社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

相手の知識モデルを使用した議論システムに基づく戦略的説得対話

横浜　静夏†　　高橋　和子†

† 関西学院大学理工学部　〒 669-1337 三田市学園 2 丁目 1 番地
E-mail: †{yokohama-shizuka,ktaka}@kwansei.ac.jp

あらまし　本発表では，説得対話において，対話相手の知識予測を使うことで，説得に失敗しないような戦略を与え，さらに，嘘と疑われる発言に対してそれを指摘する方法について述べる．まず，議論システムに基づいて対話進行に伴う各エージェントの知識や信念の変化を表現した対話モデルを作成する．このモデル上で対話相手の知識予測を使った説得のための戦略を提案する．次に，この予測を用いて嘘の指摘ができるようにプロトコルを拡張し，説得者が嘘をつくことで説得に成功する場合と嘘があばかれてしまう場合の対話例を表現する．また，この対話モデルの性質や問題点についても議論する．

キーワード　議論システム，対話，説得，相手のモデル，戦略，嘘

Strategic Persuasion Dialogue Based on an Argumentation System Using an Opponent Model

Shizuka YOKOHAMA† and Kazuko TAKAHASHI†

† School of Science &Technology, Kwansei Gakuin University　2-1, Gakuen, Sanda, 669-1337, JAPAN
E-mail: †{yokohama-shizuka,ktaka}@kwansei.ac.jp

Abstract　This presentation proposes a strategy for non-failing persuasion, and provides a protocol for pointing out a dishonest argument, using a predicted knowledge of an opponent. First, a dialogue model is constructed in which changes of an agent's knowledge and belief are represented based on an argumentation system. A strategy for non-failing persuasion using a predicted opponent's knowledge is proposed on this model. Next, the protocol is extended to point out a dishonest argument. A dialogue in which persuasion succeeds owing to a dishonest argument as well as the one in which dishonesty is exposed is represented using this protocol. Moreover, properties and problems of this model are discussed.

Key words　argumentation system, dialogue, persuasion, opponent model, strategy, lie

1. はじめに

エージェント間の対話において互いの合意を得るにあたり，意見の衝突や矛盾を解決することは重要である．説得は対話の種類の 1 つであり，そのような衝突解消・合意形成のための対話である．

Amgoud らは議論システムを使用した対話モデルを提案した [1]．彼らは，対話進行に伴い変わりゆく知識を扱う上で，各エージェントが現在の知識から構成された自身の議論システムを持ち，その受理可能な論証を現在の自身の信念とすることで，各エージェントの知識状態を表現した．しかし，このモデルには相手の現在の知識状態を予測するという観点が欠落している．

例えば以下のような，学生が研究室を選ぶ状況を考える．アリスはボブと同じ研究室に入りたいと考えている．アリスは意欲の高い学生で，厳しい研究室に配属されたいと考えており，チャーリー研に入りたい．彼女はチャーリーが，厳しくも面倒見がよい教授であることを知っている．一方ボブは，面倒見のよい研究室には入りたいが，厳しい研究室には絶対に入りたくない．またボブはチャーリーについて何も知らない．

(例 1)　もしアリスがボブの持つ知識について何も知らない場合，アリスは「チャーリー教授は厳しい先生だから，一緒にチャーリー研に入ろう」と言うかもしれない．その結果，(ボブは厳しい研究室を嫌うので) ボブの説得に失敗する．

(例 2)　もしアリスが，「ボブが厳しい教授を嫌う」ことを知っていた場合，アリスは「チャーリー教授は面倒見のよい先生だから，一緒にチャーリー研に入ろう」とだけ言うことで，ボブの説得に成功できるかもしれない．

(例 3)　もしボブが，『アリスは「チャーリーが厳しい事，及

びボブが厳しい先生を嫌うこと」を知っている』ということを知っている場合，(例 2) の「面倒見のよい先生だから，一緒にチャーリー研に入るべき」というアリスの発言が，「チャーリー教授は厳しいからボブは入るべきでない」という情報を隠した上での嘘発言であるということがボブにあばかれてしまう．

