確率モデルに基づく他者モデル相互適応のモデル化

Probabilistic Modeling of Mutual Adaptation of Models of Others

鴫原宏明 † アッタミミ ムハンマド † 阿部香澄 † 長井隆行 † * 大森隆司 ‡ 岡夏樹 ††
Hiroaki Shigihara, Muhammad Attamimi, Kasumi Abe, Takayuki Nagai, Takashi Oomori, Natsuki Oka

† 電気通信大学 ‡ 玉川大学 †† 京都工芸繊維大学
The University of Electro-Communications, Tamagawa University, Kyoto Institute of Technology

Abstract: Intimacy is a very important factor not only for the communication between humans but also for the communication between humans and robots. In this research we propose an action decision model based on other's favor, which can be estimated using the model of others. We examine the mutual adaptation process of two agents, each of which has its own model of others, through the interaction between them.

1 はじめに

最近の HRI 研究では，人とロボットが長期的に関わるために，興味や飽きといった新奇性の軸ではなく，親近性の軸と呼ぶ関係構築が重要であることが明らかにされつつある [1, 2]．また心理学においては，人同士の親密な友人関係が，適応や精神的健康を高めることが多くの研究によって示されており，その重要性が指摘されている [3]．これらのことは，HRI において長期に渡りより良いインタラクションを実現するために，ロボットと人が親密な友好関係を形成することが重要であることを物語っている．人と人のコミュニケーションにおいては，コミュニケーション相手の心的モデルである「他者モデル」が非常に重要な役割を果たしており，様々な手がかりを利用して相手の心的状態を推定し，行動を予測していると考えられる．従って，他者モデルのパラメータを推定し運用することが，円滑なコミュニケーションや関係構築に不可欠であると考える．

しかしながら，他者モデルが実際にどのようなモデルで，どのように働いているのかについてはあまり明らかにされていないのが現状である．そこで本研究では，ロボットに搭載するために，他者との関係を構築するための他者モデルを検討する．そしてその他者モデルをもった者同士が，どのようにモデルを相互適応させるのかについて明らかにすることを目指す．これは，モデルの相互適応こそが，二者の関係を構築するための重要なプロセスであると考えられるためである．またこうしたプロセスの背後には，行動決定の方策，つまりはどのような規範に基づいて行動を決定するのかという問題がある．人同士のインタラクションにおいては，友人との間に親密な関係を築こうという共同目標が，友人からの好意に影響することが明らかにされている [3]．そのため，ロボットが人と親密な友人関係を形成するためには，人の好意を推定し，それを高めるような目標を持って行動を決定することが必要だと考えられる．

これまでにも，他者の推定する自分の意図を推定することで，作業効率の向上を目指す研究 [4] や，他者の行動決定方策を推定するモデルを構築することで，自律的に利他的行動を生み出すことを目的とした研究 [5] は行われている．しかしこれらの研究は，あくまで作業効率の向上が主な目的であり，人とロボットによる親密な関係の構築を目的として，モデルの構築および行動決定手法を検討する研究は見られない．平川 [6] らは，エージェントがユーザーの飽きを推定して行動戦略を変更することにより，継続的なインタラクションの実現を目指した．しかし，ユーザーの飽きに即時的に対応するだけでは，長期的な関係性の構築は困難である [2]．

本稿では人の好意に影響を及ぼす因子を挙げ，その因果関係を確率モデルで表現し，これをエージェントの行動決定モデルであると同時に，他者モデルであると考える．このモデルを使った強化学習により，他者モデルが相互に適応し，関係性が構築されていくプロセスをシミュレートする．この際に，どのような報酬を最大化するかというメタ戦略が，最終的に構築される関係に対してどのような意味を持つかについても検討する．

