

## HAI研究のモデル化への挑戦

## Challenge for Modeling Human-Agent Interaction

今井 倫太\*1

Michita Imai

慶應義塾大学 理工学部

Faculty of Science and Technology, Keio University

This paper proposes a channel theory for recognizing an agency. We employ the theory of information flow as a basis of our theory. There are three types of channels in our theory: signal interpretation channels, model reference channel, and model sharing channel. The paper interprets examples of Human-Agent Interaction on the three channels. Our theory succeeded in explaining the difference of interaction based on the difference of the used channel.

## 1. はじめに

人と機械のインタラクションの研究は、様々なデザインやサービスが提案される中で発展してきた。しかしながら、研究としては、個別の事例が提案されるのみで体型だった理論として蓄積されていない。インタラクションが、人というモデル化しづらい対象を含んだ現象である事も理論体系が構築されない理由の一つであると考えられる。人とエージェント間のインタラクション (HAI) の研究でも、同様に、様々なエージェントのデザインの提案やインタラクションの事例の報告が行われている一方で、体型だった理論の研究は行われていないのが実情である。しかしながら、HAI 研究は、デザイン自由度の大きいコンピュータとは異なり、エージェントといった自律性を持つ認知主体と人とのインタラクションを対象としている。認知主体同士のインタラクションとは何なのかを理論的に明らかにすることで、人と関わるエージェントのデザイン自体の体系化が可能であると思われる。本稿では、人とエージェント、環境間の情報の流れの観点から HAI のモデル化を検討する。

認知主体は、環境から必要な情報を能動的に取り出し、行動を行う。さらに、他者とコミュニケーションすることで、情報を伝えたり、情報を共有したりすることができる。本稿では、人とエージェント、環境の三つの中で生じる情報の流れのパターンを考察することで、エージェントの知覚行動および、エージェントから人へのメッセージが、人にどのように伝わるのかをモデル化する。

人・エージェント・環境間の情報の流れを整理することで、どのようなタイプのインタラクションがどのような条件で発生するのか、ロボットの行動や発話が人にどのような影響を与えるのか、どのようなエージェントをデザインしたらいいのかをインタラクションのタイプ別に考える事が可能になる。また、情報の流れに詳細な世界モデルを与える事で、シミュレーションによる HAI 研究も可能になると考えられる。

本稿では、HAI における情報の流れをモデル化するためにインフォメーションフロー [Barwise 97] を用いる。インフォメーションフローは異なるロジックを持つモデル間に生じる情報の流れを表す理論である。人・エージェント・環境といったそれぞれ異なるロジックを持つ物の間の情報の流れを記述することが可能である。本稿では、インフォメーションフローの詳

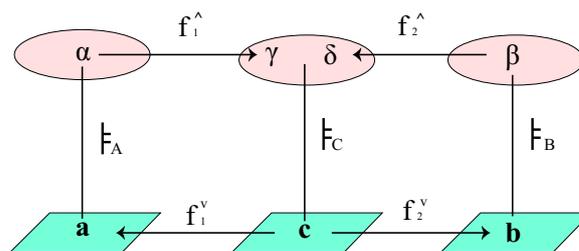


図 1: Channel on Information flow

細に触れるのではなく、HAI における情報の流れを分類するために必要な側面に焦点をあて説明を行う。個別の HAI をモデル化したい研究者や、ある条件化での HAI をシミュレーションしたい研究者は、本稿では触れなかったインフォメーションフローにおける詳細な記述方法を別途学ぶ必要がある。本稿では、人・エージェント・環境の情報の流れのモデルを、主体認知チャンネル理論 (Theory of channels for recognizing an agency) と呼ぶことにする。

## 2. Information flow

インフォメーションフローでは、異なるロジック (局在論理, local logic) 間の情報的な関係 (同一情報の表現, 情報の伝達) を記述することができる。本節では、主体認知チャンネル理論を理解する上で必要となるインフォメーションフローの概略を説明する。詳細は、[Barwise 97] をお読み頂きたい。

