

相互信頼感形成の共関心擦り合わせモデル A Concern Alignment Model for Interactive Trust Management

片桐恭弘¹⁾, 石崎雅人²⁾, 伝康晴³⁾, 高梨克也⁴⁾, 榎本美香⁵⁾, 岡田将吾⁶⁾
Yasuhiro Katagiri, Masato Ishizaki, Yasuharu Den, Katsuya Takanashi,
Mika Enomoto, Shogo Okada

¹⁾ 公立はこだて未来大学, ²⁾ 東京大学, ³⁾ 千葉大学, ⁴⁾ 京都大学,

⁵⁾ 東京工科大学, ⁶⁾ 東京工業大学

Future University Hakodate, The University of Tokyo, Chiba University, Kyoto University, Tokyo University
of Technology, Tokyo Institute of Technology

katagiri@fun.ac.jp

1. はじめに

対話はわれわれの日常生活で他者と合意を形成して協調的行動を実現するために中心的役割を果たす。合意はあるグループの人々が将来のある時点である共同行為を遂行するという選択に対する共有されたコミットメントである[1]。共同行為の成功はそれぞれのメンバーが割り当てられた行為を適切に遂行することにかかっているため、合意形成には必然的にグループメンバー間での信頼がかかっている。そのために対話インタラクションでは、情報共有と合意形成に加えて、会話参加者間の信頼感構築および維持が起こっている。本稿では、複数エージェントによる協調行動を支える信頼について考察を加えた後、会話による合意形成を通じた相互信頼感構築プロセスを記述する関心擦り合わせ (concern alignment) 概念[2, 3]を紹介する。関心擦り合わせ概念に基づく特定健診保健指導対話の分析に基づいて、関心擦り合わせ過程の計算モデル化の可能性について議論する。

2. 協調行動を支える信頼

人が他者に対して抱く信頼について、山岸[4]は、囚人のジレンマゲームでは裏切り行為を選択することが合理的であるはずなのに、人間同士では協力をを選択することが多いという実験結果に基づいて、信頼を他者と協力的な協調行動を産み出すための社会心理メカニズムとして位置づけ、そのタイプ分類を提示している。

信頼を人間に備わる一種の心理的メカニズムと想定すると、信頼を「確とした根拠無しに他者が一定の行為を行うと想定し、その想定を前提として自己の行為を選択することを支える心的メカニズム」のように規定することができる。

信頼は、その対象あるいは段階に従って次のように分類できる。

(1) 能力に対する信頼

信頼の対象となる他者が問題とする行為を遂行する能力を有しているということに関する期待・想定。医者が有能である、人工物が壊れずに仕様通りに動作する等に関する信頼が含まれる。

(2) 誠実性に対する信頼

信頼の対象となる他者が問題となる行為を遂行する能力を備えていることは前提として、実際にその行為を遂行することに関する期待・想定。相手が約束を守るか、嘘をつかないか、自分を裏切らないか等のレベルでの信頼が相当する。

(3) 投資としての信頼

ある程度の失敗は覚悟の上で、長期的に協調行動をとっていくという方針をお互いに持っている、あるいは作り出せるという期待・想定。

本稿で取り上げるのはこれらの中で、(2) 誠実性に対する信頼である。

「信頼」のメカニズムは任意の他者に対して無条件に起動されるわけではない。まずはじめに、相手が信頼できるかどうかの判断が必要である。そのために、相手の外見から得られる情報、対話をはじめとする各種の相互作用を通じて得られる情報、第三者から得られる間接的な評判情報など多様な情報が収集され利用されている。さらに、対話のような他者との相互作用は、相手から自分に対する信頼を勝ち取るためのデバイスとしても機能している。情報の交換によって協調的行動を意図するのに必要な信頼を獲得した後は、合意した行動を実際に選択して遂行することを通じて、長期的な相互信頼感が維持される。