2 提案モデル概要

本研究では，人とロボットが特定のタスクを行うことを想定し，モデル化を行う．人とロボットのインタラクションの概念を，図 1 に示す． 本稿におけるインタ

*連絡先：国立大学法人電気通信大学
〒 182-0021 東京都調布市調布ケ丘 1-5-1
E-mail:hchie@apple.ee.uec.ac.jp

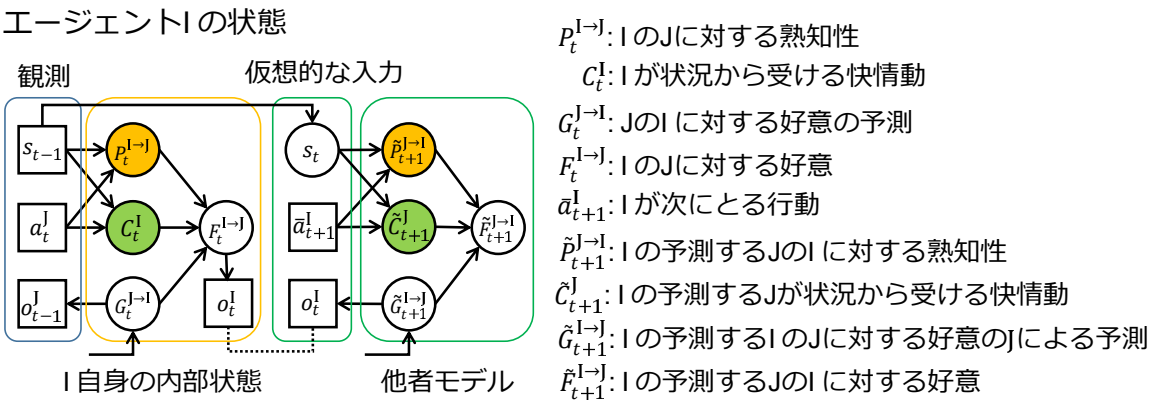


図 2: ベイジアンネットワークによる他者モデル

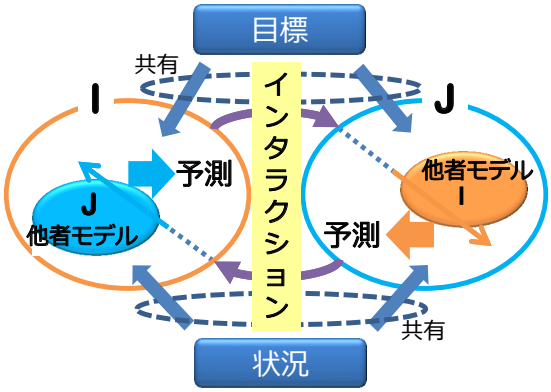


図 1: HRI 概念図

ラクションでは，可観測な因子として表情と音声の韻律，行動とそれにより変化する環境 (状況) を扱う．また，好意に影響を与える非観測な内部状態として，以下の 3 つを考える．

熟知性 熟知性とは，相手について良く知っていると感じている度合いであり，これが高いと相手への魅力の一因である「安心感」が高まり，好意に発展するとされている [7].

相手からの好意 人には，好意が返ってくるとわかっている相手に対して好意を持ちやすくなる好意の返報性という性質があることが指摘されている [8]. これは，相手から感じる好意が，相手への好意に影響を与えることを示唆している．

快情動 人は高い報酬を受けることができると考えられる他者に対して好意を持ちやすいことが，研究によって示されている．ここでは，最も原始的な情動である快情動を報酬とし，これが好意に影響を与えると考える．

これら 3 つの内部状態が可観測な情報を入力として変化し，その結果好意が増減するとして，その因果関係をベイジアンネットワークとして表現する．

本稿では，これをエージェント自身の動作モデル (自己モデル) として利用すると同時に，他者モデルとして他者の内部状態を推定するモデルとしても利用し，各エージェントが図 2 に示すような複合モデルを内部状態として持つと考える．

2.1 内部状態の更新

熟知性は，相手行動の予測精度によって変化すると考える．そこで，まず現在までの経験の蓄積を元に行動予測器を学習する．ここでは簡単のため，各状態に対して相手が取った行動をカウントすることで行動確率表を更新することを想定する．そして，行動予測器から出力される相手行動の予測と，実際に観測される次の相手の行動の比較結果から，熟知性を計算する．また，快情動については，ある状況での相手の行動に対する，快情動の出力確率を表した条件付き確率表を事前に持っているものと仮定し，その表に基づいて更新する．相手からの好意については，観測された相手の表情と韻律から，事前に与えられた条件付き確率表に基づいて，相手の持つ好意を推定し，更新する．これらのパラメータ（確率表）は，EM アルゴリズムによって推定することが可能であるが，後に示すシミュレーションでは基本的な振る舞いを観察するため，固定している．

