

| |
|-----|
| 解 説 |
|-----|

ロボカップによる認知ロボティクス

Cognitive Robotics in RoboCup Initiative

浅田 稔* 大阪大学大学院工学研究科

Minoru Asada** Graduate School of Engineering, Osaka University

1. はじめに

昨年7月パリで開催された第2回ロボカップ国際大会では、世界22カ国から約70チームが参加し、シミュレーション(約40チーム)、実機小型(12チーム)、実機中型(18チーム)の3つのリーグで熱戦が繰り広げられた。ロボカップの目的は、ロボティクスと人工知能における新たな標準問題を設定し、多くの研究者が動的マルチエージェント環境における協調・競合問題の解決に当たること、知的人工システムの様々な諸問題を解決することである[1]。

筆者のチームは、第1回の名古屋大会から参加し、特徴として学習機能を取り上げ、主に強化学習法を用いてロボットの行動を獲得してきた。通常「学習」の特徴としてあげられる

- (1) 容易なプログラミング
- (2) より少ない事前知識
- (3) 汎用性

は、理論的な学習アルゴリズムが理想的な環境で適用される際には、優れた特徴と写るが、実際の環境に適用していく上では、必ずしもそうとは限らない。我々の研究目標は、このような表層的な学習の特徴の実現より、むしろ自律エージェントが環境との相互作用を通して、世界をどのように表現し行動を獲得していくかといった、ロボットの認知過程に焦点を当ててきた。特に、環境因子として他のエージェントの行動が自分の行動をどのように規定していくかという過程の中に、ロボットが「自我」を見出していく道筋が解釈できるのではないかという期待がある。

本稿では、我々のここ5年間の研究の推移を認知ロボティクスの観点から見直す。最初に他者の存在の認知に関する議論を展開する。次にそれぞれのアプローチについて、筆者が「展望」[2]でのべた二つの設計論(内部/外部構造の設

計論と経時的発展のための設計論)の立場からそれらの意義を明確にする。最後の今後の課題を述べ、まとめる。

2. 他者の存在の認知

ネズミの発達の実験で、広い環境で比較的多くのネズミと接しながら成長する場合と、狭い環境で比較的少数のネズミとしか生活を共にしないネズミの成長との比較[3]で、明らかに前者が後者より脳の発達が活発で、行動の多様性が見られたと報告されている。この事は、環境の多様性が脳の発達に多大な影響を与えている例である。知能の本質に関する議論として、社会性、他者の存在、模倣などがキーワードとして挙げられている[4][5]。特に大澤は、自己を否定し得る可能性のあるものやことを「他者性」とし、未来性も含めており、時間との関連性を印象づける。ロボットが環境との相互作用を通じて知的な行動を獲得していく過程においても、環境の多様性が獲得される行動の多様性や複雑度に関連する可能性は十分にあると考えられる。つまり、知能のレベルが、環境との相互作用の複雑性にあるとすると、視覚を基にした環境との相互作用をベースとする我々の研究でも、他者の存在が、複雑さを増すであろうことや、自己の規定などに関する議論が可能と考えられる。

- (1) 自己の身体：運動機能を有する主体が、運動指令と視覚情報との直接的な相関がとれる場合(例えば、他の多くの人たちと手のひらを重ねあわせた状況を観測していて、自分の手を発見するとき、自分の手をちょっと動かすだけで分かる。)すなわち、運動指令を発した自己の運動結果が視覚で確認され、大澤の言を借りれば、運動指令を出した自己を裏切らない範囲が自己と規定し得る範囲であろう。
- (2) 静止環境：自己の運動との直接的な相関により、切り出し可能(静止環境に対するオプティカルフローは、視線を変えた方向と逆方向に流れる。)切り出せば制御可能。全ての運動は相対的なので、「静止環境」と「自己の身体」との区別は容易ではないが、重力方向などによる他の参照座標系との関係から、識別可能と