このように相手の知識の予測は，説得に有効な発言の選択や，相手の嘘を見抜く為にも利用される．相手の知識モデルを考慮に入れることにより，このような対話表現が可能になることを期待できる．

そこで本研究では，まず Amgoud らの対話モデルに，相手の予測知識を導入して改良した [2]．我々の対話モデルでは，エージェントは自身の議論システムに加え，相手の予測議論システムを構築することで，現在の相手の知識状態を予測する．加えて各エージェントが嘘指摘を行わない対話モデルにおける，相手の知識状態を加味した戦略を提案した．本論文ではこの対話モデルを，嘘指摘を含めた対話モデルに拡張する．相手への嘘の指摘は，相手の予測知識を用いることで実現している．そして説得者が嘘をついて説得に成功する場合と，嘘をついたことがあばかれてしまう場合の対話例を表現し，嘘指摘のプロトコルに関する性質や問題点を議論する．

第 2 章では，本研究の対話モデルが基礎におく，議論システムの形式的定義を与える．第 3 章では，本研究における対話モデルを形式化し，説得戦略を提案する．さらに，説得対話の例を紹介する．第 4 章では，嘘指摘可能な対話へ表現を拡張し，その性質を探る．第 5 章では，嘘指摘のプロトコルに関する問題点について議論する．第 6 章では，関連研究について述べる．第 7 章では，本研究の成果と今後の課題について述べる．

2. 議論システム

Dung は，論証と呼ばれる要素の集合とその上での二項関係の組で定義される，抽象議論システム（抽象議論フレームワーク）を提唱した [3]．抽象議論システムでは論証の内容を考慮せず，論証をノード，攻撃関係をエッジとしたグラフ表現上での意味論を与える．本研究で用いる議論システムでは，各論証をはじめに与える知識ベースに含まれる論理式から構成し，そこに，優先度の概念を導入する．

Σ を有限な命題論理式集合とし，その中の各論理式には関数 str により自然数が割り振られているとする．

［定義 1］ (論証)．Σ を命題論理式集合とし，これを知識ベースと呼ぶ．Σ は無矛盾であることや論理帰結に関して閉じていることを仮定しない．Σ 上の論証とは，根拠 H と主張 h の組 (H, h) として定義され，以下のどちらかの条件を満たす．

(i) $H = \emptyset$ かつ $h \in \Sigma$

(ii) H は 無矛盾であり，かつ，$H \vdash h, \forall h' \in H; h' \neq h$ を満たす Σ の（集合の包含関係における）極小部分集合である．ここで $\equiv$ は論理的同値を表す．

また，論証には優先度があり，論証の根拠を構成する論理式のうち，最も弱い論理式の強さ (根拠が無いときは主張の強さ) で決まる．

［定義 2］ (攻撃)．論証 $A_1 = (H_1, h_1)$ と $A_2 = (H_2, h_2)$ において，もし，$h_2 \equiv \neg h$ ただし $h \in H_1 \cup \{h_1\}$ かつ，A_2 が A_1 よりも優先度が大きいとき，A_2 は A_1 を攻撃するという．

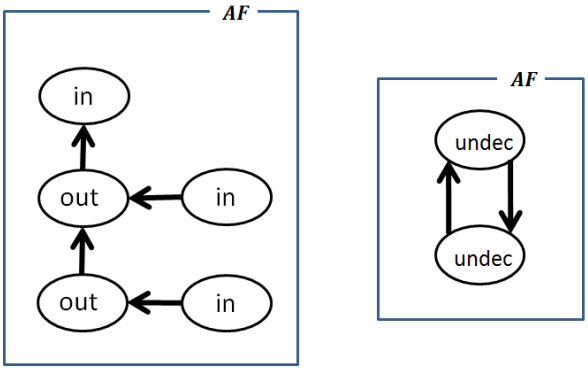

図 1　ラベリング
Fig. 1　labeling

［定義 3］ (議論システム)．議論システム $AF(\Sigma, str)$ とは，$\langle AR, AT \rangle$ である．ここで AR は Σ 上のすべての論証集合であり，AT は str に基づいた AR 上の攻撃関係である．なお，str の相違を考慮する必要がないとき，特に問題がなければ str を省略し $AF(\Sigma)$ と表記する．