2.2 モデルによる行動決定

前述の通り，人同士が親密な友人関係を築くためには，相手の自分への好意を推定し，それを高めることを目標として行動決定することが必要であると考えられる．そこで，相手が自分と同じ自己モデルを持っていると考え，自分の中に相手の自己モデルを構築することで，相手の自分への好意を推定する．この自分の

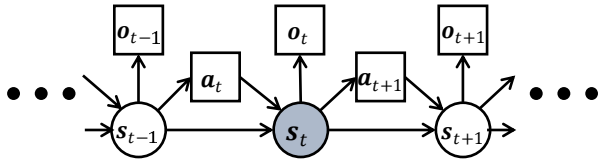


図 3: マルコフ決定過程によるモデル化

中に構築した相手の自己モデルが，他者モデルである．エージェントは自己モデルに基いて表情や音声を出力することで，相手の持つ自己モデルと他者モデルに影響を与える．

エージェント 2 者の行動を $\boldsymbol{a_t}$，表情や音声などの観測情報を $\boldsymbol{o_t}$，両エージェントと環境の状態を $\boldsymbol{s_t}$ として，それぞれ一つにまとめると，2 者のインタラクションを図 3 に示すようなマルコフ決定過程でモデル化することができる．このように表すことで，各エージェントの学習による行動決定を，強化学習 (Q-学習) によって定式化することができる．

他者モデルによる行動決定は，推定される他者の好意の増減を報酬 (好意報酬) とし，それを最大化することを目標とした強化学習として定式化する．その手法をエージェント I，J を例にして以下に述べる．次の I の行動を a^I_{t+1} とすると，他者モデル内の熟知性 $\widetilde{P}^{J\rightarrow I}_{t+1}$ の確率は $P(\widetilde{P}^{J\rightarrow I}_{t+1}|s_t, a^I_{t+1})$ として表現できる．同様に，他者モデル内の快情動 $\widetilde{C}^J_{t+1}$ の確率は $P(\widetilde{C}^J_{t+1}|s_t, a^I_{t+1})$ として表現できる．また，観測情報から J が予測する I の好意の予測 $\widetilde{G}^{I\rightarrow J}_{t+1}$ の確率は $P(\widetilde{G}^{I\rightarrow J}_{t+1}|o^I_t)$ として表現できる．

さらに，ある熟知性 $\widetilde{P}^{J\rightarrow I}_{t+1}$，快情動 $\widetilde{C}^J_{t+1}$，I から J への好意 $\widetilde{G}^{I\rightarrow J}_{t+1}$ における好意報酬 $r\widetilde{F}^{J\rightarrow I}_{t+1}$ の確率を $P(r\widetilde{F}^{J\rightarrow I}_{t+1}|\widetilde{P}^{J\rightarrow I}_{t+1}, \widetilde{C}^J_{t+1}, \widetilde{G}^{I\rightarrow J}_{t+1})$ と表現すると，次の自分の行動 $\overline{a}^I_{t+1}$ に対する好意報酬 $r\widetilde{F}^{J\rightarrow I}_{t+1}$ の期待値 $E\{r\widetilde{F}^{J\rightarrow I}_{t+1}\}$ は

$$\begin{aligned} E\{r\widetilde{F}^{J\rightarrow I}_{t+1}\} &= \sum r\widetilde{F}^{J\rightarrow I}_{t+1} \\ &\times P(r\widetilde{F}^{J\rightarrow I}_{t+1}|\widetilde{P}^{J\rightarrow I}_{t+1}, \widetilde{C}^J_{t+1}, \widetilde{G}^{I\rightarrow J}_{t+1}) \\ &\times P(\widetilde{P}^{J\rightarrow I}_{t+1}|s_t, a^I_{t+1}) \\ &\times P(\widetilde{C}^J_{t+1}|s_t, a^I_{t+1}) \\ &\times P(\widetilde{G}^{I\rightarrow J}_{t+1}|o^I_t) \end{aligned} \quad (1)$$

