

他者の信念の推定と自己の行動決定のための追跡ゲームの開発

Development of chasing game for a study of other's belief estimation and action decision

大森隆司 奥谷一陽

Takashi Omori, Kazuhiro Okutani

玉川大学工学部

School of Engineering, Tamagawa University

Abstract: In a communication scene between human, we estimate the belief of other person on outer world, and change own action based on the estimation. Belief inference is indispensable for the human-human, and for the human-agent interaction. However, there are few models that represent dynamics of belief in the interaction scene. So, in this study, we develop a chasing game that require the belief process for realizing higher level interaction, and evaluate a belief model we developed by a computer simulation. We also discuss on a possibility of cheating that makes use of the belief estimation process.

1. はじめに

コミュニケーションには、言語・身振り・表情のように外部からの直接観測が可能な過程と、意図や信念のように外部からの直接観測は不可能ではあるがかならず存在する心的な過程が存在する。前者は直接の観測という方法で情報伝達を実現し、後者は心についての推論によって相手の心的状態を理解することで自己の心的状態の変化を引き起こす。現在、コミュニケーションを目的あるいは手段とした多くの機器・ロボットが開発されているが、その開発の多くはユーザーの観測可能な部分についてのインターフェース特性の理解とそれに基づく装置の機能設計になっている[1]。しかし、コミュニケーションは相手との相互理解により相互の行動を適切に変えていくための手段であり、その本質はむしろ直接には観測できない心的相互作用にあらう。さまざまな状況で相手の発話の意図を十分に理解して適切な応答するためには、単に相手の表面的な行動の解釈だけでなく、直接観測はできない相手の信念や意図を推定しての意味解釈が必要である[2]。

これまで、心的過程を論理で記述する研究がいくつか行われてきた[1][2]。しかし現状では、それらは思考の論理的記述に終わっており、現実的な場面での運用にまでは至っていない。しかし意図や信念はコミュニケーションの重要な要因であり、これを無視することはできない。また横山は、意図を人間の

行動決定に影響する隠れ変数として定式化し、ハンターゲームという協調ゲームのコンピュータシミュレーションによりその意図推定の方式とその運用アルゴリズムを定式化し、その有用性を示した[3]。また阿部はそのモデルをロボットと子供の遊び場面に適用し、子供の興味状態の推定による同調性が相互作用に有用であると提唱している[4]。しかし、同様に重要と考えられる信念については、いまだ理論の範囲にとどまっており、現実的場面での運用についてのモデルの検討は少ない。

そこで本研究では、信念に焦点を当て、非言語コミュニケーション場面におけるそのモデルと運用方法について検討する。信念とは、行動主体が考えるところの外部世界についての事実である。個体レベルでは信念は環境を観測することで獲得され、環境の変化や行動主体の内部での信念の表現媒体（すなわち記憶）が活動の中で減衰することで消えていく。その過程で、行動主体は自身の環境についての認識（信念）に基づいて行動決定する。よって行動決定過程から信念を切り離すことはできない。

一方でコミュニケーション場面では、我々は対象の行動を観測してそこから対象のもつ信念を推定する。他者の信念の推定は他者の行動予測に直結し、その行動の予測を織り込んで自己の行動を決める方略は、現実の予測と自身の行動決定という煩雑な要素を分離させ、問題の単純化に繋がる。このように他者の信念の推定は、対人相互作用の基本的かつ重

要な問題である。

そこで本研究は、他者の信念の推定と行動予測に関わる問題点を整理するため、信念推定が必要な行動決定ゲームを開発する。そして、その世界での行動決定モデルの構築とシミュレーションによる検証を通じて、他者と自己の信念を通じた相互作用の基本的な原理を検討する。

2. 信念の現象のモデル化の試み

2.1. 本研究で扱う信念

本研究では「信念」を以下のⅠのように定義し、Ⅱ～Ⅳの性質があると考ええる。

- Ⅰ. 信念とは、行動主体が考えるところの外部世界についての事実もしくは事実からの推論である。
- Ⅱ. 信念は、外界の観測により獲得され、状況が変化するとその認識に基づき更新される。
- Ⅲ. 信念は、各行動主体が個別に持ち、観察しなくとも継続されるが、時間とともに減衰する。
- Ⅳ. 信念は、行動主体の行動決定に影響する。特定の状況では、観測対象の行動からその対象の持つ信念を推定できる。

