラインの示す信頼度が徐々に高くなり，実演者の行為が認識できている．



(a) Top view of the behavior

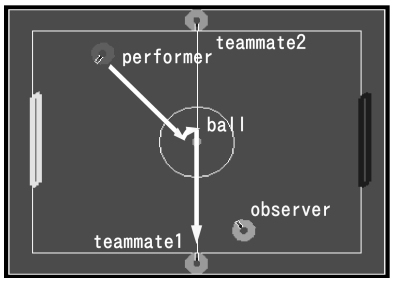


(b) State value V_i of each bahavior module

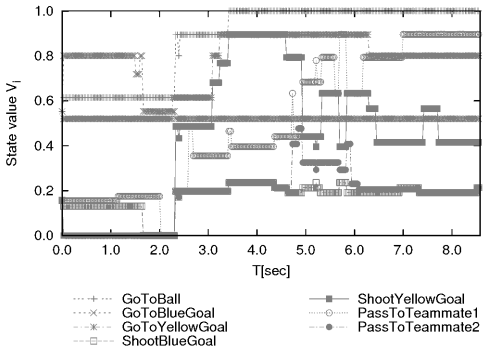


(c) Reliability g_i of each behavior module

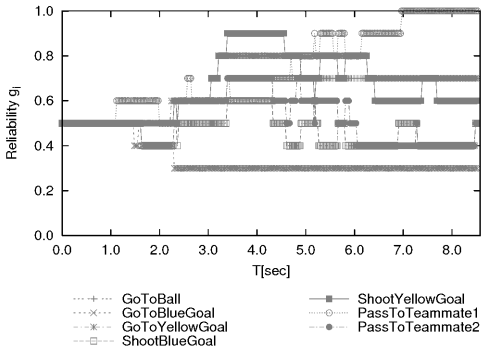
Fig. 15　Behavior recognition of the performer trying to shoot ball to the blue goal



(a) Top view of the behavior



(b) State value V_i of each bahavior module



(c) Reliability g_i of each behavior module

Fig. 16　Behavior recognition of the performer trying to pass a ball to teammate1

Fig.16(b), (c) は，実演者がチームメート 1 にパスする行動を行なった際に，観察者が自身の行為モジュールを用いて推定した状態価値の変化と信頼度の変化を示しており，最終的に正しい推測結果であるモジュールの信頼度が高くなっており，実演者の行為が正しく認識できていると言える．

これまでの実験では，ひとつのタスクに対して観察者は同じ位置から実演者を観察していたが，ここでは観察者の配置をフィールド上で様々な位置に変えて，行為認識を行なった．Fig.17 は実演者が ShootBlueGoal 行動を行なった際に，観察者の観測する位置を様々に変えて行為認識するタスクを行なった．正しく行為認識ができた地点は赤い丸，誤った認識をした地点には緑のバツを記している．同様に，Fig.18 は，実演者が PassToTeammate1 を行なった際の結果である．どちらの結果においても，広い領域で認識が可能となっている．

どちらの結果においても，実演者が最終目標とする物体を背に向ける領域，つまり，ShootBlueGoal では青いゴールの手前付近，PassToTeammate1 ではチームメート 1 の周辺において誤認識が見られる．これは，本来実演者行為の認識のための状態変数を構成する物体が観察者の視野外となることで，その物体に関する推定値の誤差が大きくなってしまい，結果として，状態価値が他の行為に対して相対的に低くなってしまったためであった．また，それ以外の場所でも誤認識が見られるが，それらの場所においても同じように，観察者の視野外にある物体の状態変数の推定誤差

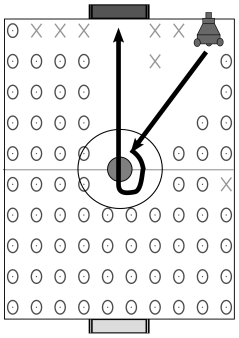


Fig. 17　Behavior recognition by observing the performer from various points while performer is trying to shoot ball to the blue goal