によって計算することができる．

この好意報酬の期待値を用いて Q-学習を行い，ある状況における行動価値表 Q-table を次式により更新する．

$$\begin{aligned} Q(s(t);a(t)) \leftarrow (1-\alpha)Q(s(t);a(t)) \\ +\alpha(E\{rFr_o)\} + maxQ(s(t+1);a')) \end{aligned} \quad (2)$$

ここで，α $(0 < \alpha < 1)$ は過去の経験をどれだけ重視するかを表す学習率であり，γ $(0 < \gamma < 1)$ は遠い未来の報酬をどれだけ重視するかを表す割引率である．つまり，$Q(s(t),a(t))$ は，状態 $s(t)$ において行動 $a(t)$ を取り，その後報酬の期待値 $E\{rFrs\}$ が最大となる行動を選択した場合に期待される価値を表す．

他者の好意報酬を推定し，それに基いて Q-学習により現在の状況における各行動の価値を計算して価値が最大となる行動を選択することで，他者の好意を最も高めるような行動決定をすることが可能となる．

一方で，タスク自体にも報酬が存在するために，行動価値をタスクの報酬に基づいて更新することも可能である．これは，従来の強化学習の枠組みであり，他者の内部状態を一切考慮せずに振る舞うことに相当する．また，他者の好意とタスクの報酬の重み和を考えることも可能であり，どのような重みで双方を考慮すべきかを調整するのが，メタ戦略である．

3 シミュレーション実験概要

提案モデルの動作を確かめるため，繰り返し囚人のジレンマ (IPD) ゲームを用いたシミュレーション実験を行う．

3.1 囚人のジレンマゲーム

囚人のジレンマゲームは，ゲーム理論の枠組みの一つであり，利他的な行動を理論的に研究するために考案されたものである．I と J のエージェント 2 者が 1 ステップに一度，同時に「協力」か「裏切り」のどちらかを選び，その結果によって表 1 のような報酬を得るとする．このゲームにおいて，協調行動は利他的な行動と考えられる．例えば相手が協力行動を選ぶと仮定すると，自分が協力行動を選んだとき，裏切り行動を選んだときと比べて，相手が受ける報酬は 3 大きくなる．逆に，相手が裏切り行動を選ぶと仮定すると，自分が協力行動を選んだとき，裏切り行動を選んだときと比べて，相手が受ける報酬は 4 大きくなる．両者の報酬の和を最大にしようと考えるならば，互いに協力行動を選択する方が良いが，自身の報酬だけを最大にしようと考えるならば，裏切り行動を選択する方が良いということになる．そのため，それぞれがタスクによる自身の報酬を最大化するような最適解を探すと，両者が裏切りを選択し，互いに損をすることとなる．

3.2 事前設定パラメータ

本稿において行ったシミュレーションでは，報酬が大きいほど快情動が大きくなると考えて，快情動の生起確率を表 2 のように設定した．また，Q-学習のパラメータとして，学習率を 0.5，割引率を 0.999，探索のために行動決定時に 0.2 の確率で行動がランダムに置き換わるように設定した．

表 1: 囚人のジレンマゲームの報酬表

I＼J	協力	裏切り
協力	3＼3	0＼5
裏切り	5＼0	1＼1

表 2: 快情動の条件付き確率

報酬＼快情動	6	5	4	3	2	1	0
5	0.4	0.3	0.3	0	0	0	0
3	0	0.3	0.3	0.4	0	0	0
1	0	0	0	0.4	0.3	0.3	0
0	0	0	0	0	0.3	0.3	0.4

表 3: 条件 1

	メタ戦略 I	メタ戦略 J	自己モデル I	他者モデル I	自己モデル J	他者モデル J
条件 1.1	タスク重視	タスク重視	ポジティブ	ポジティブ	ポジティブ	ポジティブ
条件 1.2	重みづけ差なし	重みづけ差なし	ポジティブ	ポジティブ	ポジティブ	ポジティブ
条件 1.3	好意重視	好意重視	ポジティブ	ポジティブ	ポジティブ	ポジティブ