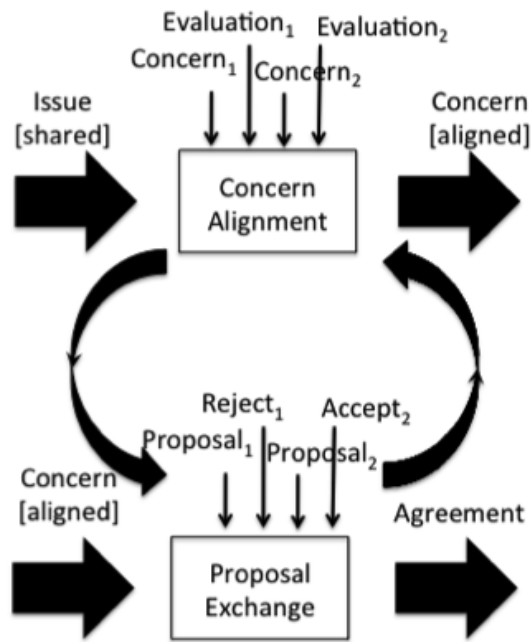


図1 対話を通じた信頼感構築における関心擦り合わせ過程

信頼を産み出す情報源

- (a) インタラクションに入る前に得られる外部的情報に基づく直観的判断
- (b) 単一のインタラクションセッションの中で獲得される信頼
- (c) インタラクションの繰り返しの蓄積として形成される安定した信頼

信頼を産み出す外部的情報

- (a-1) 第一印象に基づく信頼
- (a-2) 同一集団帰属認識に基づく信頼
- (a-3) 評判に基づく信頼

信頼判断にかかわる情報は多岐にわたるが、対話を通じて交換される情報は上記の分類では(b)に含まれる。

信頼を産み出す情報をこのように分類すると、最初に述べた囚人のジレンマゲーム状況は、相手に関する情報を人工的に最小化した上で行動判断を求めるものである。その中でも特に、上記(a)のしかも極めて貧しい情報のみで信頼を判断する状況設定となっている。囚人のジレンマゲーム状況では通常の場合での人間同士のインタラクションにある相互作用を通じた相手の情報は得られない。さらに、繰り返し型でない単発型であれば、行為選択の歴史の情報も利用できない。このように囚人のジレンマゲーム状況は人間同士の信頼を考える設定としては十分に適切とは言えないだろう。

信頼を他者の行動選択に関する期待ととらえると、それは一義的には他者行動の予測可能性に帰着される。エージェントの行動選択を、各エージェントが自分の持つユーティリティ構造に基づいて、ユーティリティを最大化する行為を選択する過程ととらえれば、インタラクションを通じて行動の予測可能性を高めるためにとられる行動としては、段階的に以下のようなものが想定できるだろう。

相手の行動選択の予測可能性要素

- (s-1) 相手の価値構造を知っていれば相手の行動選択の予測可能性が高まる
- (s-2) 自分の価値構造を知っていてそれに沿った行動をしてくれると予想できる
- (s-3) 自分の価値構造を推測してくれる

自分の行動選択の予測可能性を高める要素

- (o-1) 相手に自分の価値構造を伝えて自分の行動選択の予測可能性を高める
- (o-2) 相手の価値構造を知ってそれに沿った行動をとる
- (o-3) 相手の価値構造を推測する

3. 相互信頼感と対話: 関心擦り合わせ

対話インタラクションを通じて人々が信頼を構築する典型的設定として、合意形成過程を取り上げる。人々が合意形成のために対話を行う時には、情報共有、(将来の協調的行動に関する)合意の形成のふたつに加えて、相互信頼感形成が起きている。相互信頼感が形成されることによって、人々は、確とした根拠無しに他者が一定の行為、すなわち合意された協調的行動のうちの他者の分担分、を行うと想定し、その想定を前提として自己の行為、すなわち合意された協調的行動のうちの自己の分担分、を選択する。そのような選択の結果として協調的行動が実現する。