2.2. 見えない他者についての信念

本研究では、対象が見えなくなった時の、その対象の動きや位置に関する信念を扱う。対象が見えなくなった時、人はその対象についての知識を用いてその対象の現在状態を推定する。例えば我々は、物質の保存則により、見えなくとも物体はそこにあると推定する。遮蔽物が取り除かれたとき、そこにその物体がない、あるいは別の物体が現れたなら、我々は驚く。驚くということは予想が外れたということ、我々は物体の存在を予測している。

しかし、このような物体に対する仮定は、対象が行動する主体に替わったとき、大きく変わる。行動主体は自ら考えて行動を起こすため、物質についての保存則は成り立たず、他の規則に基づく推定が必要になる。ここで必要なのは、他の行動主体の行動についての信念である。ここで問題を複雑にするのは、推定される相手もまた自身の行動や相手についての信念を持ち、それが行動に影響を及ぼすことである。我々は、相手を直接には観測できない間の信念の変化もまた推定していく必要がある。

例えば、行動主体 A, B が存在するとしよう。A は現在時刻 t までに観測した B の行動と外界の状況から、B の持つ信念を推定し、その信念と B についての知識（モデル）を参照して B の次の時刻以降の行

動を予測する。A はその予測に基づいて行動する、すなわち予測を信じているので、これは A が B について推定した事柄を信念として持つということになる。ところがこの A の B についての推定には、B が何を認識しているか、すなわち B の信念も含まれている。A はさらに、B の迷い、すなわち信念の揺らぎもまた B の行動から推定することもできよう。

2.3. 行動決定過程の信念

見えなくなった対象の追跡という問題を考えてみよう。行動主体は、対象の存在に対する信念、さらには対象の行動に対するモデルを持つことで、より効果的な追跡(逃走)行動が選択できよう。ここでモデルとは、相手はこういう場合はこう行動するという知識である。主体は、相手はこのモデルに従って行動するという信念を持つことで、自身の内部での相手の予測の方法が決まる。このモデルは、見えている間に相手を観測して学習・獲得し、見えない時にそれを運用することになる。しかしその背後には、相手が見えない期間もそれに従うという認識、すなわち相手の行動決定過程の安定性についての信念がある。

しかし予測の正確さには問題がある。相手の行動決定過程についての信念が間違っていれば行動予測は誤りを含み、その修正も相手が観測できないため難しい。再び観測ができて、見えない間にどのような行動を経て今の位置にたどり着いたかはわからない。予測性の向上には、対象の信念推定が正確である必要がある。そのためには、対象の信念過程についての十分なモデルが必要となろう。

行動決定の信念に関わる指標として、行動の一貫性を考える。「一貫性」とは、エージェントが目的に対して効率的かつ継続的に行動していることの指標である。一貫性の高い行動とは、目的地に向かって寄り道することなく最短経路で向かうことであり、低い行動とはノイズを多く含んだ行動となる。この指標は、継続した観測によって形成される。信念の立場からは、その行動主体が持つ自身の行動決定の正しさについての確信の指標となり、他者の信念推定としてはその行動決定過程がしばらくは継続されるであろうという予測につながる。

2.4. 提案する信念-行動モデル

図 1 は本稿で考える信念-行動モデルの概略である。エージェントは、対象を観測できない場合は、それまでの信念値を減衰も含みつつ更新し、信念値・対象のモデル・対象の一貫性推定値に基づいて行動リストから行動 $m(t)$ を選択して実行する。対