4　シミュレーション結果

提案モデルは，自己モデル及び他者モデルの条件付き確率と，報酬に対する重みづけによるメタ戦略のパラメータを，任意に変更できるように設計されている．各エージェントが持つモデルによる動作の違いを検討するために，自己モデル及び他者モデルの条件付き確率パラメータとして,「快情動」「熟知性」「相手からの好意」が高いほど好意が高くなるポジティブモデルと，それらが低いほど好意が高くなるネガティブモデルの 2 種類を用意した．そして，以下の 3 つの異なる条件でシミュレーションを行った．

4.1　条件 1：モデル，メタ戦略一致条件

条件 1 では，等しい自己モデル及び他者モデルを持つエージェント I，J の 2 者が，同様のメタ戦略を用いて繰り返し囚人のジレンマゲームを行う．

Table 3 に示す各条件で 1000step のシミュレーションを 10 回ずつ行い，エージェント I，J それぞれが得た相手への好意の平均と標準偏差の時間推移を Fig. 4 に示す．

また，シミュレーションによって各エージェントが得た 1step 当たりの平均報酬を Table 4 に示す．

Fig. 4 を見ると，提案モデルを用いてそれぞれが好意報酬を重視したメタ戦略を取ることで，互いに好意を高め合っていることがわかる．また，Table 4 より，両エージェントが獲得した平均報酬は，タスク報酬を重視すると 1 に近く，好意報酬を重視すると 3 に近くなることがわかる．これは，タスク重視時には両者が自身の受ける報酬を優先した結果，相互裏切り状態に収束し，好意重視時には相手が受ける報酬を優先した結果，相互協力状態に収束したためだと考えられる．

4.2　条件 2：モデル不一致条件

条件 2 では，異なる自己モデル及び他者モデルを持つエージェント I，J の 2 者が，同様のメタ戦略を用いて繰り返し囚人のジレンマゲームを行う．表 5 に示す各条件で 1000 ステップのシミュレーションを 10 回ずつ行い，エージェント I，J それぞれが得た相手への好意の平均と標準偏差の時間推移を図 5 に示す．

表 4: 条件 1 における平均報酬

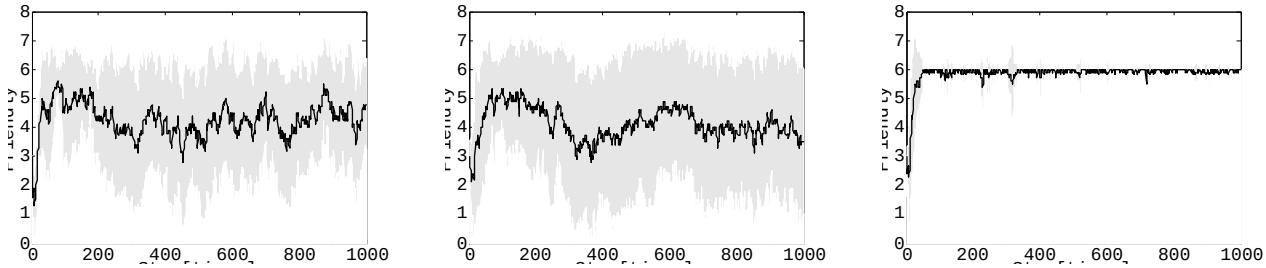
	エージェント I	エージェント J
条件 1.1	1.387	1.672
条件 1.2	2.218	2.247
条件 1.3	2.960	2.680

図 5 の結果から，自分の他者モデルと相手の自己モデルが全く異なっている場合，エージェントは相手の好意を推定することができず，好意重視のメタ戦略を取っていても，相手の自分への好意を高められないことが分かる．また，自分と相手の自己モデルが異なっていても，自分の他者モデルと相手の自己モデルが等しければ，相手の好意を推定し，高めるような行動選択が可能となることが分かる．