協調的行動が実現するにいたる上記の過程の中心となる合意形成を通じた相互信頼感構築の過程として、関心擦り合わせ(concern alignment)を想定する。

3.1 合意形成・相互信頼感の機能要素

会話進行における合意形成と相互信頼感構築の構造をとらえるために、会話内で談話イベントの果たす機能要素として、以下の3種類の概念を設定する。

論点(issue) 会話コミュニケーションの中で合意を形成すべき事柄。

関心擦り合わせ		
- C-solicit	:	concern solicitation
- C-introduce	:	concern introduction
- C-eval/positive	:	subjective positive evaluation
- C-eval/negative	:	subjective negative evaluation
- C-elaborate	:	concern elaboration
提案交換		
- P-solicit	:	proposal solicitation
- P-introduce	:	proposal introduction
- P-accept	:	proposal acceptance
- P-reject	:	proposal rejection
- P-elaborate	:	proposal elaboration

表1 関心擦り合わせの対話行為

関心(concern) 論点に関して会話参加者同士が何らかの合意を形成するに当って、会話参加者がそれぞれ有している主観的な価値判断基準。

提案(proposal) 合意の候補となる提案。

合意形成の過程では、会話参加者は個々の論点に関して自他の関心を考慮しながら、合意の候補となる提案を提示し、参加者間で吟味・交渉を通じて合意に至る。

グループでランチに行くという例を考えよう。ランチの場所を決めるという目標が論点(issue)となる。ある者はラーメン、和食など食事のタイプにこだわり、他の者は値段がいくらかを気にする。また別の者は場所が近いことを望むかもしれない。これらが参加者の関心(concern)である。関心の情報を交換した後に、具体的なレストランの名前が挙げられる。これが合意形成のための提案(proposal)に相当する。

現実の合意形成では具体的合意形成の議論に入る以前に、そもそも論点が何かという事自体について合意を形成する必要がある場合もある。その場合のように、論点・関心・提案をひとつの単位として合意形成過程が多層の階層構造を作る可能性もある。

3.2 関心擦り合わせの過程

論点・関心・提案の概念を想定すると、会話を通じた合意形成および相互信頼感形成過程は、概念的には図1に示すように、関心擦り合わせと提案交換との二つのステップに分けてとらえることができる[2, 3]。最初に論点が定まるとまず会話参加者は各自は自分の関心の提示と他者の関心提示に対する賛同・不賛同の反応表明を通じて、会話参加者それぞれの関心を把握する。これが第一ステップの

関心擦り合わせ過程である。次に、擦り合わせた関心に基づいて各自は具体的提案を提示して合意内容の交渉に入る。こらが第二ステップの提案交換である。

関心擦り合わせでは以下の3種類のプロセスが生起する。

関心導入 論点に対して会話参加者は自分が重要と考える要因を提示する。最終的に合意されるべき提案に含まれる行為の備えるべき特徴、性質の形をとることが多い。

関心に対する評価反応 関心が導入されると他の対話参加者はそれに対する正あるいは負の評価を反応を示す。評価反応は言語的表明だけでなく、非言語的に表現されることもある。

漸進的擦り合わせ 導入された関心が正に評価されて関心のレベルでの共有が達成されたならば、会話参加者は具体的な提案交換のステップに移ることができる。導入された関心が負に評価された場合には、関心を限定する、変形するなどの修正を加えて妥協点を探ることになる。

これらの関心擦り合わせと提案交換の過程を記述するために表1に示すような対話行為[5, 6]を設定する。

4. 関心擦り合わせ過程構造分析

対話による関心擦り合わせ過程を通じた相互信頼感構築を現実の事例に基づいて確認する目的で、特定保険指導対話の分析を行っている[7, 8]。特定保健指導では、健康診断によってメタボと判定された人々が、健康指導として保健師から生活改善のアドバイスを受ける。数人の受講者を対象とした講義形式の講習と、受講者ひとりずつを対象とした個別の対面指導とからなる。前者の講習では、保健師が受講者に対して肥満に関わる一般的な健康医学

A-B: C-introduce:(stop smoking)	⇒ C-eval/negative:(no intention)
A-B: C-introduce:(reduce smoking)	⇒ C-eval/negative:(already tried)
A-B: C-introduce:(use non-smoking pipe)	⇒ C-eval/negative:(tongue tingling)
B-A: C-introduce:(cost money)	⇒ C-eval/positive: (acknowledge)
B-A: C-introduce:(choose tobacco rather than eating)	⇒ C-eval/negative:(not good)
B-A: C-introduce:(consider when short on money)	⇒ C-eval/positive: (good)
B-A: C-introduce:(withdrawal syndrome)	⇒ C-eval/positive: (acknowledge)
B-A: C-introduce:(smoker communication)	⇒ C-eval/positive: (acknowledge)
↓	
A-B: P-introduce:(consider stop smoking when prices go up)	
B-A: P-accept: (stop smoking when prices go up)	

