自己の状態価値に基づく他者の意図推定

Inferring other's intention based on estimated state value of self

○ 河又 輝泰 (阪大院)　　正 高橋 泰岳 (阪大, 阪大 FRC)
田村 豊武 (阪大)　　正 浅田 稔 (阪大, 阪大 FRC)

Teruyasu KAWAMATA, Osaka University, 2-1, Yamadaoka, Suita, Osaka
Yasutake TAKAHASHI, HANDAI Frontier Research Center, Osaka University
Tom TAMURA, Osaka University, 2-1, Yamadaoka, Suita, Osaka
Minoru ASADA, HANDAI Frontier Research Center, Osaka University

Inferring intention of other agents is one of the most imortant issues in realizing cooperative behaviors in multiagent environment. This paper proposes a method where an observer infers the intention of other agent (performer) from its own estimated state values instead of conventional three-dimensional reconstruction of other agent's behaviors. This is because conventional methods often require precise knowledge of the camera parameters and the environment. Furthermore, the difference of the viewpoints between them may disturb the observer in its inferring intention of the performer even though the both take the same behavior in the physically same situation. However, the tendency of the estimated state value changes can be same if the both attempt to accomplish the same task. This inference method requires two steps. At first, the observer learns several behaiviors to accomplish various kinds of task, each of which may correspond to the other's intention to do it. Second, the observer estimates the variation of the state value of the performer through observation and figures out the behavior that has the same tendency of the value changes. The experiments results in computer simulation and the real robot tasks are shown.

Key Words: intention inference, state value, reinforcement learning

1 はじめに

近年ではロボットに関する研究や開発が多く成されており，今後は人間の生活に近い状況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，ロボットに人間と協調して行動することの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行なわれている [2)3)5)]．人間，あるいは他のロボットと協調的な行動をとるためには他者の行動予測が非常に重要である．他者の行動を適切に予測可能であると，その行動に対して協調行動をとれるように自身の行動を決定できるためである．行動は意図に基づき目標状態を達成するように生成されるため，他者の意図を理解することは他者の行動予測の指針となる．

他者の行動認識や行動予測を扱った研究として，Inamura et al.[1)] 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．運動認識においては，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 を他者の行動として認識をしている．また，鮫島ら [4)] は強化学習を応用した MOSAIC によって，観察から他者の行動の意図推定を実現している．この手法では，観察者は複数の予測器を持った行動モジュールを有し，他者の行動を観察したとき，自身のそれぞれの行動モジュールの予測と比較し，もっとも誤差の小さい予測をした学習モジュールの意図を他者の行動のそれとみなす手法である．

しかし，これらの従来手法では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ので，同じ意図であるはずの状態遷移系列であっても，予測器が最適とする状態遷移系列と異なると，同一の意図として認識することはできない．これは相手の意図を自身の唯一の行為としての状態遷移系列によって捕らえてしまうことの欠点である．また，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，状態認識に大きな誤差が生じる．鮫島らや従来の手法のように，他者の行動意図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認識が，意図推定に大きく影響を及ぼす可能性がある．このような問題に対し，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，そのような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．

このような問題に対して，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る．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と目標状態に向かうことになり，状態価値が向上する．つまり，同じ意図を持って行動している限り，行為が異なっても状態価値の値は概ね増加していく傾向にあるので，この傾向から他者の意図を推定できる．また，明示的な三次元再構成による座標変換を行うのではなく，相対情報を保持する状態変数を利用することにより，状態価値の変化情報を保持することでよりロバストな認識を実現する．本論文ではサッカーロボットのシミュレータ及び実機において意図推定の実験を行ない，提案手法の有効性を示す．

2 提案手法

本研究で提案する手法は強化学習における状態価値を用いたものであり，これを基にして他者の意図を認識する．ここでは，状態価値について簡単に説明した後，意図認識の手法について記す．

2.1 状態価値

Fig.1 は強化学習の基本概念を示した図である．まず，エージェントは環境の状態 s_t と方策 π に基づいて行動 a_t を決定する．環境は現在の状態 s_t と行動 a_t によって，次