局在論理は、ある分類域 (classification) 上に成り立つ帰結関係である。分類域  $A$  とは、 $A = \langle tok(A), typ(A), \models_A \rangle$  の三つで構成される。 $A$  のトークンの集合  $tok(A)$ 、 $A$  のトークンを分類する対象の集合  $typ(A)$ 、 $tok(A)$  と  $typ(A)$  の間の二項関係  $\models_A$  である。分類域は、 $A$  のあるトークン (ある状況  $a$  もしくは、ある世界  $a$ 、ある物体  $a$ 、ある人物  $a$ ) において、 $A$  のある分類タイプ (ある出来事  $\alpha$ 、ある事実  $\alpha$ 、ある物体のタイプ  $\alpha$ 、ある人物のタイプ  $\alpha$ ) が成立するというものである。分類域は、可能世界における世界モデルと論理式、状況意味論における状況と状況タイプをさらに一般化した物である。

局在論理は、ある分類域の上のタイプ間の帰結関係を構成する。局在論理は次の論理構造を基本とする。 $A$  のタイプの集合

連絡先: 今井 倫太, 慶應義塾大学 理工学部, 〒 223-8522 横浜市港北区日吉 3-1-4-1, michita@ayu.ics.keio.ac.jp

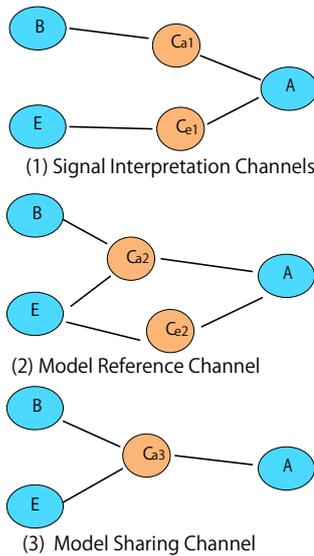


図 2: Channels for recognizing an agency

$\Gamma, \Delta$  があり, ある  $A$  のトークン  $a$  において,  $\Gamma$  に属する全てのタイプが  $a$  に当てはまれば,  $\Delta$  に属するすくなくとも一つのタイプに  $a$  が当てはまる. このような  $\Gamma, \Delta$  の対をシーケンと呼び,  $a$  はシークエント  $\langle \Gamma, \Delta \rangle$  を満足すると呼ぶ. ある局在論理  $L$  のシークエントは,  $\vdash_L$  と書かれ, また  $L$  の制約とも呼ばれる. タイプ間の関係を  $\gamma \vdash_L \delta$  と記述できる (あるトークン  $c$  がタイプ  $\gamma$  に当てはまるとき,  $c$  は  $\delta$  に当てはまる).

異なる分類域間の関係は, 情報同型写像 (inforphism) によって定義される. 情報同型写像は, タイプ間の写像  $f^\wedge$  とトークン間の写像  $f^\vee$  の対である. 図 1 の分類域  $A$  と  $C$  は, 情報同型写像  $f_1$  によって定義され,  $f_1^\wedge(\alpha) = \gamma, f_1^\vee(c) = a$  が成り立っている. 分類域  $C$  と  $B$  の間も  $f_2$  により同様の関係がなりたつ.

図 1 に示す分類域  $C$  は分類域  $A$  と  $B$  を繋ぐチャンネルである. 一般にチャンネルは, 情報同型写像  $\{f_i : A_i \rightleftharpoons C\}_{i \in I}$  によって複数の分類域を結合することができる. 例えば, 分類域  $A$  をある山の状態を表すものとする. 山である事実「煙が上がる」がトークン  $a$  であるとする. トークン  $a$  (煙が上がる) は状況タイプ  $smoking$  に当てはまる ( $a \models_A smoking$ ). 煙を見た人 (図 1 の分類域  $B$ ) が, 山火事を推測するのか, 何かの合図 (戦国時代の狼煙) を知るのかは, その人が同調する制約によって異なる.

図 1 の  $C$  で  $B$  が同調する制約が,  $f_1^\wedge(smoking) \vdash_C f_1^\wedge(wildfire)$  であるのか,  $f_1^\wedge(smoking) \vdash_C f_2^\wedge(attack(enemy))$  によって,  $a \models_A wildfire$  という情報を持つのか,  $b \models_B attack(enemy)$  という行動を取るのが異なってくる.

### 3. 主体認知チャンネル

本稿では, 人が同調するチャンネル上の制約の観点から, 主体を認知するチャンネルを考える. HAI では, 自律的に振舞うエージェントを人がどのように捉えるかを研究してきた. 多くは, エージェントに対する人のスタンスの違い (エージェントを意図を持った主体と捉えるか, デザインされた機械と捉えるか等) で HAI の質が変わる事が明らかになりつつある. 一

方で, 具体的にエージェントの振る舞いにどのような意味を人が見出すのか, エージェントと人との間でどのような情報が交わされるのかという観点からモデル化する試みは少ないように思われる. 主体認知チャンネルでは, エージェントとインタラクションする際に人が同調する制約を整理し, どのような情報が伝わるのかについて考察する.