図2 特定健診保健指導対話における関心/提案の系列構造分析の一例

的情報を具体的に紹介する。その後、対面指導で受講者個別の事情に合わせたアドバイスが与えられる。具体的には、受講者の健康診断結果に基づいて生活改善の具体的目標および計画を、受講者と保健師との合意の下に作成し、それを文書にまとめる。さらに、一定期間後(2週間, 3ヶ月)に電話確認あるいは面接を設定して、文書に記した計画の実施を求める。

収録した保健士と保健指導受講者との対話に対して、表1に示す対話行為ラベル付与による構造分析を行っている。その結果の一例を図2に示す。この対話では、保健士と保健指導受講者との間で喫煙を巡る対話が展開されている。禁煙、減煙、禁煙パイプに始まり、喫煙所コミュニケーションにいたるまで、さまざまな話題が交換され、最後に保健士から提案が提示される。丁寧な関心擦り合わせがあったために、保健士の提案は保健指導受講者からすぐに受諾されている。実際、受講者はほとんど重なるように提案を繰り返して同意を表明した。

図2の構造分析において、禁煙、減煙、禁煙パイプに始まり、喫煙所コミュニケーションまで交換された要素は、保健士および受講者によるそれぞれの関心の表明と解釈される。それらの関心提示の直後には相手からの正あるいは負の評価が続いている。負の評価の場合には、たとえば禁煙を減煙に緩和するなど関連する関心に修正するという関心擦り合わせが起こっている。

このような過程を通じて関心の擦り合わせが起こり、その後「タバコの値段が上がったならば禁煙する」という提案が保健士から提示されて合意に至っている。この例では関心擦り合わせの部分と提案交換の部分との区分が明確である。

最終的に合意された提案は、生活改善という観点からするとかなり弱い内容である。おそらく関心擦り合わせの過程から、保健士は強い提案では合

意の可能性が低いという感触を得て、このように弱い内容の提案を選択したものと解釈できる。その結果として、受講者からの同意は容易に引き出すのに成功している。関心擦り合わせの結果、合意し易い(さらには実行し易い)と想定される内容の提案に落ち着く事ができたと言えるだろう。

図2に示す関心擦り合わせの対話構造分析によって、信頼感構築のための行為選択予測性を高めるために対話参加者がお互いに自分の価値構造の表明と相手の価値構造の推定を積み重ねている過程が捉えられている。

5. 関心擦り合わせを通じた相互信頼感構築のモデル化

関心擦り合わせ過程は、multi-issue negotiation[9]ととらえることができる。図3に示すように、対話インタラクションに参加するエージェントはそれぞれ自分の価値構造として、さまざまなissueに対する重要度評定を持っている。レストラン選択の例では、あるエージェントは料理タイプを重要としているのに対して、別のエージェントは値段を重要視している。両者が対話の過程を通じてまずそれらのissue重要度を関心(concern)として交換することによって、自らの価値構造を表明し、他者の価値構造の推定を行う。

5.1 関心情報交換

関心擦り合わせの中で関心情報交換部分は以下のように特徴付けられる。

- (c1) 特定のissueに関する自分の重要度評定を表明する。
- (c2) 相手のissue重要度表明から相手の価値構造を推定する。
- (c3) 相手のissue重要度表明に基づいて自分のissue重要度評定を変更する。