象を観測できた場合は、観測情報で信念を更新し、対象のモデルを獲得していく。

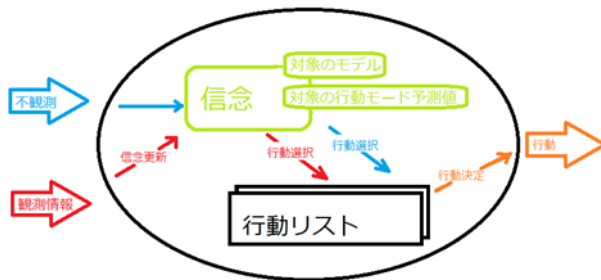


図 1. 本稿で考える信念－行動モデル

2. 5. 信念を要する課題

本研究では、信念の特性を検討してモデル構築につなぐために、信念の処理を必要とするゲームを開発する。そのゲームのテーマとして、「追跡」を考える。課題は、(1) 我々が目指す現実的な場面での信念の解明とその計算アルゴリズムの構築に必要とされる特性をゲームが持っているか、さらに(2)そのゲームの課題を解決するために必要とされる信念の計算アルゴリズムはどういうものであるか、ということを明らかにすることである。

まず、人対人での追跡課題を考えよう。遮蔽となる障害物(以後、障害物はすべて遮蔽物でもある)が無い場合には、「追跡者」は「逃走者」を観測して接近行動を行い、「逃走者」は「追跡者」を観測して逃避行動を行なう。このときは、次の時刻の対象の行動の予測はする必要がなく、対象に関する信念は必要とされない。

一方、障害物がある場合は、互いが遮蔽によって見えなくなり、接近(逃避)する対象の観測情報がなくなり、そのままでは行動決定ができない。ここで行動主体が対象に関する信念を持つなら、最後に相手を観測した地点に対象が存在すると信じて、あるいはその後に対象が移動をしたことを予測して信じ、行動を継続できる。すなわち、障害物のある追跡課題には信念が必要である。

3. 実験課題用エージェント

これまでの信念に関する議論をうけて、信念を持って行動するエージェントによる追跡課題を以下のように構築した。

3. 1. 追跡課題の規則

図 2 に示すように、四角形のグリッドワールド(以下、GW)に、追跡課題を行う追跡エージェント P(以

下、 P)と、逃走エージェント Q(以下、 Q)がいて、 P は Q を追跡し、 Q は P から逃走する。GWの中には障害物があり、 P と Q は間に障害物が入ると相互に観測できなくなる。各エージェントは、障害物がない限り上、下、左、右、停止の5方向の行動を選び、 P と Q の座標が重なると、各試行は終了する。

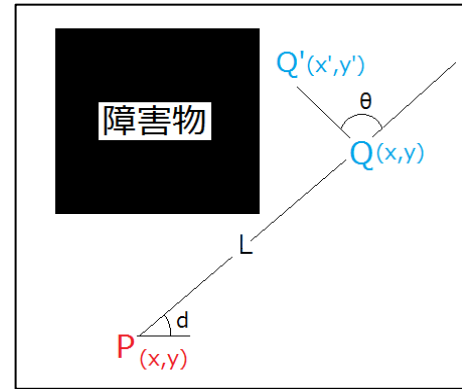


図 2 .二体のエージェントの関係を表す座標系

3. 2. 構成要素

各エージェントの構成要素を以下のように定める。

I. 対象との距離 L , 方位 d

II. 報酬・罰に対する信念 S_R

エージェントの持つ報酬・罰に関する知識。本シミュレーションでは、事前に類似した別の追跡課題で強化学習による学習した Q 値を知識 $S_R(L, d)$ として使用する。

III. 対象の行動モデル S_0

エージェントが持つ対象の行動のモデル。対象との距離と方位の空間での対象の行動 M の確率分布である。このモデルにより対象の行動を確率的に予測する。本シミュレーションではこのモデルは事前に他の課題で学習しておく。

$$M(t) = S_0(L, d)$$

V. 対象の行動方向 θ

対象 Q が時刻 $t-1$ から時刻 t 間の行動 $m(t)$ の、自己から対象への-方位に対する角度の開き θ_t (図 2)。 θ は -1 から 1 の間を取り、より $\theta_t=0$ に近い方向への行動が追跡者から効率よく逃れる Q の望ましい行動を表すと考ええる。

VI. 対象の行動への一貫性指標の推定特性 α, β

対象の行動の一貫性指標 K の推定に用いる。 α は、時刻 t における K の推定値の減衰率であり、その値が低いほど信念は早く減衰する。 β は時刻 t における K の推定値の増加率であり、対象の時刻 t における行動 $m(t)$ の方向 θ_t を信念にどれほ