議論システム $\mathcal{AF} = \langle AR, AT \rangle$ が与えられたとき，各論証に in，out，または undec のラベル付けを行うラベリング関数を考える [4]．

［定義 4］ (基礎ラベリング)．議論システム $\mathcal{AF} = \langle AR, AT \rangle$ を考える．以下を満たすラベリングのうち，ラベルが in である論証の数が (集合の包含関係において) 極小なものを基礎ラベリングといい，$Lab^{\mathcal{AF}}$ と表記する．

（1） A のラベルが in$\Leftrightarrow$ A を攻撃するすべての論証のラベルが out である．

（2） A のラベルが out$\Leftrightarrow$ A を攻撃する論証のうち，ラベルが in のものが存在する．

なおこの定義より，論証 A が何からも攻撃を受けていないときは A のラベルは in となる．また，任意の議論システムに対し，必ず基礎ラベリングが 1 つだけ存在する [4]．

一般にラベリングは様々な種類が定義されているが，本論文では基礎ラベリングの意味として用いる．

ある議論システム $\mathcal{AF}$ 及びその基礎ラベリング $Lab^{\mathcal{AF}}$ において，in にラベル付けされている論証を集めた集合を $in(Lab^{\mathcal{AF}})$ と表記する．すなわち $in(Lab^{\mathcal{AF}}) = \{A | Lab^{\mathcal{AF}}(A) = in\}$ である．なおこれは，Dung の抽象議論システムの意味論における **(基礎) 外延**と一致する [4]．

［定義 5］ (信念)．議論システム $\mathcal{AF}$ を考える．このときラベルが in であるすべての論証を構成する論理式集合を信念と呼ぶ．すなわち，$Bel(\mathcal{AF}) = \bigcup_{(H,h) \in in(Lab^{\mathcal{AF}})} H \cup \{h\}$

図 1 のグラフにおいて，ノードは論証，エッジは攻撃関係を表す $\mathcal{AF}$ 表現であるとする．ノード内の in，out，undec はそのノードのラベルを表す．

3. 対話モデル

3.1 議論システムに基づく対話モデル

説得者 P がその相手 C に対し，議題 ρ を説得するための対話を考える．説得者 P は自身の有限な知識ベース Σ_P を持ち，対話中，自身の知識ベースとそれまでに相手から得た情報で，自身の議論システム $\mathcal{AF}_P^{d_k}$ を作成する．加えて，P は自分が考える C の予測知識ベース Π_C を持ち，これを用いて相手の予測議論システム $\mathcal{PAF}_C^{d_k}$ も作成する．また C も同様である．

相手から発言をうけたとき，これらの議論システムは更新される．エージェントは自身の議論システムを再構築した際，そこに存在する論証を基本的に次の手として発言できる．次に，エージェントは，次の発言として許された手の中から，相手の予測知識ベースを用いて最善の手を選ぶ．

また本研究では，$\Pi_C \subseteqq \Sigma_C$，$\Pi_C \subseteqq \Sigma_P$ を仮定する．なお議題 ρ は Σ_P の要素である．

［定義 6］ (対話環境)．L を命題論理式集合とする．対話環境とは以下を満たす 4 つ組 $\langle Participant, KG, PKG, \rho\rangle$ である．

- $Participant = \{P, C\}$
- $KG : Participant \to 2^L$
- $PKG : Participant \to 2^L$

かつ，$\forall X, Y \in Participant (PKG(Y) \subseteqq KG(X))$

- $\rho \in KG(P)$

以後，X，Y を，「任意のエージェントとその相手において」の意として用いる．また，$KG(X)$ は X の知識ベース (Σ_X)，$PKG(X)$ は Y の持つ X の予測知識ベース (Π_X) と表す．

p を論理式，S を論理式集合とするとき，エージェントは以下の 4 つの act を使い対話を行う．言明 $assert(S, p)$, は根拠つきで主張を発言し，補足 $assertS(S, p)$ は根拠のみを発言する．質問 $challenge(p)$ は相手に理由を問い，パス $pass$ は，何の情報も発言しないままにその番をパスをする．