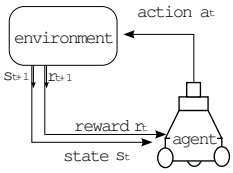


Fig.1 A basic model of agent-environment interaction

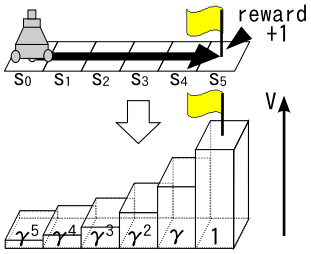


Fig.2 Sketch of state value propagation

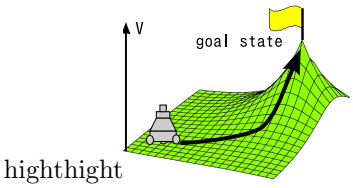


Fig.3 Sketch of a state value function

状態 s_{t+1} に遷移し，エージェントは報酬 r_{t+1} を受け取る．強化学習におけるエージェントの目標はエージェントが最終的に受け取ることのできる報酬を最大化することであり，そのために，方策 π を更新し改善する．

方策を更新するための指針として，ある状態 s を起点として将来期待される報酬量である状態価値が用いられる．状態価値 $V(s)$ は次のように計算される．

$$V(s) = \sum_{t=0}^{\infty} \gamma^t r_t \tag{1}$$

ここで，γ は減衰率である．Fig.2 のように，s_0 から状態遷移を繰り返して s_5 へ遷移し，s_5 で報酬 $+1$ を受け取ると状態価値はゴール状態である s_5 で最も高く，その状態から離れるに従って値が小さくなっていく．さらに学習を繰り返すと，Fig.3 のように状態価値はゴール状態を頂点とした山のような形状になる．

2.2 意図認識の基本概念

例えば，Fig.4 に示すようなボールに近付くタスクを考える．ここで，状態 $\boldsymbol{s}$ はエージェントの位置座標 (s_1, s_2) で構成されるとし，報酬はボールのある状態に達した時にのみ正の値が与えられるものとする．このような目標を達成するための方策は，path1 や path2 のような経路をとるもの以外にも多数存在する．他者が行動するのを観察者が観察し，その状態遷移から他者の意図を推定するためには，そのような状態遷移を生成し得る全ての方策を知らなくてはならない．

これに対して，Fig.5 に示すようにそれぞれの状態を状態価値に写像しその遷移を見ると，どのような方策によって生成される状態遷移系列であっても状態価値が上昇するという傾向は同じである．状態価値はあるゴール状態へ向かう程値が大きくなる性質を持っているので，他者の行動を観察して，その状態価値が大きくなることによって他者の意図を認識することができる．

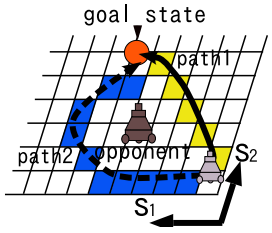


Fig.4 Sketch of different behaviors for one intention

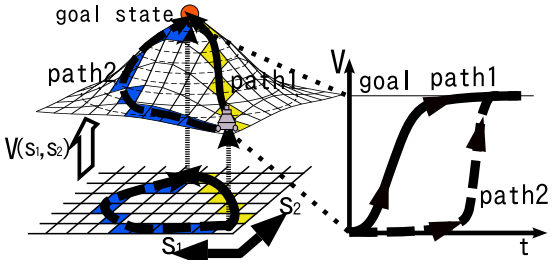


Fig.5 Method for inferring intention by the change of state value

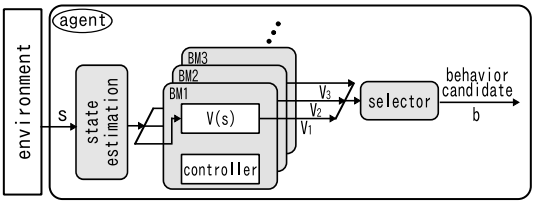


Fig.6 A system for inferring other's intention

2.3 意図識別の手法

観察者が他者の行動意図を推定するシステムを Fig.6 に示す．まず，観察者は他者を含んだ環境から状態 s を入力として受け取り，他者と物体との相対的な距離や角度などの状態を粗く推定する．エージェントは複数の行動モジュール (Behavior Module; BM) を持っており，各々が別な行動の意図に対応する．これらの行動モジュールは推定された状態を受け取り，各々の状態価値関数によって状態を状態価値に写像して出力する．各行動モジュールが出力した状態価値は選択器 (selector) に渡され，状態価値の時間変化を見て，状態価値が大きくなるものを実演者の行動意図として出力する．

ここで，尤もらしい他者の意図を識別するための指標として，以下に示す信頼度 g_i を用いた．

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

ここで $V_i(s_t)$ は状態 s_t における行動モジュール 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．行動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 実験設定