人とエージェントが, ある場所に居る状況を想定する. 人の認知における分類域  $A$ , エージェントの認知における分類域  $B$ , 環境における出来事の種類域  $E$  の間のチャンネルについて考えることにする (図 2). 図 2 は, 人・エージェント・環境間のインタラクションで人が持つチャンネルの三つのパターンを列挙してある. ただし, 図は分類域とチャンネルのつながりを表すものであり, それぞれのノードはトークンとタイプを含むものとする.

一つ目は, 人がエージェントの発話や振る舞いを解釈するチャンネル  $C_{a1}$  と, 環境の出来事に対するチャンネル  $C_{e1}$  を独立に持っている場合である. これをシグナル解釈チャンネルと呼ぶことにする. シグナル解釈チャンネルでは, エージェントの発話や振る舞いを, 周囲の環境とは無関係に人が解釈・理解する情報の流れを示している.

二つ目は, エージェントと環境を組としてエージェントの発話や振る舞いから情報を得るチャンネル  $C_{a2}$  である. これをモデル参照チャンネルと呼ぶことにする. 環境に対するエージェントの振る舞いのモデルに同調した制約を人が持っている状態を指す. 一見, モデル参照チャンネルでは, 環境に対して機械的に反応をするエージェントを扱うものに見える. しかし, モデル参照チャンネルで, 意図的な振る舞いと機械的な振る舞いのモデルを区別する意味は無い. あくまでも人がエージェントの発話や振る舞いを理解する際に, エージェントが持つモデルを利用する事を表す.

三つ目は, 人とエージェントがお互いのモデルを相互に参照し, エージェントの発話や振る舞いを解釈理解するチャンネル  $C_{a3}$  である. これを, 共有モデルチャンネルと呼ぶことにする. 共有モデルチャンネルでは, 人が理解・解釈したエージェントの発話・振る舞いが, エージェント側にも伝わる. さらに, エージェントが, 人の発話や振る舞いを理解していることについて人が気づいているその事自体をエージェントが情報として持つ事も許している.

HAI 研究では, 意図を持つと思わせる振る舞いをするエージェントを構築する事を目的とするものがある. エージェントの意図を人に気づかせるだけでは, モデル参照チャンネルで可能である. 共有モデルチャンネルでは, ただ単にエージェントが意図を持つ振る舞いをするだけではなく, 人が知覚しているエージェント象自体に反応する能力が求められる.

### 4. インタラクション例

主体認知チャンネルの具体例として, 移動ロボットについて考える. 移動ロボットには, 掃除ロボットルンバ, ITACO ロボット [Ono 00], 自律移動椅子 [寺田 07] を想定している. ルンバは, 部屋を掃除する事が目的のロボットであり, 人とのコミュニケーションは, 人からルンバへの命令や, ルンバのエラー状態を人に知らせることに限定されている. もし, ルンバを購入したばかりであり人がルンバの動作のモデルを持っていない場合, シグナル解釈チャンネルで, ロボットのピープ音を人は解釈することが想像できる. しかしチャンネル  $C_{a1}$  の下ではピープ音自体からは, 何を意味しているのか人は知ることができない. 一方で, もし人がルンバの動作を熟知している場合,

ルンバのピーブ音からモデル参照チャンネル  $C_{a2}$  を通してルンバが障害物にスタックしている状況  $type(E)$  を知ることができる。ルンバ自体は、人の振る舞いを理解する機能を持たないので共有モデルチャンネルは生じない。

ITACO ロボット [Ono 00] の研究では、移動ロボットがゴミ箱にスタックした際にゴミ箱を移動するように人を依頼する実験を行った。その際に、ロボットと関係性を気づいていない実験被験者が、ロボットの合成音声を理解できないという状況が発生した。この現象は、シグナル解釈チャンネル  $C_{a1}$  で理解することができる。自分とは関係の無い移動ロボットが突然現れ、不明瞭な合成音声で発話した場合、ロボットのモデルを参照するチャンネル  $C_{a2}$  はなく、音声自体を直接解釈しようとする。そして不明瞭な音のために理解に失敗することになる。一方で、ITACO ロボットのエージェントマイグレーション機能\*1によって被験者がロボットと関係を持つ場合は、被験者はロボットの発話を理解することができている。この現象は、ユーザが意識的にロボットの周囲状況を観察するなかで、ロボットの前進したいという内部モデルにモデル参照チャンネル  $C_{a2}$  を通して被験者が同調できた結果であると考えられる。ITACO ロボットも、ロボット側が人のモデルを参照していないのでモデル共有チャンネルは発生しない。