ど反映するかを定める。

$$K_{PQ t} = \alpha(K_{PQ t-1} - 0.5) + 0.5 + \beta \cos \theta_t$$

IV. 一貫性指標 K

エージェントの行動の揺らぎを制御するパラメータで、0～1の間の値を取る。エージェントごとに個別に設定し、 S_R が表す状態価値にシグモイド関数を適用して行動決定の確率分布を作る際に、ボルツマン分布における温度 T の逆数と同様に、 K が大きい時には積極的かつ継続的な「一貫性の高い」行動、 K が小さな時にばらつきのある「一貫性の低い」行動を発生させる。

本研究では、一貫性指標 K は相手の行動から推定が可能と考える。対象の時刻 t の行動 m_t の行動方向 θ_t と、推定パラメータ α β を用い、 P が推定する Q の一貫性指標は以下のように推定される。 K は、対象の行動がない場合には中央値 0.5 に収束する。

エージェントの行動モデルはすべて共通であり、個々のエージェントの個性は一貫性指標に反映される。それを推定することで、個々のエージェントの行動特性が予測に反映される。

後述するシミュレーションでは、 P と Q にはそれぞれ自身の行動の一貫性指標 K_P と K_Q があり、さらに P と Q がそれぞれの他者の一貫性を推定した K_{PQ} と K_{QP} が原理的には含まれている。本研究のシミュレーションでは、 K_P と K_Q はそれぞれ外部から設定し、 P による K_{PQ} の推定を評価する。

前節までに示した規則と構成要素によるエージェントの振る舞いのフローチャートを図 3 に示す。追跡と逃走の両エージェントがこれに基づいて交互に行動決定を繰り返し、互いの座標が重なった時点で追跡の成功とする。