3.1 実験環境

Fig.7 は実験環境の概観を示している．観察者 (observer) は実演者 (performer) が行動するのを観察し，その行動の意図が何であるかを識別する．実験はコンピュータシミュレーションと実機によって行なった．Fig.8(a) は本実験で用いたシミュレータの概観であり，Fig.8(b) は実機の実験で用いたロボットの概観である．ロボットはロボ

カップの中型リーグで使われているものであり，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．本研究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．

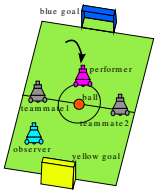


Fig.7 Sketch of the experimental environment

3.2 行動と状態変数

実験で実演者がとり得る行動の候補とその状態変数を **Table 1** に示す．意図推定の実験では，観察者はこれらの行動に対する状態価値をあらかじめ学習によって獲得済であるとする．

状態変数の取り方の概略図を Fig.9 に示す．Fig.9(c),(d) に示す実演者の状態推定に用いられる状態変数は，Fig.9(a),(b) に示す実際の状態変数とは正確には一致しないが，相対的な関係が粗く推定できるように設定している．

4 実験

4.1 既知の行為に対する意図推定

この実験では，実演者は **Table 1** の中のいずれかの行動をとり，観察者はその行動が何であるかを識別する．ここでの実演者の行為は，観察者が状態価値を学習する際に用いた行為と全く同じものである．Fig.10 は実演者が青ゴールにシュートする行動を行なった際の意図識別の実験を示している．Fig.10(a) は実際の実演者の行動の軌跡を示し，Fig.10(b) はその際の信頼度の変化をグラフに示した．青ゴールにシュートする行動の信頼度を表す緑の線が最も高くなっており，観察者は実演者の意図を正しく識別できていることが示されている．

Fig.11 は実機において実演者が青ゴールにシュートする行動を行った際の意図識別の実験を示している．シミュレーション結果同様に，青ゴールにシュートする行動の信頼度を表す緑の線が最も高くなっており，実機においても観察者は実演者の意図を正しく識別できていることが示されている．

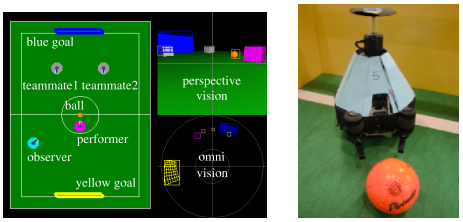
同様に，Fig.12 は実演者がチームメート 1 にパスする行動を行なった際の意図識別の実験を示している．Fig.12(b) で PassToTeammate1 の信頼度が最も高くなっており，正しく意図を識別できていることを示している．

4.2 未知の行為に対する意図推定

前節で実演者の行為は観察者にとって既知の行為であったが，ここでは観察者にとっては未知である異なる行為によって他者が行動する．実演者は Fig.10(a) で示すようなぎこちない動きではなく，Fig.13(a) で示すようになめらかな動きで青ゴールにシュートを行なう．Fig.13(b) の信頼度のグラフにおいて ShootBlueGoal の値が最も高くなっており，観察者にとって未知の行為に対しても適切に意図が識別できていることが示されている．

4.3 三次元再構成による手法との比較

この実験では，観察者が正面カメラから得た画像を基に三次元再構成によって実演者の状態を復元し，その状態遷移の尤度の高さから意図を推定する手法と提案手法において，意図の識別率を比較した．その結果が **Table 2** である．状態遷移に基づく方法に比べて，提案手法ではより正しく意図推定が行なえていることがわかる．



(a) (b)

Fig.8 Experiment environment (a)A simulation environment (b)A real robot

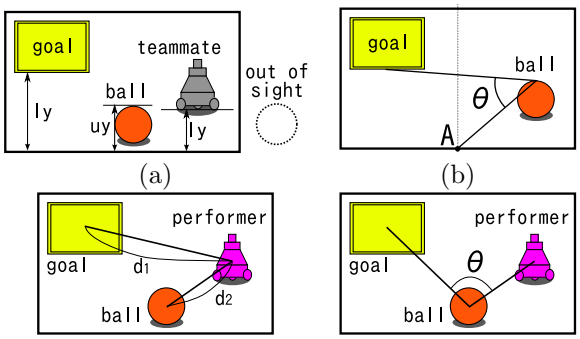


Fig.9 State variables used for learning and estimation. (a)State variables representing distances to objects. (b)A state variable θ representing the positional relationship between objects. (c)Estimated state variables representing distances. (d)An estimated state variables θ representing the positional relation among objects.