なお，以後，$assert(S, p)$ または $assertS(S, p)$ であることを $assert/S(S, p)$ と略記する．

また，1 つの act から論理式集合を返す関数 $formula$ は，$assert(S, p)$ なら $S \cup \{p\}$，$assertS(S, p)$ なら S，その他の act なら $\emptyset$ を返すとする．

［定義 7］ (手 (move))．手とは，エージェント X と act T の組 (X, T) である．

［定義 8］ (対話)．議題 $\rho \in \Sigma_P$ に関する，説得者 P と相手 C 間の対話 d_k とは，手の有限列 $[m_0, \ldots, m_{k-1}]$ である．ここで各 $m_i = (X_i, T_i)$ $(0 \leqq i \leqq k-1)$ は以下の条件を満たす．

(i) $X_0 = P$ かつ T_0 かつ $assert(S, \rho)$.

(ii) i が偶数なら $X_i = P$，i が奇数なら $X_i = C$

(iii) m_i は可能手である．

可能手とは後に定義する，対話プロトコルに従った手のことである．

［定義 9］ (終了対話)．議題 $\rho \in \Sigma_P$ における，説得者 P と相手 C 間の対話 $[m_0, \ldots, m_{k-1}]$ において，$m_{k-2} = (X, pass)$ かつ $m_{k-1} = (Y, pass)$ であるとき，この対話を終了対話と呼ぶ．

［例 1］ $d_5 = [(P, assert(\emptyset, \rho), (C, assert(\{a, a \to \neg\rho\}, \neg\rho), (P, assertS(\{b, b \to \rho\}, \rho), (C, pass), (P, pass),]$ は終了対話である．

エージェントのコミットメントストアとは，対話進行に伴いエージェントがそれまでに開示した論理式の集合である．

［定義 10］ (コミットメントストア)．対話 $d_k = [m_0, \ldots, m_{k-1}]$，$m_i = (X_i, T_i)$ $(i = 0, \ldots, k-1)$ において，d_k における X のコミットメントストア $CS_X^{d_k}$ とは，もし $k = 0$ ならば $\emptyset$，もし $k \neq 0$ ならば $\bigcup_{i=0,\ldots,k-1, X_i = X} formula(T_i)$ である．

［定義 11］ (対話における議論システム)．対話 $d_k = [m_0, \ldots, m_{k-1}]$ において，対話 d_k におけるエージェント X の議論システムを $AF(\Sigma_X \cup CS_Y^{d_k})$ と定義し，$\mathcal{AF}_X^{d_k}$ と表記する．また，対話 d_k における，X が持つ Y の予測議論システムを $AF(\Pi_Y \cup CS_X^{d_k} \cup CS_Y^{d_k})$ と定義し，$\mathcal{PAF}_Y^{d_k}$ と表記する．

各 act には，対話 d_k の時点でそれが選択可能かどうかの前提条件が定められており，それを満たす手を可能手という．

［定義 12］ (可能手)．$d_k = [m_0, \ldots, m_{k-1}]$ において，この時点の X の議論システムを $\mathcal{AF}_X^{d_k} = \langle AR_X^{d_k}, AT_X^{d_k}\rangle$ とする．対話 d_k における，エージェント X の各 act の前提条件を以下に定義する．もし m_k 及び d_k が以下の前提条件を満たすとき，m_k を d_k における X の可能手という．なお，$0 \leqq i \leqq k-1$ とする．

m_k，d_k に対し，まず $pass$ 以外の act に共通する前提条件を以下に示す:

- $pass$ 以外の act に対し: $m_k \neq m_i$.

