相互信念モデルに基づくインラクションシミュレーション

Modeling and Simulation of Human Interactions based on Mutual Belief

菅野太郎 [1]　Tom Hope[2]　飯塚勝哉 [3]　古田一雄 [1]

Taro Kanno[1] , Katsuya Iiduka[2], and Kazuo Furuta[1]

[1] 東京大学大学院工学系研究科
[1]Department of Systems Innovation, the University of Tokyo
[2] 産業技術総合研究所
[2]National Institute of Advanced Industrial Science and Technology
[3] 東京大学大学院新領域創成科学研究科
[3] Graduate School of Frontier Sciences, the University of Tokyo

Abstract: This paper presents the modeling and simulation of human-human interaction based on a concept of mutual beliefs. The proposed model captures the four important aspects of human interactions: beliefs structure, mental states and cognitive components, cognitive and inference processes, and metacognitive manipulations. We implemented the model with Bayesian belief network and carried out some test simulations. The result shows that some basic aspects of human interactions as well as the effectiveness of mutual beliefs could be well simulated. We conclude by discussing the possibility of the application of this model and simulation to human-agent interaction.

1. 緒言

HAI（Human Agent Interaction）を実現する上で人同士のインタラクションに学ぶべきことは少なくなかろう。一方、人同士のインタラクション研究においてもそのメカニズムが解明されているとは言えない。本研究では HAI への応用を視野に入れつつ、人同士の協調におけるインタラクションの認知メカニズムに焦点をあて、相互信念に基づく協調行動における認知のモデリングとモデルに基づくコミュニケーション生成の計算機シミュレーションの開発を行った。

筆者らはすでに相互信念の概念を導入した人の協調行動における認知モデルの概念モデルを提案している[1]。次節ではその概念モデルについて紹介するとともに、協調タスク遂行時における人同士のインタラクションの駆動パターンを実験室実験における発話分析によって抽出し概念モデルを用いて整理した拡張認知モデルについて説明する。３節では、ベイジアンネットワークを用いた拡張モデルの実装について説明し、４、５節で二人組の協調行動を想定したインタラクション（コミュニケーション）生成の計算機シミュレーションの例を示す。６節で HAI への拡張可能性について議論し、本稿をまとめる。

2. チーム認知モデル

発達心理学における心の理論[2]や、哲学における協調行動における意図[3,4]（we-intention や joint-intention、 collaborative intention）に関する議論や reflexivity モデル[5]、脳科学におけるミラーニューロンの発見、これらの理論や生理学的知見が示す人間の認知行動の特質は、人には再帰的に他人の心的状態を理解する仕組みが備わっているということである。我々はこのような再帰性を相互信念の概念を用いて人の様々な心的状態や認知プロセスに適用することで、図１に示すような協調行動における認知（チーム認知）の概念モデルを提案した。

図１の各層は上から、１）主体となる人（エージェント）の認知（モデル）、２）相手の認知に対する信念（モデル）、３）相手の自分の認知に対する信念（モデル）、をそれぞれ表しており、各層には状況認識（SA）や知識、メンタルモデル、感情といった様々な心的状態や認知的コンポーネントを含み得る。各層には、例えば１層目では知覚-状態認識-意図形成-行動といった認知プロセス、２、３層の信念層では信念を獲得するための推論プロセスがあり、また、各層を参照・比較し、各プロセス・状態を修正、補完するといったメタ認知的操作も存在する。これら

の１）相互信念構造と２）心的状態・認知コンポーネント、３）認知・推論プロセス、４）メタ認知操作、から概念モデルは構成され、これらの各状態、プロセスの結果を基にインタラクションが駆動される様をモデル化している。