5 おわりに

状態価値による意図推定手法を提案した．また，シミュレーション及び実機における実験において提案手法の有効性を示した．

参考文献

[1] 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.

[2] 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.

[3] 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.

[4] 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.

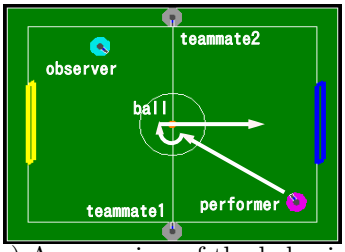
[5] 福田敏男, 山本修平, 関山浩介. 他者評価を用いた強化学習による合理的集団行動の獲得. 第 15 回インテリジェントシステムシンポジウム, pp. 391–396, 9 2005.

Table 2 Results of inferring intention

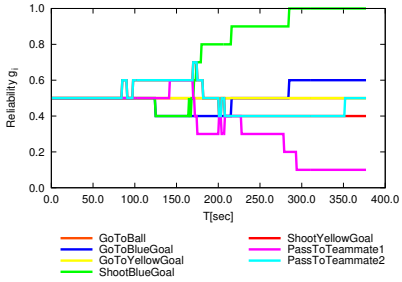
	提案手法	状態遷移に基づく方法
ShootBlueGoal	84%	24%
ShootBlueGoal2	82%	25%
ShootBlueGoal3	78%	11%
ShootYellowGoal	86%	20%
PassToTeammate1	76%	34%

Table 1 Behavior Modules and state variables

Module	State variables
GoToBall	ball position y on the image of perspective camera
GoToBlue	blue goal position y on the image of perspective camera
GoToYellow	yellow goal position y on the image of perspective camera
ShootBlue	ball position y, blue goal position y, and angle between them θ on the image
ShootYellow	ball position y, yellow goal position y, and angle between them θ on the image
PassToTeammate1	ball position y, teammate 1 position y, and angle between them θ on the image
PassToTeammate2	ball position y, teammate 2 position y, and angle between them θ on the image

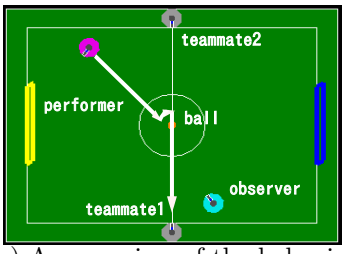


(a) An overview of the behavior

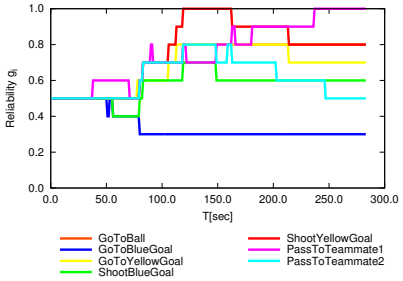


(b) Reliability g_i of each behavior module

Fig.10 Result of inferring intention of the performer trying to shoot ball to the blue goal



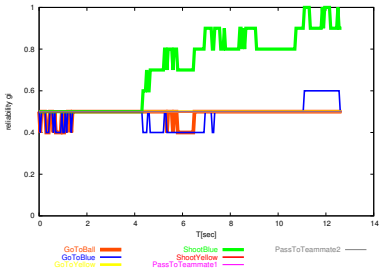
(a) An overview of the behavior



(b) Reliability g_i of each behavior module

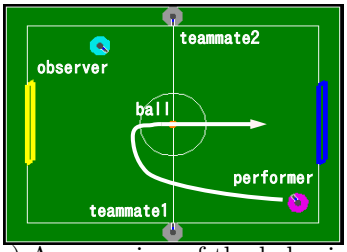
Fig.12 Result of inferring intention of the performer trying to pass a ball to teammate1 with real robots

(a) An overview of the behavior

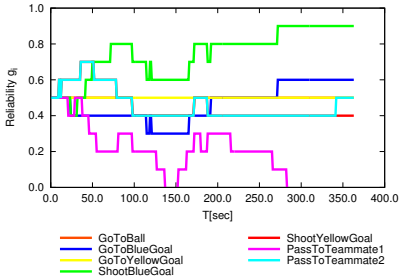


(b) Reliability g_i of each behavior module

Fig.11 Result of inferring intention of the performer trying to shoot ball to the blue goal



(a) An overview of the behavior



(b) Reliability g_i of each behavior module

Fig.13 Result of inferring intention of the performer trying to shoot ball to the blue goal