原稿受付 1998年9月4日

キーワード：ロボカップ、やさしいタスクからの学習、発達、認知、自我

*〒565-0871 吹田市山田丘2-1

*2-1 Yamadaoka, Suita, Osaka 565-0871

なる。

- (3) 受動エージェント：自己や他者の運動の帰結により、静止環境内で運動したり、停止したりするもの。自らは状態変化を起こさないもの。ロボカップでは、ボールが相当する。
- (4) 他の能動エージェント：直接的には、自己の運動との単純な相関を持たない。大澤の意味では、自己を否定する可能性のあるもの。視覚情報からの相関では、単純な相関を持たないものとして、最初、排他的に扱われるが、自己の運動との高度な相関(協調、競合、干渉)により、規定される対象。行動選択の複雑さは一挙に増す。

以上の議論を踏まえながら、これまでの我々のアプローチの検証を試みる。

3. 強化学習の単純な応用

視覚情報をもとにシュート行動を獲得するために、最初に行った強化学習実験 [7]。強化学習の手法として離散化された状態空間と行動空間を必要とする 1 ステップ Q 学習を用いた。対象はボールとゴールでタスクはボールをゴールにシュートすること。

- 内部構造：Q 学習モデル
- 環境設計：LEM (Learning from Easy Missions) 学習を加速するために容易な状況(ボールと学習ロボットをゴール近くに設置)から徐々に困難な状況へ学習をスケジューリング。経時的発展は設計者のスケジューリング。
- シミュレーションと実機の関係：シミュレーションで獲得した政策を実機で再現。
- 達成点：強化学習の視覚ロボットへの適用。状態空間は、視覚情報に基づいてプログラマーが設計。行動空間は当初、前進や転回などのモーター指令としていたが、実環境では 1 回の行動が必ずしも状態遷移を引き起こすと限らず、学習が進まなかった。そこで、状態遷移を引き起こすまで同一のモーター指令を続け、その一連の指令を一つの行動として再定義し、行動空間とした。
- 課題：プログラマーが設計した状態空間がロボットにとって最適となっている保証がない。

4. 状態空間の自律的構成(その1)

外界の様子をどのように記述するかは、ロボット自身が環境との相互作用の中で決定すべき事項であることを踏まえ、成功事例をもとに、センサー情報から状態空間を自律的に分割する手法を提案し、上と同じタスクに適用 [6]。

- 内部構造：成功事例のクラスタリングの後 Q 学習
- 環境設計：事例をオフラインで収集するフェーズと作成した状態空間に基づく Q 学習フェーズの 2 段階の学

習環境。前者で LEM (Learning from Easy Missions) 学習を加速するために容易な状況(ボールとゴールがほぼ正面に見える位置)から徐々に困難な状況(ボールとゴールが画面上で左右に分かれている場合名)へ学習をスケジューリング。経時的発展は設計者のスケジューリング。

- シミュレーションと実機の関係：シミュレーションと実機で別の実験。
- 達成点：センサー情報(ボールの位置、大きさ、ゴールの位置、大きさ、向き)で構成される状態空間をタスクに応じて自律的に分割。見かけのセンサー情報が大きく異なっても、ゴールへ到達するモーター指令が同一なら同じ状態とみなす「行動に基づく状態空間の構成」。
- 課題：なるべき多くのデータを一様に収集する必要あり。オフラインなので環境の変化(例えば、ボールの大きさが倍になるなど)に対応できず。

5. 状態空間の自律的分割(その2)

オンラインでセンサー入力、状態空間の構成、強化学習のループを永遠に続ける手法の提案と実験 [8]。タスクは上と同じ

- 内部構造：状態と行動による状態変化の関数近似による状態空間の構成とそれによる Q 学習の無限ループ。近似誤差が状態分割や統合のトリガー
- 環境設計：最初の実験と同じ。オンライン学習による経時的発展の内部観測可能。
- シミュレーションと実機の関係：シミュレーションと実機で別の実験。
- 達成点：ボールとゴールの画像特徴の主成分解析により、必要な情報の選択。環境の動的変化(例えば、ボールの大きさを突然倍にするなど)に対応可能。このときそれまでの学習結果を効率的再利用。
- 課題：オンラインなので、記憶無し。以上三つに共通するのが、より動的で複雑な環境への適用がないこと。