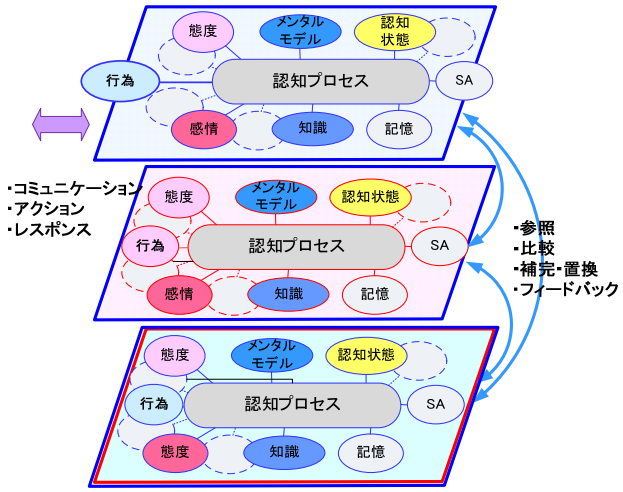

図 1 チーム認知モデル

2.1 インタラクションジェノタイプ

本節ではインタラクションの駆動理由について説明する。HAI におけるインタラクション生成の実現や、その基礎となる発話・インタラクションの生成プロセスを考える上で発話内容や発話の機能（遂行動詞）といった観察可能な要素とそれを発する理由・原因、機構となる観察できない要素を区別することは重要である。本研究ではそれぞれインラクション Phenotype、Genotype と呼ぶ。概念図を図 2 に示す。上二段が Phenotype、下二段が Genotype にあたる。どのように表現したのかといった発話内容や、どう伝え、相手に働きかけたかという発話の機能は Phenotype に類される。Genotype には Phenotype に至った理由やそのための認知プロセス、それを支える脳のメカニズムや、パースペクティブテイキングや感情の伝染作用といった様々な協調装置[6] が類される。

所謂人間同士のインタラクションやヒューマンマシン研究における発話・会話分析や振舞い分析は上二段を主に扱うもので、Genotype を分析するためには別途実験後のインタビューや生理実験といった他の手法が必要となる。先行研究で二人組でゲームを行う協調実験を行い。実験中のビデオ・音声記録を行い、さらに実験後に実験中のビデオを被験者に見せ、インタラクション（コミュニケーション）の背後にある理由を相互信念の観点から説明させたり、観察者に背後理由を推論させたりすることでインタラクション Genotype の抽出・分類を行った[7]。結果を表 1 に示す。

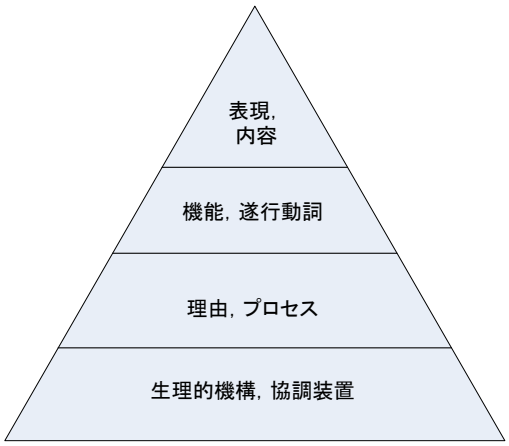

図 2 インタラクションにおける
Phenotype と Genotype

表 1 インタラクション Genotype

Genotype		Phenotype
分類	理由	(遂行動詞)
1。 自身の認知プロセスや相手の心的状態・プロセスの推論のため	- 必要・十分な情報や知識がない - 相手の心的状態・プロセスの推定に自信がない、など	Query Confirm
2。 相手の認知プロセスや推論プロセスを助けるため	- とにかく共有しておく - 相手が必要・十分な情報がない - 前もって情報を提供しておく、など	Inform
3。 相手の認知プロセスや推論を修正するため	- 齟齬の回避・回復 - 誤解の訂正、など	Inform Query Confirm