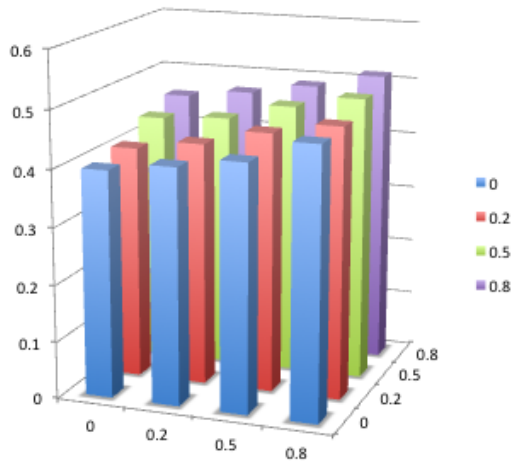
Agent A	
料理タイプ: 和食	High
料理タイプ: インド	Low
⋮	
値段: 高	Neutral
値段: 中	Neutral
値段: 低	Neutral
場所: 近所	Low
場所: 盛り場	Neutral
⋮	

⇔

Agent B	
料理タイプ: 和食	Low
料理タイプ: インド	Low
⋮	
値段: 高	Low
値段: 中	Low
値段: 低	High
場所: 近所	High
場所: 盛り場	Low
⋮	

図 3 Multi-Issue Negotiation としての関心擦り合わせ過程

One-shot alignment



Complete alignment

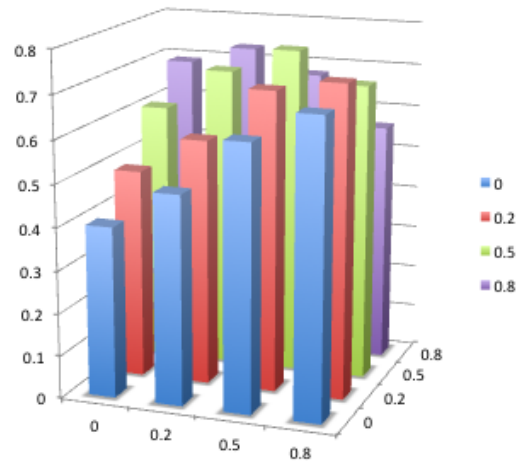


図 4 関心擦り合わせによる合計相対利得

すると関心擦り合わせの中の関心情報交換は以下のようにモデル化可能である。

I : issue の集合

\mathcal{V} : 価値構造の集合

エージェント i の価値構造 v_i

$$v_i \in \mathcal{V} : I \rightarrow \mathcal{R}$$

\mathcal{A} : 行為の集合

u : エージェントの行為に対するユーティリティ

エージェント i の行為 a に対するユーティリティ

$$u : \mathcal{V} \times \mathcal{A} \rightarrow \mathcal{R}$$

(c1) の特定の issue に関する自分の重要度評定の表明は、エージェント i が issue s に対する価値 $v_i(s)$ を提示することに相当する。(c2) の相手の issue 重要度表明から相手の価値構造推定は、(c1) に従ってエージェント i が相手エージェント j の issue s に対する価値の推定値 $\hat{v}_j(s)$ を求めることに相当する。(c3) の相手の issue 重要度表明に基づく自分の issue 重要度評

定変更は、(c2) に従ってエージェント i が issue s に対する自分の価値の推定を修正する重み $w_i(s)$ を設定することに相当する。

5.2 提案交換

関心擦り合わせの中で提案交換部分は以下のように特徴付けられる。

- (p1) 自分の価値構造 (issue 重要度) と相手の価値構造推定値に基づいて共同行為提案を行う。
- (p2) 相手の共同行為提案に基づいて提案構造の変更を行う。
- (p3) 相手の共同行為提案に基づいて受諾/拒絶/代替提案を行う。