6. ゴールキーパーとの競合行動

ゴールキーパーを含めた環境でのシュート行動の獲得では、個別に Q 学習で獲得された複数の行動を統合する 3 つの手法(単純和、サブサンプリング、再学習)の比較 [9]。

- 内部構造：個別行動の統合のための再 Q 学習
- 環境設計：サブタスクへの分割と、再学習領域への高頻度設置、更に、ゴールキーパーの行動を制御(低速から高速へ)することにより学習を加速化。経時的発展は設計者のスケジューリング。
- シミュレーションと実機の関係：シミュレーションで獲得された政策の実機での再現。

- 達成点：単純和やサブサンクションでは対処できない統合による非統一の状態の発見及び、再学習による対処。
- 課題：再学習時間が長いこと。モジュール構造による加速化あり [10]。動的変化への対応が不十分。

7. 複数ロボット環境下での行動学習

これまで述べてきた手法では、他の能動エージェントが存在しない簡単なタスク（ボールをシュートするのみ）で、状態空間を自律的に構成できたが、味方や敵がいる環境には、次元が極端に増加し、直接適用することが困難である。基本的な問題点は、マルチエージェント環境では、学習者が一人であったとしても、他のエージェントの行動政策が未知の場合、容易に状態空間を構成できないことである。そこで、学習者の観測と行動を通して、学習者と他者の行動の関係を局所予測モデルとして推定し、その結果をもとに強化学習をおこなうことを考えた。これを簡単なパスとシュートの協調行動に適用した [11]。

7.1 アーキテクチャ

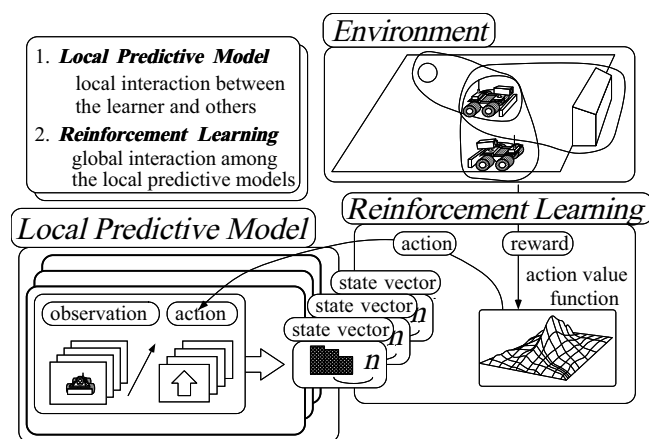


図1 提案するアーキテクチャ

図1は各ロボットに与えられる行動獲得のためのアーキテクチャである。はじめに、学習者はセンサ情報だけでなく、学習者自身の行動のシーケンスも考慮して局所予測モデルを構築する。ここでのポイントは、観測のみによって対象のモデルを推定するのではなく、自分の行動との関わりを通じて、対象の動きを予測することである。この結果、自分との関わり（味方（協調）、敵（競争）、審判（干渉?））が判別可能となる。次に推定された状態ベクトルをもとに、協調行動獲得のための学習を開始する。

7.2 学習スケジューリングと実験結果

最初にシューターとパスナーは、ボール、ゴール、そして互いの局所予測モデルを、計算機のシミュレーション上で構築する。次に、シューターを静止させた状況下で、パッ

サーは行動の学習を実機で開始する。パスナーの学習が終了した時点で、パスナーの行動政策を固定し、シューターの学習を開始する。パスナーは、ボールをシューターにパスしたときに報酬 1 を受け取り、シューターはボールをゴールにシュートしたときに、報酬 1 を受け取る。さらに、ロボット間で衝突が発生した場合、-0.3 の報酬が与えられる。このようにして、シミュレーションと実機の実験を結びつけ、実機での学習の効率化を図った。