3. モデルの実装

3.1 認知・推論プロセスと認知コンポーネント

不確かで限定的な情報に基づく状況認識における

人間の認知・推論を扱うために、本研究では認知・推論プロセス、つまり提案モデルの各層をベイジアンネットワーク(BBN)を用いて実装した。後述するシミュレーションでは、例題として車の故障診断に関する認知・推論プロセスを用いた。BBN の各ノードは観測可能な兆候および構成機器の状態（に関する認知・推論）を、リンクはノード間の因果関係を表している。各ノードの持つ確率は事象生起に対する確信度を表し、これによって確率値で認知・推論プロセスを定量的に扱うことができる。意識的に事象生起を認識しているノードの集合を U とし以下の式で与える。ここで P_i は事象 i への確信度を、T は「生起認識」のための確信度の閾値を表している。

$$U = \sum_i \{i \mid P_i \geq T\} \quad (1)$$

BBN のネットワーク構造、ノード間の条件付き確率分布、ノード事象の事前確率分布の 3 つの要素を各主体（エージェント）が持つ知識とみなすことができる。シミュレーションでは、知覚やコミュニケーションから得られる入力情報や層間の相互作用によってこの確信度を操作し、推論を行うことで、各層の認知プロセスを模擬する。

3.2 相互作用

自分の認知プロセスと相手の認知プロセス・状態に関する推論が独立して行われているとは考えにくい。２節で説明したように、各プロセスの結果を参照、比較するといったメタ認知的操作を通した認知・推論結果の修正や、無意識的に自身の認知と推論結果と置き換える（例えば自分が見たものを相手も見たものと思い込む、またはその逆（補完）など）といった効果もあると考えられる。このような層間の相互作用も各 BBN の対象ノードの確信度の値を以下のように操作することによって模擬する。

思い込み : $P_i = \alpha P_1$（α は影響度） (2)
補完 : $P_1 = \beta P_2$（β は影響度） (3)

3.3 コミュニケーション生成

表 1 に示すように実験からインタラクション Genotype として３パターンが確認された。本研究ではそのうち表 1 の分類 3 の Genotype を実装した。すなわち、各層の生起認識事象の集合をそれぞれ U_1、U_2、U_3 とすると、これらの比較によってコミュニケーションを生成させる。コミュニケーション生成ルールを以下に示す。(4)は自身の１層と２層目の生起認識事象が異なり相手の認識が間違っていると考えられる場合は、自身の認識事象を伝達することで相手の１層目の修正を試みることを表し、(5)は相手の自分に対する信念が誤っていると考えられる場合は、自身の認識事象を伝達することで相手の２層目の修正を試みることを表す。

$U_1 \neq U_2 \wedge U_2$=false *then* Inform(U_1) to Modify(U_{1b}) (4)
$U_1 \neq U_3$ *then* Inform(U_1) to Modify(U_{2b}) (5)

4. シミュレーション

二人のエージェントA、Bによる車の故障診断を例題に挙げ、各エージェントが持つ特性とモデルに基づくコミュニケーションによってチームの認知（状況認識共有）がどのように変化するかをシミュレートした。今回のシミュレーションでは、エージェント A、B の持つ知識（ネットワーク構造、確率分布）は同じと設定し、A、B の各層の認知プロセスを図 3 のような BBN で実装した。

９つのノードを観察可能な徴候ノードと設定し、徴候ノードに関する情報をコミュニケーションで伝達するものとした。徴候の観察パターンはエージェント毎にシナリオベースで記述し、各エージェントが断片的に情報を収集しながら互いにコミュニケーションを行うことで、二者間の共有情況認識が醸成されていく様が模擬できる。

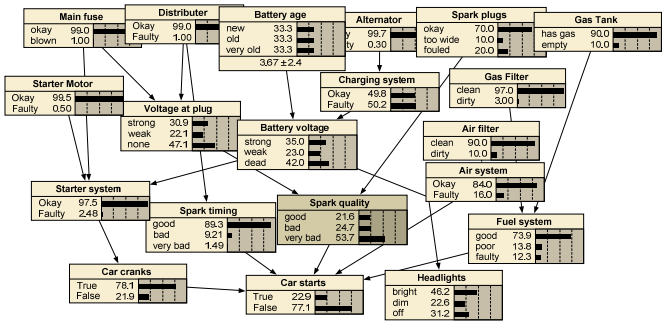

図 3 故障診断における認知モデルの BBN

4.1 シミュレーション設定

シミュレーションのパラメータとして、思い込みや補完といった認知操作におけるパラメータ（層間の影響度）、各認知操作におけるエラー率、認知操作・コミュニケーションの実行率、知識の 3 つを設定した。これらのパラメータでエージェントの個性を表現できる。今回のシミュレーションでは、エージェント A、B の個性は均質とし、パラメータを表 2 のように設定した。