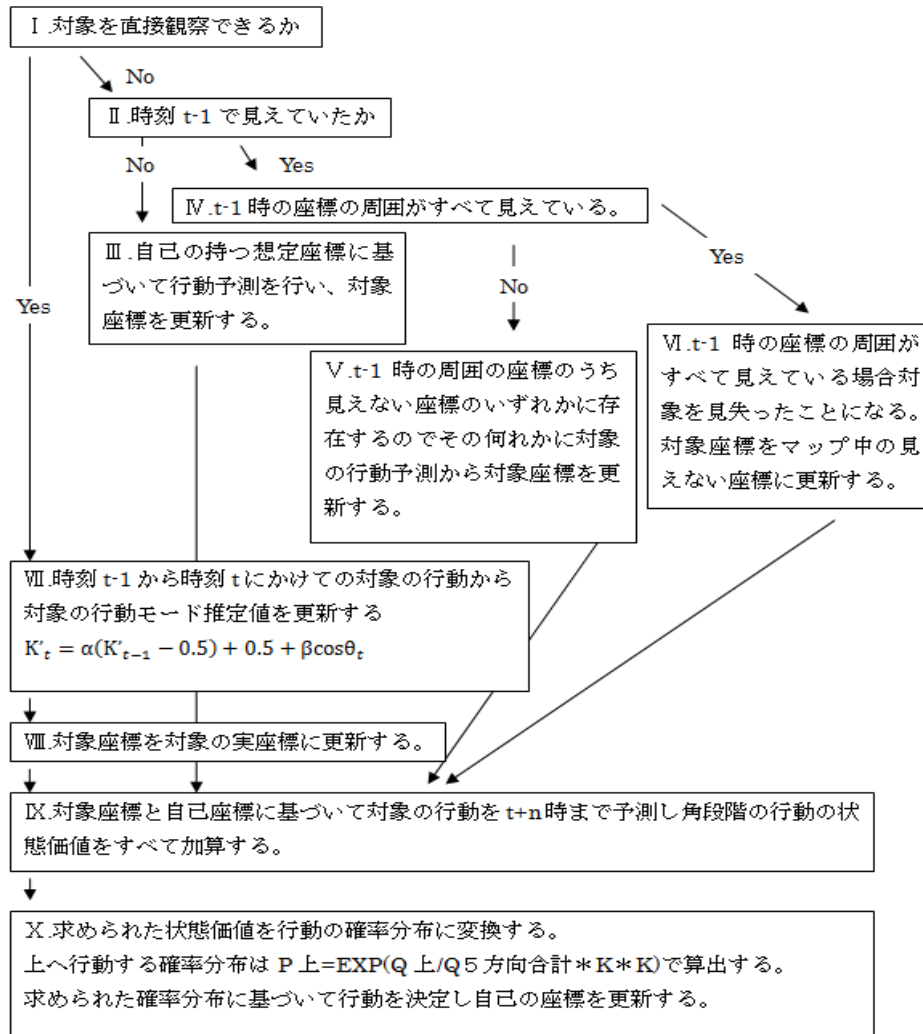


図 3 エージェントの行動決定過程のフロー

シミュレーション実験

3.3. 実験1：一貫性指標の推定

GW は縦横 50×50 マスとして障害物は置かず、すべての時刻に対象が見えるようにした。P と Q との一貫性指標はともに $K_P=0.9$, $K_Q=0.9$ として、その推定特性は $\alpha=0.1$ $\beta=0.8$ とした。課題は、P に Q の一貫性特性 K_Q の推定を行わせることである。想定では、Q の一貫性指標 $K_Q=0.9$ を P が推定し、推定値が $K_{PQ}=0.9$ に近づくと期待される。

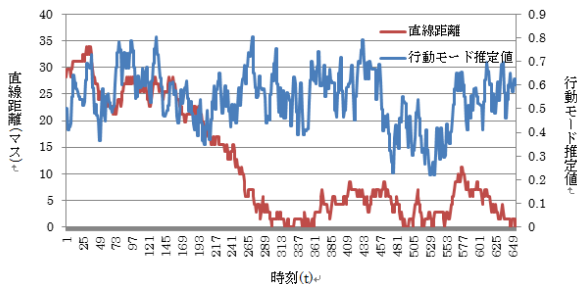


図4 実験1：一貫性指標の推定シミュレーション

図4に推定シミュレーションの結果を示す。想定と大きく異なり、推定値 K_{PQ} は実際の値 $K_Q=0.9$ には届かず、大きく値の振れる不安定なものであった。推定値は最高で $K_{PQ}=0.8$ までであった。この原因は、Q が P から離れる中で自発的に GW の角に寄って行って追い詰められ、本来進むべき $\theta_t=0$ の方向に動けなかったことによる。

3.4. 実験2：一貫性指標の追従

実験1と同様に、GW の設定は縦横 50×50 マスで障害物は置かず、すべての時刻に対象が見えるようにした。P の一貫性指標を $K_P=0.5$ として、その推定特性は $\alpha=0.1$ $\beta=0.8$ とした。実験1と異なり、Q の一貫性指標 K_Q が $t=100$ まで $K_Q=0.9$ であるが、それ以後は $K_Q=0.1$ に変更した。この変更に対して、追跡エージェントが推定する K_Q の推定値 K_{PQ} がその変更に従うことを期待している。

結果を図5に示す。シミュレーションの開始 $t=0$ から $t=100$ までは推定値 K_{PQ} に振れながらも上昇しており、 $t=100$ に $K_Q=0.1$ に変更して以降は推定値 K_{PQ} に減少している。しかし、実験1と同様の原因から、安定して推定できているとは言えない。