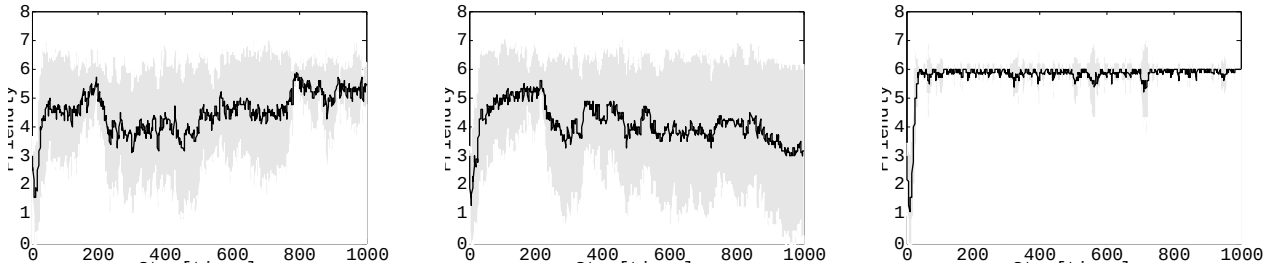
4.3　条件 3：メタ戦略不一致条件

条件 3 では，等しい自己モデル及び他者モデルを持つエージェント I，J2 者が，異なるメタ戦略を用いて繰り返し囚人のジレンマゲームを行う．表 6 に示す各条件で 1000 ステップのシミュレーションを 10 回ずつ行い，エージェント I，J それぞれが得た相手への好意の平均と標準偏差の時間推移を図 6 に示す．また，シミュレーションによって各エージェントが得た 1 ステップ当たりの平均報酬を表 7 に示す．

図 6 を見ると，異なるメタ戦略を持つ相手に対しても，好意重視戦略を取ることで好意を高められることがわかる．また，表 7 より，条件 3.2 で両エージェントの獲得する報酬の差が最大となることから，タスク重視戦略は自身が報酬を得ることを優先する戦略，好意重視戦略は相手が報酬を得ることを優先する戦略であることがわかる．



(a) Condition 1.1 I　(c) Condition 1.2 I　(e) Condition 1.3 I



(b) Condition 1.1 J　(d) Condition 1.2 J　(f) Condition 1.3 J

図 4: 条件 1 における好意の平均と標準偏差

表 5: 条件 2

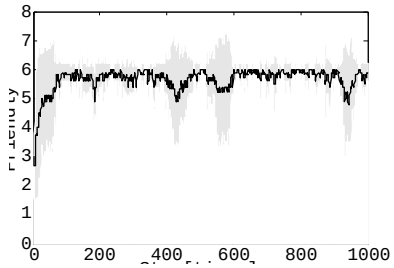
	メタ戦略 I	メタ戦略 J	自己モデル I	他者モデル I	自己モデル J	他者モデル J
条件 2.1	好意重視	好意重視	ポジティブ	ポジティブ	ネガティブ	ポジティブ
条件 2.2	好意重視	好意重視	ポジティブ	ネガティブ	ネガティブ	ポジティブ

5 議論

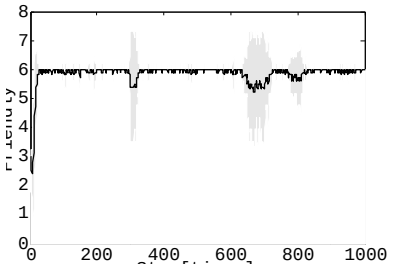
提案モデルを用いてシミュレーションを行った結果，等しいモデルを持つ 2 者であれば，好意報酬を重視して強化学習を行うことで相手の好意を高めるような行動選択が可能となることがわかった．しかし，本来人同士のインタラクションでは，自分と相手のモデルが常に一致しているとは限らない．また，今回のシミュレーションでは，インタラクション中における他者モデルの内部状態を表すベイジアンネットワークのパラメータは一定であったが，人同士のインタラクションにおいては相手に適応して常に更新しているものと考えられる．そのため，不明なモデルを持つ相手に合わせて，エージェントが自らモデルを更新し，適応するようなアルゴリズムが今後必要となる．提案したモデルでは，他者モデルをベイジアンネットワークで表現しているため，可観測な情報を入力として，EM アルゴリズムによって学習し更新することが可能である．