表 2　パラメータ設定

パラメータ	内容	番号	値
知覚による情報入力	1 層目への入力	1-1	90%
	2 層目への入力	1-2	70%
	3 層目への入力	1-3	50%
思い込み	1→2 層目	2-1	20%、40%、60%
	1→3 層目	2-2	2-1＋20%
	1→2(コミュニケーション後)	2-3	2-1×1。5 (max は 100%)
	1→3(上と同様)	2-4	2-2×1。5
補完	U_1=U_2	3-1	40%
	U_1≠U_2で修正	3-2	3-1×2
実行率	補完	4-1	100%
	コミュニケーション	4-2	100% 、 75% 、 50%、25%
エラー率	―	5	0%、5%、25%

4.2　シミュレーションの流れ

シミュレーションの概要を図 4 に示す。シナリオで規定されている行動を実線四角（赤色）で、相手とのコミュニケーションを破線四角（青色）で、主体内部のメタ認知や層間の相互作用を破線丸(灰色)で表している。コミュニケーションによって相手から獲得する情報の入力はそれぞれ実線四角で表される操作に相当する操作を行う。図 4 に示すサイクルをシナリオで規定されたステップ行いシミュレーションは終了する。シミュレーション結果として各エージェントのコミュニケーション履歴と各層の各ノードの確信度の時間変化を得る。

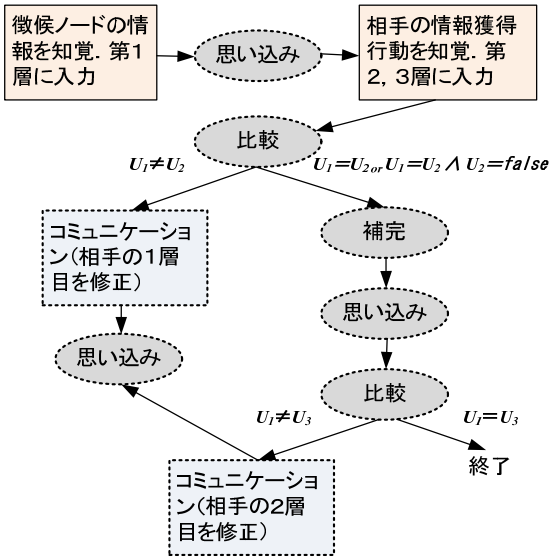

図 4 シミュレーション概要

4.3　評価指標

インタラクションの評価は状況認識の共有の程度で測る。相互信念を導入した場合、様々な共有の定義が可能となる[8]。本研究では、i) 個人の 1 層目の認知の正解度の平均(各個人が実際に起こっている状況をどの程度正しく認識しているか) と、ii) 自分の信念と相手の認知の共有度、iii) 正解度と共有度の平均で評価する。正解度と共有度 は以下の式(6)、(7)で表わされる。U_{xn}は、エージェント x の n 層目の「生起認識」集合を表す。また、U_0は実際に生起している事象の集合を表す。

$$\text{正解度} = \left(\frac{U_{A1} \cap U_0}{U_0} + \frac{U_{B1} \cap U_0}{U_0} \right) \times \frac{1}{2} \qquad (6)$$

$$\text{共有度} = \left(\frac{U_{B1} \cap U_{A2}}{U_{B1}} + \frac{U_{A2} \cap U_{B1}}{U_{A2}} \right) \times \frac{1}{2} \qquad (7)$$

(信念の網羅性)　(信念の正確さ)

5. 結果と考察

5.1　結果

5.1.1 パラメータ設定とインタラクションの評価

表 3 は、エラー率、「思い込み」の影響度、コミュニケーション実行率を変化させた際のインタラクションの評価結果を示している。一般的にコミュニケーション実行率が低いほど正解度が高い反面、共有度が低くなり、逆に実行率が高いほど共有度は高くなる傾向が観察された。「思い込み」の影響度に注目すると、影響度が小さいとき（20%）は、コミュニケーション実行率が極端に高かったり低かったりした場合に共有度が低下している。