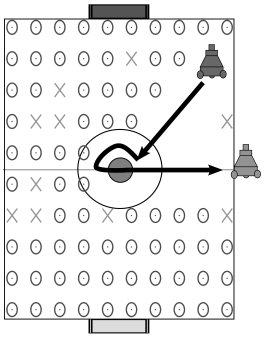
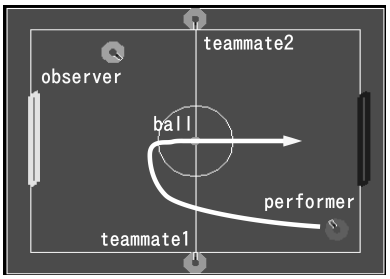


Fig. 18　Behavior recognition by observing the performer from various points while performer is trying to pass a ball to teammate1

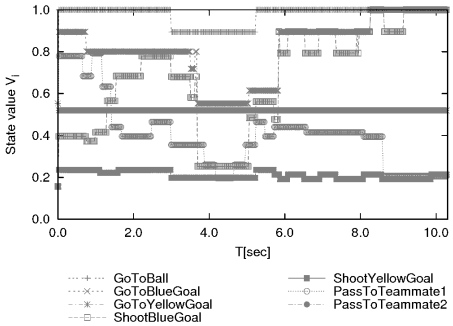
により，逆に状態価値を大きく見積もってしまったためであった．本実験では，観察者の視野から見えない物体の位置に関してはかなり粗い推定をし，この推定の粗さが行為認識の誤認識を引き起こす原因となっている．これは実演者の視野が限られていることによる部分観測問題ではあるが，同じように視野を限られた人間の場合には，視野外にある物体に関してもある程度の予測ができるモデルを持っており，その予測を基に状態を推定していることが考えられ，ロボットにおいてもそのようなモデルを持たせることで誤認識の問題が緩和されると考えられる．

3．3．2　自己と異なる行動系列をとる他者行為認識実験

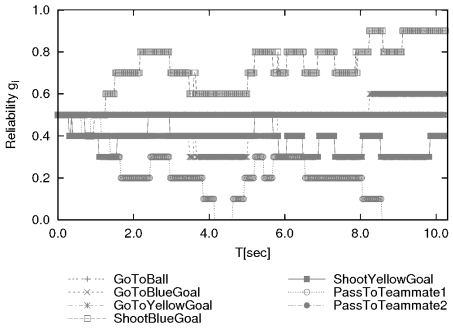
これまでの実験では，実演者が行なった行動は観察者が学習段階で行なった行動と全く同じものであった．これに対し，Fig.19 に示すのは学習者が学習した行動とは異なる行動パターンであり，Fig.15(a) で示した軌跡に比べ Fig.19(a) で示したシュート行動はなめらかな動きで行なわれている．この行動を観察した際



(a) Top view of the behavior



(b) State value V_i of each bahavior module



(c) Reliability g_i of each behavior module

Fig. 19　Behavior recognition of the performer trying to shoot ball to the blue goal

の各モジュールの状態価値と信頼度の値を Fig.19(b), (c) に示す．グラフより，この場合も実際の他者の意図と一致したモジュールの信頼度が高くなっており，他者の意図を適切に推定できていることがわかる．

3．4　三次元再構成を利用した状態遷移に基づく行為認識方法との比較実験

提案手法の有効性を検証するために，3 次元再構成により自己が観測している状態から実演者がが観測している状態を復元し，状態遷移に基づいた行為認識方法との比較実験を行なった．

3．4．1　状態遷移に基づく行為認識法

状態遷移に基づく行為認識方法では，まず，他者の状態を観測し，それを三次元再構成によって座標変換したのちに状態遷移の尤度を計算し，比較する手法である．本実験で用いた三次元再構成の手法では，

Fig.20 に示すように，正面カメラで捉えた画像に透視図を仮定して距離に関する情報を得，また，Fig.21 に示すように，正面カメラで捉えた画像から他者と物体との位置関係に関する情報を得る．これらの情報を併せて他者の状態を推定するものである．