ユーティリティ最大化による行為選択の原理に従えば、エージェント i の単独での行為選択 act は

$$act = \arg \max_{a \in \mathcal{A}} u(v_i, a)$$

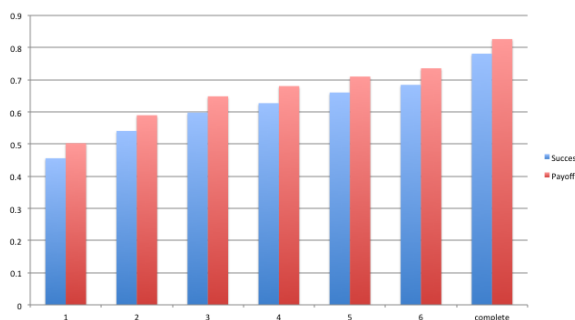


図 5 関心擦り合わせ回数の効果

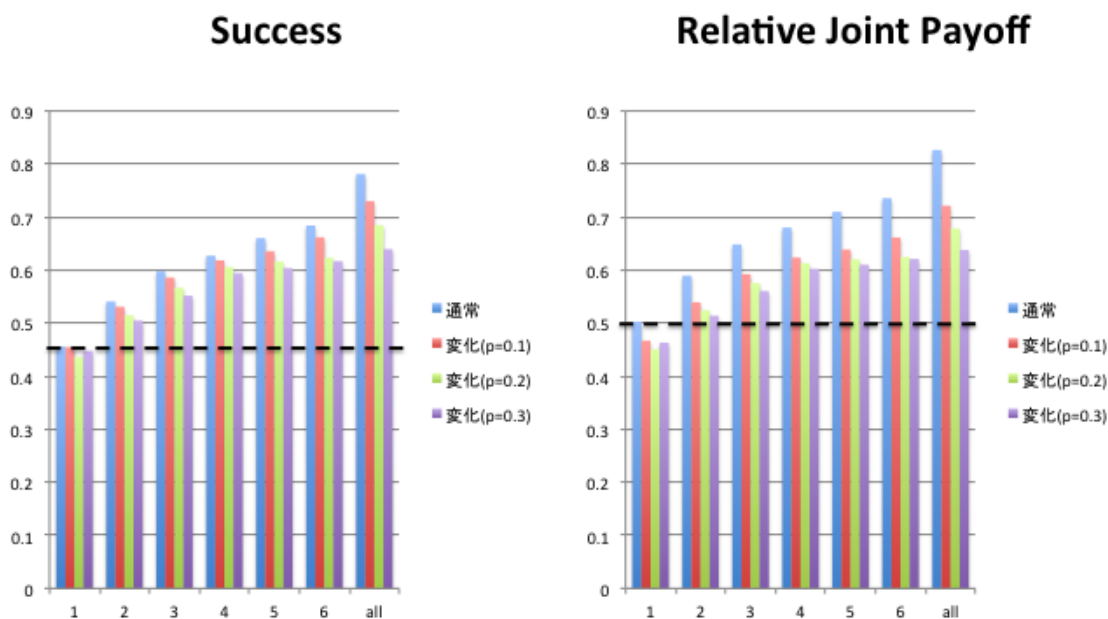


図 6 関心擦り合わせの履歴としての信頼感構築

となる。共同行為選択では相手の価値構造も考慮するため、関心擦り合わせを経た後の行為選択提案(p1)は以下のように捉えることができるだろう。

$$act = \arg \max_{a \in A} u((1 - \alpha)w_i v_i + \alpha \hat{v}_j, a)$$

エージェント*i*は自分の価値構造の中で相手の価値構造に合わせるとした部分は修正重み w_i を適用し、また全体として相手の価値構造推定を一定比率 α だけ取り入れて行為*a*のユーティリティを評定する。提案レベルでの交渉(p2)(p3)は相手の提案に基づく α, w_i の修正と見なすことができる。

5.3 関心擦り合わせと信頼

レストランの例を用いて一対の agent 間での Multi-issue negotiation のシミュレーションを行った。

- Coordination game 設定

- Concern table を random 生成
- 関心擦り合わせが1回だけ (One-shot alignment) と可能なすべての関心を擦り合わせ (Complete alignment) とを比較
- 繰り返し数 $n = 10,000$
- Parameter として、行為選択の際の他者の concern 考慮の一般的重みの範囲を $\alpha = \{0, 0.2, 0.5, 0.8\}$ とし、個別的重み w_i はすべて 1 に固定。

図4には、合計相対利得として行動選択の結果得られる利得の合計を各自が個別に最適行為選択をした場合との比を示す。関心擦り合わせが1回よりは完全関心擦り合わせの方が得られる合計相対利得が高いこと、擦り合わせ1回の場合には α が高いほど合計相対利得が高くなるが、完全関心擦り合わせでは $\alpha = 0.5$ が最適となることが分かる。