5.1.2 コミュニケーション意図と解釈

図 5 はエラー率 5%、「思い込み」の影響度 40%、実行率 75%における相手の認知に対する信念の網羅性と正確さの時間変化を、表 4 はその際実行されたコミュニケーションの詳細を示している。

タイムステップ 2、3 においてコミュニケーションの意図が読み取れず、意図とは違う解釈を行うことで信念の網羅性が下がっていることがわかる。また、タイムステップ 14、15 では、受け手は意図に沿った解釈を行ったがそのフィードバックがなく送り手の信念の網羅性が低下している。その後タイムステップ 15、16 で、A が 3 層目を働かすことにより齟齬が発見され網羅性の低下を回復している。

表 3　チーム認知の評価

		コミュニケーション実行率			
		100%	75%	50%	25%
エラーなし 影響度 2-1 40%	正解度	0.67	0. 70	0. 74	0. 78
	共有度	0. 98	0. 96	0. 96	0. 92
	平均	0. 82	0. 83	0. 85	0. 85
エラー率 25% 影響度 2-1 40%	正解度	0. 23	0. 17	0. 21	0. 29
	共有度	0. 91	0. 83	0. 91	0. 83
	平均	0. 57	0. 50	0. 56	0. 56
エラー率 25% 影響度 2-1 60%	正解度	0. 31	0. 31	0. 58	0. 53
	共有度	0. 88	0. 91	0. 86	0. 87
	平均	0. 59	0. 61	0. 72	0. 70
エラー率 25% 影響度 2-1 20%	正解度	0. 21	0. 21	0. 31	0. 35
	共有度	0. 83	0. 90	0. 85	0. 73
	平均	0. 52	0. 55	0. 58	0. 54

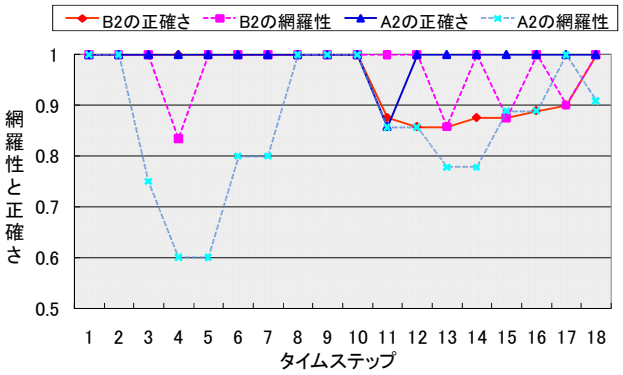

図 5 網羅性と共有度の変遷

表 3　コミュニケーションの詳細

タイムステップ	意図と解釈
2	AはBの2層目の修正を意図しコミュニケーション
3	Bは1層目を修正、B1-A2でずれが発生（Aの意図通りに解釈せず）
14	BはAの1層目の修正を意図しコミュニケーション
15	Aは意図通りに解釈したが、BがAの解釈を未確認。A1-B2でずれが発生
16	Aは1層目と3層目の比較からA1-B2のずれに気づき、Bの2層目を修正を意図しコミュニケーション→回復

5.2　考察

コミュニケーション実行率が低いと、各自の状況認識は外部環境からの情報獲得行動により依存し、認識におけるエラーが少なければ正解度は上がる。一方、相手に対する信念形成が個人内で閉じているため状況認識の共有は促進されない（共有度低）。シミュレーション結果はインタラクションにおける基本的な現象がよく模擬されているといえる。シミュレーションで行ったパラメータ操作は被験者を用いた実験等で制御することは実質不可能であり、シミュレーションの有効性を示している。