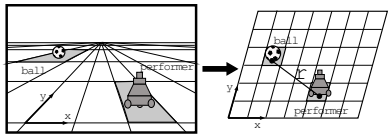


Fig. 20　A coordinate transformation on distance

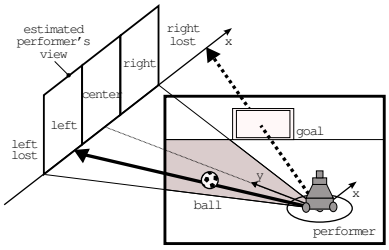


Fig. 21　A coordinate transformation on angle

三次元再構成によって座標変換された他者の状態から，状態遷移の尤度を計算し，得られた状態遷移系列に対し，尤度 θ は

$$\theta = \prod_{t=-T+1}^{0} \exp(p_t) \quad (3)$$

$$p_t = \Pr\{s_t = s' | s_{t-1} = s, a_{t-1} = a\} \quad (4)$$

で計算し，その値の最も大きいものを他者の意図と見なす．table ここでは $T = 5$ と設定している．

3．4．2　実験結果

実験では，観察者と実演者の初期配置をランダムに変えて 500 試行の行為認識のタスクを行ない，正しく意図認識できた割合によってそれぞれの手法を評価した．その結果を **Table 2** に示す．ここで，ShootBlueGoal2 とは，Fig.19(a) に示す行動であり，ShootBlueGoal3 とは，動きの軌跡は Fig.15(a) に示す行動と同じであるが，常に青いゴールを正面に見るように移動する行動である．

Table 2 から，状態遷移に基づく方法の認識率が極めて低いことがわかる．これは，三次元再構成の誤差が行為認識に大きく影響することを示している．三次元再構成の手法は，環境，タスク，他者の機構等，多くのパラメータなどの事前知識を必要とする手法であるが，それらの値が全て正確に得られないと状態遷移によって他者の意図を推定をすることが困難であると言える．これは，わずかな誤差であっても，観察者の知る行為の状態遷移と比べると，尤度が低くなってしまうことが原因である．これに対し，提案手法では約 8 割の確率で正しく行為認識が行なえている．

Table 2　Performances of Behavior Recognition

	提案手法	状態遷移に基づく方法
ShootBlueGoal	84%	24%
ShootBlueGoal2	82%	25%
ShootBlueGoal3	78%	11%
ShootYellowGoal	86%	20%
PassToTeammate1	76%	34%

また，ここでは同じ行為に対して観察者と異なる行動系列を実演者がとる場合の行為認識も行なっている．状態遷移に基づく方法では，ShootBlueGoal と ShootBlueGoal2 の認識率はそれほど大きく変わらないが，ShootBlueGoal3 で認識率がさらに大きく落ちている．これは，ShootBlueGoal と ShootBlueGoal2 は運動の軌跡は異なるものの，状態の遷移としてはほとんど同じであるのに対し，ShootBlueGoal3 に関しては，常に青ゴールの方向を向いて移動する行動であり，観察者の知るシュート行動の状態遷移と大きく異なるのが原因であった．これに対し，提案手法では全てのシュート行動において認識率があまり変動しない．2．2 で述べたように，状態遷移として比較すると大きく異なるが Fig.4 で示すように，状態価値関数によって状態価値に写像し，さらにその時間変化を見ることで，異なる状態遷移を行う同じ行為を統一的に認識できていることを示している．

4．　おわりに

従来の多くの研究では状態遷移系列によって行為認識を行なっていた．しかし，現実問題として，同じ意図を達成するために複数の行動パターンがあり，従来手法では，一つの行為に対して一つの適切な実現パターンを当てはめることしかできず，同じ行為であるはずの別な行動系列を別の行為と見なしてしまうという問題があった．

この問題を解決するために，本論文では他者行為の認識に，状態価値を用いる手法を提案した．状態価値であれば状態遷移系列によらず，状態価値が高くなるものがその行為であると広く捉えることができる．この手法の有効性を示すための具体的なタスクとして