寺田らが研究している自律移動椅子 [寺田 07] の実験では、 $C_{a1}$ ,  $C_{a2}$ ,  $C_{a3}$  の全てが生じており興味深い。被験者が部屋に入ると、被験者に座って欲しい事を椅子が動きで表現する。突然動き出した椅子の動きは、被験者にとって意味不明であると思われる。シグナル解釈チャンネル  $C_{a1}$  を通して椅子の動きのタイプを直接見つけようにも見つける事ができない状態である。しかし、暫く椅子の動きを観察すると、被験者の動きや周囲の机に反応しながら動く椅子に何らかのモデルの存在を感じモデル参照チャンネル  $C_{a2}$  で椅子の動きのタイプを見ようとする。さらに観察を続けると、被験者の行動を読むかのように動く椅子に対して、椅子と被験者がお互いの行動を参照し、モデルの共有が生じていると感じると思われる。お互いのモデルの共有は、共有モデルチャンネル  $C_{a3}$  の元での情報伝達であり、チャンネルを通してお互いが知り得た情報は、相手も一緒に気がついていてと知覚している状態である。共有モデルチャンネル  $C_{a3}$  で得られる情報は、相手が気づいていることに自分も気づいていることを表現する入れ子構造のタイプで表現される。

一方、人同士のコミュニケーションについて考察する。人同士のコミュニケーションを共有モデルチャンネル  $C_{a3}$  で全て説明可能かと言うとそうではない。例えば、ステージの壇上の人のアナウンスや教壇の上の教員の話は、モデル参照チャンネル  $C_{a2}$  による情報伝播である面が強い。聞き手は、壇上の人の話を、相手のモデルの想定元で聞くことになる。一方で、聞き手側が壇上の人に参照されていると思う意識は少ない。ゆえに共有モデルチャンネル  $C_{a3}$  ではなく、モデル参照チャンネル  $C_{a2}$  が支配した情報伝達である。また、赤の他人の発話や振る舞いは、特段の理由がなければ真剣に理解する必要がないので、相手に注意を向けたとしてもシグナル解釈チャンネル  $C_{a1}$  による情報伝播になる場合が多いと思われる。

## 5. まとめ

本稿では、インフォメーションフローをベースに、HAI で生じる現象を主体認知チャンネルという形で整理した。エージェントの発話や振る舞いをシグナルとして捉え、それのみから情報を抽出するシグナル解釈チャンネル、エージェントと環境のインタラクションのモデルの観点からエージェントの発話や振る舞いの意味を抽出するモデル参照チャンネル、人、エージェント、環境間のインタラクションにおける相互のモデル共有を基本とした共有モデルチャンネルの3種類のチャンネルでHAI研究の事例の解釈を行った。

今後は、各チャンネルを考慮した情報抽出や推論の詳細を定式化することで、HAI で起きるインタラクション現象内で具体的にやりとりされる情報の伝播を解析する予定である。また、HAI において究極の姿でもある共有モデルチャンネルの実現方法を、チャンネルが構成される過程をより詳細に追う事で考察していく予定である。

## 参考文献

- [Barwise 97] Barwise, J., Seligman, J. Information Flow: the Logic of Distributed Systems. Cambridge, UK: Cambridge University Press.(1997).
- [Ono 00] Ono, T., Imai, M, Reading a Robot's Mind: A Model of Utterance Understanding based on the Theory of Mind Mechanism, Proceedings of AAAI-2000, pp.142-148, (2000).
- [寺田 07] 寺田 和憲, 社本 高史, 伊藤 昭: 心の理論の枠組を利用した人工物から人間への意図伝達, ヒューマンインタフェース学会論文誌, Vol. 9, No. 1, pp. 23-22 (2007).

\*1 ユーザとすでに関係を形成したCGキャラクターがロボットのディスプレイに移動して表示され、ユーザとロボット間の関係形成を助ける。