今後このような他者モデルの適応アルゴリズムを実装することで，エージェント同士や，エージェント対ユーザによるシミュレーションを行った際に，どのような現象が起こるのかを調べたいと考えている．例えば，互いに他者モデルが十分に適応できていない状態で，好意を重視し相手に合わせて行動しようとすると，他者モデルの適応が上手くいかず，好意を持つことが難しくなるといった現象を再現できると考えられる．また，他者モデルが十分に適応できていないときには，強化学習は行わず，他者モデルの適応を目的とした探索行動のみを行うなどといった，モデルの適応と強化学習のバランスに関する戦略を，シミュレーションによって検討したいと考えている．

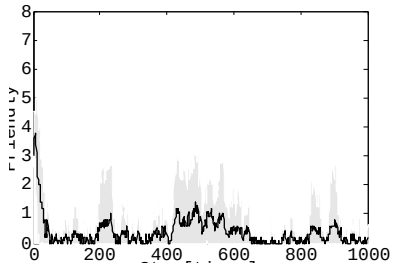
好意重視戦略を取ることで好意を高めることができるという結果が示されたが，特定のメタ戦略を取り続けると，行動パターンが一定化するという問題がある．今回のシミュレーションでは人の飽きを考慮していないため，一定の行動パターンに収束しても好意は高い値に留まり続けたが，人相手では飽きにより時間とと



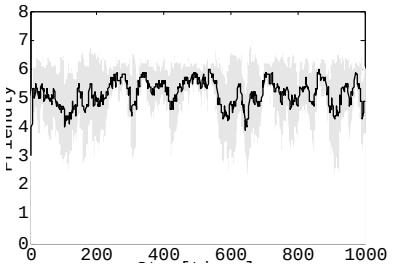
(a) Condition 2.1 I



(c) Condition 2.2 I



(b) Condition 2.1 J



(d) Condition 2.2 J

図 5: 条件 2 における好意の平均と標準偏差

表 6: 条件 3

	メタ戦略 I	メタ戦略 J	自己モデル I	他者モデル I	自己モデル J	他者モデル J
条件 3.1	タスク重視	重みづけ差なし	ポジティブ	ポジティブ	ポジティブ	ポジティブ
条件 3.2	タスク重視	好意重視	ポジティブ	ポジティブ	ポジティブ	ポジティブ
条件 3.3	重みづけ差なし	好意重視	ポジティブ	ポジティブ	ポジティブ	ポジティブ

もに好意が下がることが考えられる．そのため人の飽きを考慮してメタ戦略を調整するメタメタ戦略についても，今後シミュレーションや対ユーザ実験を通して検討する必要があると考えている．

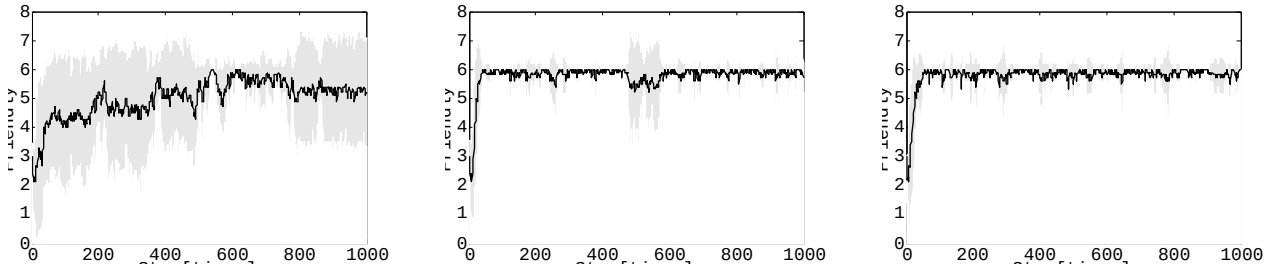
6 まとめ

本稿では，他者と親密な関係性を築くことを目的として，好意に至る因果関係モデルの構築と，他者モデルの相互作用を考慮した行動決定手法を提案し，その動作をシミュレーションによって検討した．