ロボットのサッカーを例題とし，他者の行為認識をシミュレーションによって検証した．その結果，本手法によって観察して得た情報のみから座標変換を用いずに，視点の差を越えた広い領域において行為認識が可能であることを示した．また，同じ行為であるが実現パターンの異なる行動を，他者が実行するのを観察し，適切な行為として推定ができることを示した．本論文で提案した手法は完全知覚問題におけるマルコフ過程を仮定している．実環境下のロボットのセンサには精度やオクルージョン等の制約があり，不完全知覚問題を扱わなくてはならない．不完全知覚問題に対する提案手法の有効性や拡張等が今後の課題である．

参考文献

1) Jonathan H. Connell and Sridhar Mahadevan. *ROBOT LEARNING*. Kluwer Academic Publishers, 1993.

2) V. Gallese and A. Goldman. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, Vol. 2, No. 12, pp. 493–501, 12 1998.

3) Tetsunari Inamura, Yoshihiko Nakamura, and Iwaki Toshima. Embodied symbol emergence based on mimesis theory. *International Journal of Robotics Research*, Vol. 23, No. 4, pp. 363–377, 2004.

4) Noda Itsuki. Hierarchical hidden markov modeling for team-play in multiple agents. In *Proc. of IEEE Conf. on System, Man and Cybernetics 2003*, pp. 38–45, 8 2003.

5) Marc Jeannerod. Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage*, Vol. 14, pp. S103–S109, 2001.

6) Doya K., Sugimoto N., Wolpert D.M., and Kawato M. Selecting optimal behaviors based on contexts. In *International Symposium on Emergent Mechanisms of Communication*, pp. 19–23, 2003.

7) Y. Nagayuki, S. Ishii, and K. Doya. Multi-agent reinforcement learning: An approach based on the other agent's internal model. In *Proc. of the International Conference on Multi-Agent Systems*, 2000.

8) Erhan Oztop, Mitsuo Kawato, and Michael Arbib. Mirror neurons and imitation: a computationally guided review. *Neural Networks*, Vol. 19, No. 3, pp. 254–271, Apr 2006.

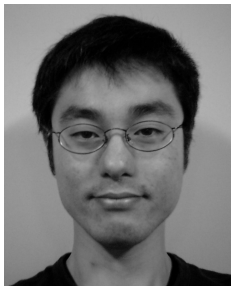
9) Tohyama S., Omori T., Oka N., and Morikawa K. Identification and learning of other's action strategies in cooperative task. In *Proc. of 8-th International Conference on Artificial Life and Robotics(AROB8th'03)*, pp. 40–43, 2003.

10) R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.

11) 鮫島和行, 杉本徳和. モジュール強化学習と意図. 人工知能学会誌, Vol. 20, No. 4, pp. 441–448, 7 2005.

12) 福田敏男, 山本修平, 関山浩介. 他者評価を用いた強化学習による合理的集団行動の獲得. 第 15 回インテリジェントシステムシンポジウム, pp. 391–396, 9 2005.

著者紹介



高橋 泰岳 （たかはし　やすたけ）[正会員]

1994 年大阪大学大学院工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．現在大阪大学大学院工学研究科知能・機能創成工学専攻助教．この間 2006 年 6 月より 2007 年 9 月までドイツ Fraunhofer IAIS 客員研究員．ロボカップ中型機リーグや知能ロボットの行動獲得に関する研究に従事．人工知能学会，日本ロボット学会，知能情報ファジィ学会などの会員．

河又 輝泰 （かわまた　てるやす）[非会員]

2004 年大阪大学工学部応用理工学科卒業．2006 年大阪大学大学院工学研究科知能・機能創成工学専攻修了．2006 年松下電器産業株式会社 (現パナソニック株式会社) 入社．現在，パナソニックモバイルコミュニケーションズ株式会社，技術開発センターで，携帯電話端末開発業務に携わる．

浅田 稔 （あさだ　みのる）[非会員]

1982 年大阪大学大学院基礎工学研究科後期課程修了．1995 年大阪大学工学部教授．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．2005 年より JST ERATO 浅田共創知能システムプロジェクト研究総括．認知発達ロボティクスの研究に従事．本学会論文賞 (1996)，文部科学大臣賞 (2001) など受賞多数．