- 内部構造：正準変量解析による状態ベクトル推定と構成された状態空間による Q 学習
- 環境設計：各エージェントのタスクの割り当て。状態ベクトル推定のための多様な行動の再現。学習者と非学習者の交代スケジューリング。シミュレーションと実機との結合のスケジューリング、これまでの学習高速化技法の流用。経時的発展は設計者の全体スケジューリング。
- シミュレーションと実機の関係：シミュレーションで獲得された政策をもとに実機での状態ベクトル推定。有機的結合。
- 達成点：マルチエージェント環境に強化学習を適用するための状態ベクトル推定手法の提案と実現。
- 課題：オフラインの状態推定フェーズが必要。より多くのエージェントが存在する場合、エージェント間の相互作用の複雑さに後段の Q 学習で対応困難、更に、学習者が一人なので学習時間の増大。

7.3 もう一つの GK との対峙

ゴールキーパを対象として、シュータがシュート行動を獲得する過程に、ゴールキーパの行動を制御して学習を高速化する際、7と同じ手法を用いて、シュータの状態ベクトルを推定し、最初から状態ベクトルをすべてはるのではなく、主要な成分から順次加えていくが、そのタイミングを制御した [10]。

- 内部構造：7と同じ。但し、環境設計に同期して学習に利用する状態ベクトルを詳細化。
- 環境設計：ゴールキーパの速度を低速から高速に変化させるタイミングと状態ベクトルを詳細化するタイミングを制御して学習を加速。経時的発展は設計者の全体スケジューリング。
- 達成点：制御しない場合と比較して、学習時間の短縮（同じパフォーマンス到達時 1/10）とパフォーマンスの向上。
- 課題：基本的にオフライン推定なので、最も詳細な状態ベクトルを最初の時点で推定する必要がある。本来、オンラインでベクトルを詳細化したい。

7.4 共進化による協調行動の創発

複数ロボットの同時学習手法の一つとして、遺伝的プログラミングによる協調・競合行動の共進化による創発実験

を行った [12] . 2 台のロボット間でのパス・アンド・シュートの協調行動を創発させるために、3 台目のロボットを仮想敵として配置することで、3 台の同時学習による共進化を経時的発展として観測した .

- 内部構造：行動決定の木構造による表現と遺伝的プログラミングによる進化 .
- 環境設計：3 台目のロボットの配置及び同時学習などのスケジューリング . 経時的発展は進化の過程そのもの .
- 達成点：3 台の同時学習による多様な行動パターンの創発 .
- 課題：適合度関数の設計及び初期配置依存性の高さ .

8. 討 論

他者の認知が、ロボットの環境表現の多様性を引き起こし、その結果として、自分自身を内部表現できれば「自己」が確立しないかと考えている . そのために、必要な設計として、内部構造と環境構造の二つの側面があることを指摘した . 後者は、これまで明示的に言及されなかったが、認知ロボティクスの観点からは、発達を促す機構の体系化という意味で必要不可欠と考えられる . 上で述べてきた研究例では、基本的に学習を加速する目的で環境設計しているが、設計者の憶測の域をでていない . 動物学における「臨界期」的なタイミングをどのように設計するかは今後の課題の一つである .

認知過程として、現在まで我々がやってきたことは、環境表現の回数として、高いものとして「他者」が表現したことである . 欺瞞が自己とか自意識と関連することが、動物行動学 [13] などでも論議されているので、学習、発達過程で「フェイント」できるサッカーロボットができないものかと考えている . このためには、相手の複雑な行動を十分予測可能な状態ベクトルが推定されただけでなく、行為系列の分節化によるより複雑な内部表現 (より高度な抽象化) が必要である . 今後の具体的な問題として、以下を挙げる .