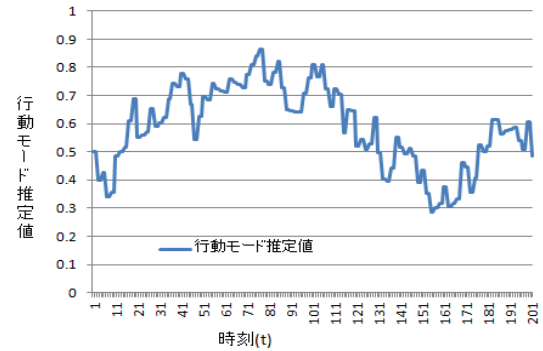


図5 実験2：一貫性指標推定の追従実験

4. 考察

シミュレーション実験では、対象の行動方向のみから対象の行動の一貫性を推定した。考慮に入れている状況が単純に過ぎ、行動の意図が逃走にもかかわらず、推定値が荒れる結果となった。追跡中に相手を角や壁に追い込めている場面ではすでに追跡にほぼ成功しており、相手の次の時刻の動きの予測は必要なくなっている。これは、十分な距離を保った状態で逃走エージェントが自発的に角に向かう傾向そのものに問題がある。本来なら行動に制限がかかって不利になる角や壁には近づかないのが自然である。角・壁に関する行動制限の知識があるべき考えるのは自然であり、我々のエージェントモデルにも環境に対するより多くの知識を与える必要がある。

一貫性指標の推定値を信念に持つことは、対象の行動から一貫性に関わる信念を読み取ることにになり、これまでよりもより強力な行動予測が実現できると期待される。例えば「だまし」がある。「だまし」とは、相手の内的過程を誘導して通常は取らない行動選択をさせ、その後にはその決定とは条件が異なる行動をこちらが取ることで実現される。

今回の課題では、例えば P が自己の一貫性を高く見せかけるためにしばらく $K_P=0.99$ で行動を続け、Q の持つ P についての K_P の推定値 K_{QP} が十分大きくなるように誘導しておく。そして障害物で互いが見えなくなった時、P が長期のプランニングを行う。この瞬間には、Q は P の一貫性は高いと推定しているため現在の逃走行動を相当時間続けると予測される。そこで P は、Q による自分についての行動予測を裏切り、障害物の裏側から回り込んで待ち伏せする、といった行動である。

このようなだまし行動が実現できるためには、追跡エージェントが相手を「走らせる」や「待ち伏せる」といったプランについての価値を認識できる仕掛けが必要になる。そのため、たとえば行動の「エネルギー」を考える可能性があろう。行動によりエ

エネルギーを消費すれば動けなくなることを学習することで、エネルギーを温存することの価値が解り不必要な行動の減少が顕著に表れるであろう。従来、このような価値は課題を作る設計者が意図的にエージェントに組み込んでいる。しかし HAI として人間の理解をすすめるには、なぜ人間がそのような行動に価値を置くのか、ということも含めて原理を検討する必要がある。

5. まとめ

行動決定における信念の機能を明らかにするため、信念の機能がなければ適切に解けないゲーム課題を提案し、それを解くためのエージェントモデルを提案した。信念を導入することで、エージェントは他者の信念、あるいは他者の自己に対する信念までも推定することで自身の行動決定をよりの確なものにできるし、さらには「だまし」のような相手の判断の誘導までできる可能性を示した。「だまし」がかならずしも HAI に必要という訳ではないが、相手の信念を考えて相手の判断を誘導する程度の処理ができるなら、高いレベルの HAI を実現することにつながるであろう。今後は、本研究の成果を踏まえた上位の行動決定システムの検討を行っていく予定である。

謝辞

本研究は、科研費新学術領域研究「伝達創成機構 21120001」21120010 の助成を受け実施した。また玉川大学岡田浩之教授にはこの研究の初期の段階のアイデアについて有益なコメントをいただいた。関係各位に深く感謝する。

参考文献

- [1]菅野太郎、古田一雄,チーム協調モデルに基づく HA インタラクション戦略 -自然な人間-エージェント協調へ向けて-, HAI2006、1-2A-2,(2006)
- [2]大沢一郎：対話行為における知的行為者の計算モデル, コンピュータソフトウェア, Vol.6, No.4, pp348-359,(1989)
- [3]横山絢美, 大森隆司：協調課題における意図推定に基づく行動決定過程のモデル的解析, 電子情報通信学会論文誌 A, Vol. J92-A, No.11, pp.734-742, 2009
- [4]阿部香澄, 岩崎安希子, 中村友昭, 長井隆行, 横山絢美, 下斗米貴之, 岡田浩之, 大森隆司：子供と遊ぶロボット：他者の状態推定に基づく行動決定モデルの適用, HAI2011, 1-2B-4, 2011