次に，すべての act に共通する 2 つの前提条件を以下に示す:

- すべての act に対し: m_{k-2} も m_{k-1} も $pass$ でない．
- $assert/S(S, p)$ の act に対し: $(S, p) \in AR_X^{d_k}$

これらに加えて，act ごとには以下の前提条件が必要になる：

- 言明 $assert(S, p)$:
  - $k \neq 0$ ならば， $p = \rho$
  - $k \neq 0$ ならば，$\neg p \in CS_Y^{d_k}$ かつ $(X, assertS(S, p)) \neq m_i$
- 補足 $assertS(S, p)$:
  - $p \in CS_X^{d_k}$ かつ $(X, assert(S, p)) \neq m_i$
- 質問 $challenge(p)$:
  - $p \in CS_Y^{d_k}$ かつ ，$(X, assert/S(S, p)) \neq m_i$; $S \neq \emptyset$.
- パス $pass$:
  - $k \neq 0$.

手 $m_k = (X, T)$ を出した後，以下の更新が行われる: d_{k+1} は d_k の末尾に (X, T) を追加して得られる．さらにコミットメントストアについて，$CS_X^{d_{k+1}} = CS_X^{d_k} \cup formula(T)$ かつ $CS_Y^{d_{k+1}} = CS_Y^{d_k}$ となる．

各々の持つ知識ベースは有限であり，また $pass$ 以外の同じ手は何度も出せず，さらに $pass$ が 2 回続くと $pass$ を含めたすべての手が出せないことから，対話は必ず終了する．

［定義 13］ (正直/嘘)．対話 $d_{k+1} = [m_0, \ldots, m_k]$ において，$m_k = (X, assert/S(S, p))$ ならば $Lab^{\mathcal{AF}_X^{d_k}}((S, p)) = in$ であるとき，X の手 m_k は正直な発言である，といい，そうでないときは嘘の発言である，という．

［定義 14］ (成功/失敗)．議題 ρ に対する説得者 P と相手 C による終了対話 d_k において，$\rho \in Bel(\mathcal{AF}_C^{d_k})$ のとき P が説得成功するといい，$\neg\rho \in Bel(\mathcal{AF}_C^{d_k})$ のとき，P が説得失敗するという．

3.2 戦　　略

戦略とは，対話環境 $\langle Participant, KG, PKG, \rho\rangle$ とそれにおけるある対話 d_k を引数とし，可能手の集合から 1 つの手 $m_k = (X, T)$ を返す関数である．

本研究では P のための戦略 $\mathcal{S}_{\mathcal{NF}}$ を提案する．P は，相手 C が議題 ρ に合意する可能性のある手を積極的に選択する．逆に，C が $\neg\rho$ を受け入れてしまう可能性のある危険な手を予測し，そのような論証の発言を避ける．そして，C が ρ を信じ目的が達成されたと予測するならば，P は $pass$ を出し，それ以上情報を開示しない．

戦略 $\mathcal{S}_{\mathcal{NF}}$: $\mathcal{AF}_P^{d_k}$ と $\mathcal{PAF}_C^{d_k}$ をそれぞれ，対話 d_k における P の議論システム，P が考える C の予測議論システムとする．このとき，手 $m_k = (P, T)$ は以下のルール (1)，(2)，(3) の優先順に選択される．

（1） $d_k \neq d_0$ のとき，もし $\rho \in Bel(\mathcal{PAF}_C^{d_k})$ ならば $(P, pass)$ を選択する．

（2） $d_k = d_0$ のとき，すべての可能手 m_0 に対し，$\neg\rho \in Bel(\mathcal{PAF}_C^{[m_0]})$ となるならば，$m_0 = (P, assert(\emptyset, \rho))$ を選択する．

（3） act は，$assert(\emptyset, p)$，$assert(S, p)$(ただし $S \neq \emptyset$)，$assertS(S, p)$，$challenge(p)$，$pass$ の順で優先される．すなわち，$assert(\emptyset, p)$ が最も優先される．さらにもし T が $assert(\emptyset, p)$，$assert/S(S, p)$ のいずれかのとき，以下のルールが適用される．