- (1) 他者が何を観測しているかを推定すること . これは、他者の行動観察から他者の政策を推定し、それにより観測物を推定する段階を含む . 他者の視線に立てることは、明確な他者の意識が存在することの証明 . 本特集号では、開 [14] の空間認知と関連する .
- (2) 記憶と時間の概念の獲得 . 現在まで、推定された状態に対する適切な行動を生成する機構を構築してきたが、時間という概念は獲得されていない . 記憶もない . 時間と記憶は認知主体が、自己の存在を意識するうえで必要不可欠と考えられ、何らかの内部構造が必要 .
- (3) 我々は観測と行動が、意図的であるか否かに関わらず、アイコンタクトによるコミュニケーションを成立させ

ると考えている . そのためには、個々のロボットが獲得した環境表現が物理場を通して共有されなければならない . サッカーという限られた環境なので、語彙もかぎられるが、何らかの形で原始的な言語の発生の可能性があると考えられる . 谷 [15] も環境との相互作用から言語発生に言及している .

参 考 文 献

- [1] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, E. Osawa, and H. Matsubara. "robocup: A challenge problem of ai". *AI Magazine*, Vol. 18, pp. 73-85, 1997.
- [2] 浅田, 石黒, 国吉. 認知ロボティクスの目指すもの. 日本ロボット学会誌, Vol. 17, No. 1, 1999.
- [3] 津本 (編). 脳と発達. 朝倉出版, 1986.
- [4] 大澤. 「知能の社会性」. 日本ロボット学会誌, Vol. 14, No. 4, pp. 15-19, 1996.
- [5] 国吉. 「実世界エージェントにおける注意と視点-情報の分節・統合・共有-」. 人工知能学会誌, Vol. 10, No. 4, pp. 507-514, 1995.
- [6] 浅田, 野田, 細田. ロボットの行動獲得のための状態空間の自律的構成. 日本ロボット学会誌, Vol. 15, No. 6, pp. 886-892, 1997.
- [7] M. Asada, S. Noda, S. Tawaratumida, and K. Hosoda. Purpose behavior acquisition for a real robot by vision-based reinforcement learning. *Machine Learning*, Vol. 23, pp. 279-303, 1996.
- [8] Y. Takahashi, M. Asada, and K. Hosoda. Reasonable performance in less learning time by real robot based on incremental state space segmentation. In *Proc. of IROS'96*, pp. 1518-1524, 1996.
- [9] M. Asada, E. Uchibe, S. Noda, S. Tawaratsumida, and K. Hosoda. "coordination of multiple behaviors acquired by vision-based reinforcement learning". In *Proc. of IROS'94*, pp. 917-924, 1994.
- [10] Eiji Uchibe, Minoru Asada, and Koh Hosoda. Environmental complexity control for vision-based learning mobile robot. In *Proc. of 1998 IEEE ICRA*, pp. 1865-1870, 1998.
- [11] E. Uchibe, M. Asada, and K. Hosoda. "State space construction for behavior acquisition in multi agent environments with vision and action". In *Proc. of ICCV-98*, pp. 870-875, 1998.
- [12] E. Uchibe, M. Nakamura, and M. Asada. "Co-evolution for Cooperative Behavior Acquisition in A Multiple Mobile Robot Environment". In *Proc. of IROS'98*, 1998.
- [13] リーキー著, 馬場悠男訳. 「ヒトはいつから人間になったか」. 草思社, 1996.
- [14] 開. 認知科学とロボティクス. 日本ロボット学会誌, Vol. 17, No. 1, 1999.
- [15] 谷. 自己意識の問題に関する構成論的アプローチ. 日本ロボット学会誌, Vol. 17, No. 1, 1999.

浅田 稔

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) . ロボカップ国際委員会副委員長, ロボカップ日本委員会委員長, 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)