図5には、関心擦り合わせ回数を変化させたときの交渉成功率と合計相対利得とを示す。いずれも

関心擦り合わせ回数が多い, すなわち丁寧に関心擦り合わせを行うほど高い値が得られている。

最初に述べたように信頼を他者の行動選択に対する期待ととらえるるとすると, 上記のMulti-issue negotiationモデルと信頼との対応は以下のように考えることができる。

- 他者の価値構造推定 $\hat{v}_j(s)$ が行動選択の予測可能性を高める要素(s-1)(o-1)に対応している。
- 提案を選択する際の相手配慮 $\alpha, w_i(s)$ が他者の価値構造に合わせた行動選択要因(s-2)(o-2)に対応している。

関心擦り合わせの結果, 行動選択の際の一般の相手配慮 α が履歴として残り, そのagentのその後の行動選択の予測可能性を高めることが信頼形成に結びつくと考えられる。図6には, 関心擦り合わせの後にagentの関心構造(concern table)に確率的な微小変化を加えた後に行動選択を行った場合の合計相対利得を示す。丁寧に関心擦り合わせを行えば, 相手の関心に一定の変化が生じたとしても, 関心擦り合わせを行わない場合よりは高い合計相対利得が得られることが分かる。

6. おわりに

信頼は人間同士の協調行動を支える基本的要因である。情報システムの知能化が進展につれて知的エージェントと人間との信頼構築が重要な課題となってくる。本稿では, まず信頼概念の分析・分類を行い, 対話インタラクションを通じた相互信頼感構築に焦点を絞り, 関心擦り合わせ(concern alignment)概念の導入と, それに基づく現実の医療コミュニケーション対話の構造分析を示し, さらに関心擦り合わせ過程をmulti-issue negotiationととらえる計算モデルについて述べた。

謝辞

本研究の一部は, 日本学術振興会科学研究費補助金(基盤研究(B)「会話を通じた相互信頼感形成のマルチモーダル分析と共関心モデルの研究」(平成24年度~平成26年度, 研究代表者: 片桐 恭弘, 課題番号 24300061) によって実施したものである。

参考文献

- [1] Clark, H. H., (1996) *Using Language*, Cambridge University Press.
- [2] Katagiri, Y., Takanashi, K., Ishizaki, M., Enomoto, M., Den, Y. and Matsusaka, Y., (2011) “Concern alignment in consensus building conversations” *the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial2011)*, pp. 208–209.
- [3] Katagiri, Y., Takanashi, K., Ishizaki, M., Enomoto, M., Den, Y. and Matsusaka, Y. (2012) “Negotiation for Concern Alignment in Health Counseling Dialogues” *the 16th Workshop on the Semantics and Pragmatics of Dialogue (SemDial2012)*, pp. 173–174.
- [4] 山岸俊男 (1998) 信頼の構造—こころと社会の進化ゲーム, 東京大学出版会.
- [5] Allen, J. F. and Core, M. G., (1997) “DAMSL: Dialog act markup in several layers (Draft 2.1)” Technical report, Multiparty Discourse Group, Discourse Resource Initiative.
- [6] Bunt, H., (2006) “Dimensions in dialogue act annotation” *the 5th International Conference on Language Resources and Evaluation (LREC 2006)*.
- [7] 片桐恭弘, 高梨克也, 石崎雅人, 榎本美香, 伝康晴, 松坂要佐, (2012) 保健指導対話を対象とした相互信頼感形成過程の分析, SIG-SLUD-B103: 89-94, 人工知能学会資料.
- [8] Katagiri, Y., Takanashi, K., Ishizaki, M., Enomoto, M., Den, Y. and Matsusaka, Y., (2012) “Negotiation for Concern Alignment in Health Counseling Dialogues” *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue (SemDial2012)*, pp. 173–174.
- [9] Fatima S., Wooldridge M., Jennings N.R., (2004) “Optimal negotiation of multiple issues in incomplete information settings” AAMAS04, pp. 1080-1087.