（a） (P, T) を選んだ場合に $\rho \in Bel(\mathcal{PAF}_C^{d_k+1})$ となるならば，(P, T) を選ぶ．

（b） (P, T) を選んだ場合に $\neg\rho \in Bel(\mathcal{PAF}_C^{d_k+1})$ となるならば，(P, T) は選ばない．

もし上記のルールを適用した結果複数の候補手が存在する場合は，その中からランダムに手を 1 つ選択する．

3.3 応　用　例

ここで，1 節で紹介した研究室選択の例を考える．a, g と s をそれぞれ，「チャーリー研を選択する」，「チャーリーは面倒見がよい」，「チャーリーが厳しい」を表す論理式だとする．この対話では，P をアリス，C をボブとし，P が C に a を信じさせることを目的として説得する．

各論理式の強さを以下のように仮定する．
$str(g) = str(s) = str(s \to \neg a) = 3$，$str(g \to a) = str(s \to a) = 2$，$str(a) = str(\neg a) = 1$.
各知識ベースは以下のように与えられるとする．

$$\Sigma_P = \{g, s, g \to a, s \to a, a, s \to \neg a, \neg a\}$$
$$\Pi_P = \{g \to a\}$$
$$\Sigma_C = \{g \to a, s \to \neg a, \neg a\}$$
$$\Pi_C = \{g \to a, s \to \neg a, \neg a\}$$

最初の時点で P には 3 つの可能手が存在しており，それぞれ $assert(\emptyset, a)$, $assert(\{g, g \to a\}, a)$, $assert(\{s, s \to a\}, a)$ である．なお，最初の時点での P 自身の議論システム内おいて，これらの論証はすべて out にラベリングされているため，これらの発言はすべて嘘の発言となる．

戦略 $\mathcal{S}_{\mathcal{NF}}$ に従うと，P は最初にルール 3(a) と 3(b) により $assert(\{g, g \to a\}, a)$ を行う．P の持つ C の予測議論システム内では，論証 $(\{g, g \to a\}, a)$ は in とラベリングされ，$a \in Bel(\mathcal{PAF}_C^{d_1})$ となる．なお，$\Pi_C = \Sigma_C$ であるので，$a \in Bel(\mathcal{AF}_C^{d_1})$ も成り立つ．P はその後，戦略 $\mathcal{S}_{\mathcal{NF}}$ に基づき，C のいかなる手に対しても $pass$ を出し続け，最後には対話が終了し説得に成功する．

P が戦略を用いない場合，最初の時点で 3 つの可能手のうちどの手も出しうる．もし P が $assertS(\{s, s \to a\}, a)$ を出した場合，C の議論システム内では，論証 $(\{s, s \to \neg a\}, \neg a)$ が in とラベリングされ，$\neg a \in Bel(\mathcal{AF}_C^{d_1})$ となる．もしその後 P が $assertS(\{g, g \to a\}, a)$ を出したとしても，いかなる対話進行でも，対話終了の時点で $\neg a$ は C の信念となり P は説得に失敗してしまう．

4. 嘘指摘を含む対話

この節では，1 節 (例 3) のように，相手のエージェントが嘘をついた際に，それを指摘するための表現拡張を考える．

4.1 嘘指摘のための act

まず嘘指摘の導入の前に，相手の発言を嘘ではないかと疑うことを定義する．

［定義 15］ (疑い)．$m_{k-1} = (Y, assert/S(S', p'))$ かつ $P\mathcal{AF}_Y^{d_k}$ 内で (S', p') のラベルが out または undec であるとき，「X は (S', p') が正直な発言か疑う」，「Y が正直か疑わしい論証を発言する」という．

嘘指摘ができるようにするため，前章で定義した対話モデルに，新たな act として「$lie?(S, p)$」を導入する．

X が Y の発言 (S', p') を正直かを疑うとき，この予測する議論システム内に，(S', p') を攻撃しかつラベルが in または undec であるような論証 (S, p) があるということである．そこで X は，「論証 (S, p) の存在を隠して (S', p') を発言したのではないか？」という嘘指摘 $lie?(S, p)$ を Y に対し行う．