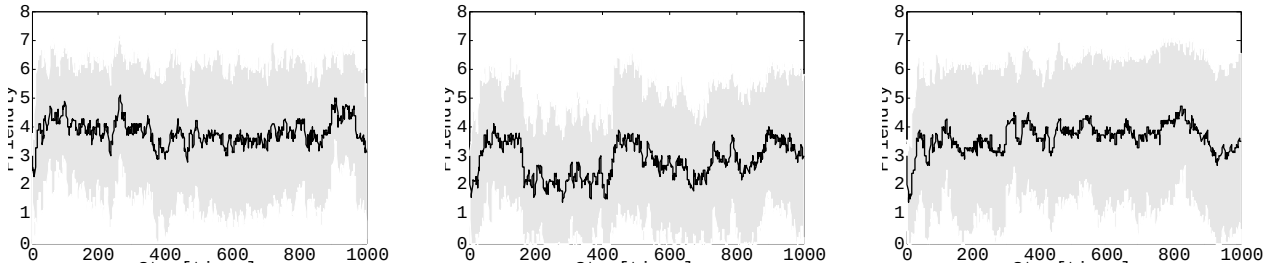
その結果，提案モデルを持つエージェント同士のインタラクションでは，好意報酬を重視したメタ戦略を取ることで，相手の好意を高めるような行動選択が可能となることが，簡単なタスクを用いたシミュレーションによってではあるが，明らかとなった．

また，例え自分と相手が異なる性質の自己モデルを持っていたとしても，自分の他者モデルを相手の自己モデルに合わせることで，好意を高めるような行動選択が可能となることが示唆された．ここでは，インタラクション中における内部状態を表すベイジアンネットワークのパラメータは一定であったが，人同士においては，インタラクションを通じて相手に合わせてモデルを更新しているものと考えられる．そこで今後は，ベイジアンネットワークの枠組みで構築された他者モデルのパラメータを，インタラクションを通じて更新し，相手の自己モデルに適応するアルゴリズムを実装するつもりである．そして，エージェント同士，エージェント対ユーザによるシミュレーションを行い，未知の自己モデルを持つ相手に対して，正しく他者モデルを適応させ，好意を高めるような行動選択が可能となるか検討したいと考えている．

最終的には，ロボット対ユーザによる実験を行い，モデルを搭載することでユーザが実際にどのように感じるかを明らかにしたいと考えている．



(a) Condition 3.1 I (c) Condition 3.2 I (e) Condition 3.3 I



(b) Condition 3.1 J (d) Condition 3.2 J (f) Condition 3.3 J

図 6: 条件 3 における好意の平均と標準偏差

謝辞

本研究は, 文部科学省 科学研究補助金 (認知的インタラクションデザイン学) の助成を受けたものである.

参考文献

表 7: 条件 3 における平均報酬

	エージェント I	エージェント J
条件 3.1	2.662	1.416
条件 3.2	3.726	1.130
条件 3.3	3.421	1.785

[1] 高橋 他: “「新奇性」と「親近性」の軸から子どもとロボットの関係性を捉える”, HAI シンポジウム 2011, 2B-2-L, 2011

[2] 阿部 他: 子供と遊ぶロボット : 心的状態の推定に基づいた行動決定モデルの適用, 日本ロボット学会誌, Vol.31, No.3(2013).

[3] 岡田 他: “親密な友人関係の形成・維持過程の動機づけモデルの構築”, 教育心理学研究, 56(4), pp.575–588, 2008

[4] 横山 他: “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, pp.734–742, 2009

[5] 牧野 他: “利他的行動と再帰的他者推定”, 生産研究, 62(3), pp.259–265, 2010

[6] 平川 他: “HAI の促進と持続に関する一考察”, HAI シンポジウム 2007, 2007

[7] 西浦 他: “同性友人関係における主観的熟知性が魅力に及ぼす影響”, 日本社会心理学会, 2011

[8] 脇本 他: “対人関係における行動傾向の類似性と親密性の相関関係”, 関西学院大学社会学部紀要, 78, pp.85–96, 1997