各コミュニケーション実行率における正解度の差はエラー率が高くなるにつれ縮まり、コミュニケーション実行率が高い方が正解度は高くなると予想されるが、今回のシミュレーションではこのコミュニケーションの効能は観察されなかった。これは、現実的には、一度判断を誤った箇所(エラーが起きた箇所)は、自分ではなかなか修正が困難でコミュニケーションによる訂正が期待されるが、本モデルではエラーがランダムに与えられているため、一度判断を誤った箇所も、次ステップで判断エラーがなければ、すぐ修正されてしまうためである。この結果、実行率の低い場合の正解度が現実的な値より高くなっていると考えられる。また、「思い込み」の影響度が小さいときは、コミュニケーション実行率が低いと一度起きた推論エラーがなかなか回復されず、コミュニケーション実行率が高すぎると、何度も相手の情報を確かめる不要なコミュニケーションを行うためどちらも信念の共有度が低くなっていると考えられる。これらの点においてモデルの改良が必要である。

また、図 4 の結果からは、意図とは違う解釈が信念の共有度の低下を招くということだけでなく、意図に沿った解釈を行っても、必要に応じて自分がどのような解釈を行ったかを相手に伝えないと信念の共有度が低下する可能性があることがわかった。これを防ぐには、伝えられた情報を解釈するだけでなく、コミュニケーションの意図を正確に読み取り、コミュニケーションのフィードバックを何らかの方法で話者へ伝えること（伝わること）が必要であろう。さらに、図 5 のタイムステップ 15 から 16 の結果が示すように、第 3 層目を働かせ 1 層目と「比較」を行うことにより、状態認識共有の低下が回復されることも示唆される。

6. 結言

提案モデルを用いて、コミュニケーションの意図と解釈を考慮したシミュレーションを行い、エージ

ェントの特性に応じた状況認識共有への影響を模擬できる可能性を確認した。シミュレーション結果から、コミュニケーション意図に沿った解釈だけでなく、その意図を読み取り必要に応じて解釈を相手に伝える重要性、即ちクローズドループコミュニケーションの重要性や、人間の冗長なチーム協調に相手の自分に対する信念を推論することが寄与することが示唆された。 今後は、エージェントの知識構造の差異やより現実的なエラーシナリオなどの他の特性を十分考慮した上で、エージェントの特性がチーム認知に与える影響を詳細に分析することが課題となる。

また、コミュニケーション生成のメカニズムは HAI における Agent のインタラクション生成のメカニズムへの応用が期待できる。HAI の場合、各層に相当するモデルが人-人インタラクションとは異なり Agent に対するメンタルモデル等が必要とされる。さらには人が Agent に対して再帰的な信念を抱くか否かもヒューマンライクなインタラクション生成の鍵となるであろう。いずれにせよ、インタラクションの Phenotype と Genotype を分類して整理することは高等なインタラクションを生成する上で有益な足がかりとなろう。また本研究で用いた相互信念による様々な共有指標は HAI を評価する際の指標としても適用することが可能であろう。

謝辞

本研究の一部は財団法人日産科学振興財団の助成によって行われた。

参考文献

[1] Kanno T. and Furuta K., Sharing Awareness, Intention, and Belief, Proc. 2nd Int. Conf. Augmented Cognition, pp.230-235 (2006)

[2] S. Baron-Cohen, Mindblindness: An Essay on Autism and Theory of Mind, MIT Press (1995)

[3] R. Tuomela and K. Miller, We-intentions, Philosophical Studies, Vol.53, pp.367-389 (1987)

[4] M.E.Bratman, Shared Cooperative Activity, the Philosophical Review,101, pp.327-341 (1992)

[5] Lefebvre A.V. Research on Bipolarity and Reflexivity, The Edwin Mellen Press (1992)

[6] Malle F.B. and Hodges D. S., Other Minds, the Guilford Press (2007)

[7] Kitahara Y., Hope T., Kanno T., and Furuta K., Developing an understanding of genotypes in studies of shared intention, Proc. 2nd Int. Conf. Applied Human Factors and Ergonomics, CD-ROM (2008)

[8] Kanno T., the Notion of Sharedness based on Mutual Belief, Proc. 12th. Int. Conf. Human-Computer Interaction, pp.1347-1351(2007)