「$lie?(S, p)$」の導入に伴い 3 節の可能手の定義において以下のように変更し，嘘指摘を含む可能手の定義をする．

- 嘘指摘 $lie?(S, p)$ の act を追加:
  - $m_{k-1} = (Y, assert/S(S', p'))$ かつ $P\mathcal{AF}_Y^{d_k}$ 内において，(S', p') を攻撃し，ラベルが in または undec であるような (S, p) が存在する．
- 言明 $assert(S, p)$ において，$k \neq 0$ のときの条件をさらに 2 つに分ける:
  - もし $k \neq 0$ かつ $m_{k-1} \neq (Y, lie?(S', p'))$ ならば，$\neg p \in CS_Y^{d_k}$ かつ $(X, assertS(S, p)) \neq m_i$
  - もし $k \neq 0$ かつ $m_{k-1} = (Y, lie?(S', p'))$ ならば，$p \equiv \neg h$; $h \in S' \cup \{p'\}$
- $assert(S, p)$ 以外の act に対し以下を追加:

− $m_{k-1} \nmid (Y, lie?(S', p'))$

このように嘘指摘の act の特徴は，これまで前提条件には自身の議論システムのみ用いていたのに対して，自身の議論システムではなく相手の予測議論システムを使うことである．さらにもう 1 つの特徴として，$lie?(S, p)$ をされた側が，その指摘に対し嘘ではないと反論できない場合，対話自体がそこで終了することがあげられる．

なお，Y の嘘指摘 $m_{k-1} = lie?(S', p')$ に対して，X はこの直後に m_{k-1} に対しての言明 $m_k = (X, assert(S, p))$; $p \equiv \neg h$ かつ $h \in S' \cup \{p'\}$ をすることによって嘘ではないと反論する必要がある．これを**弁解する**という．嘘指摘の直後に弁解が返せない場合は，対話は強制終了する．このとき，弁解できなかった側が嘘つきであるとだけ判定され，説得における成功・失敗を考えることに意味をなさないとする．

なお，X が $lie?(S, p)$ を行った後，CS_X に $S \cup \{p\}$ が追加される．

4.2 対話の結果

エージェント P は対話をすることで相手 C に議題 ρ を説得することを目的とし，目的達成のために P は正直な発言も嘘の発言もすることができる．しかしもし C に嘘がばれてしまった場合は，P は嘘つき者と見なされ対話が続けられなくなる．戦略上発言した方が説得が優位になると思われる嘘において，それがばれないとわかっているならば，戦略に基づいて問題なくそれを発言できる．しかしそれがばれる可能性のある嘘ならば，発言すべきか否か，思案できるような枠組みや戦略に改訂する必要があるだろう．

次の節では，戦略 $\mathcal{S}_{\mathcal{NF}}$ に基づいた嘘の発言が説得に有効になる例と，嘘つき者となって説得できない例を示す．

4.3 嘘指摘可能な対話例

アリス (P) とボブ (C) の対話を考える．論理式やその強さ，及び Σ_P と Π_C はすべて 3.3 節の応用例と同じであるとする．

［例 2］ C の知識ベース Σ_C，及び C の持つ P の予測知識ベース Π_P が以下の通りとする．

$$\Sigma_C = \{g \to a, s \to \neg a, \neg a\}$$
$$\Pi_P = \{g \to a\}$$

アリスが初手において，戦略 $\mathcal{S}_{\mathcal{NF}}$ に基づき $assert(\{g, g \to a\}, a)$ を行うとする．この時，この発言に対しアリスはボブから嘘指摘を受けない．なぜならば，C の予測する P の議論システム $\mathcal{PAF}_P^{d_1}$ において，論証 $(\{g, g \to a\}, a)$ を攻撃する論証がないからである．その結果，応用例 3.3 の通りアリスは説得に成功する．

［例 3］ C の知識ベース Σ_C，及び C の持つ P の予測知識ベース Π_P が以下の場合を考える．

$$\Sigma_C = \{g \to a, s \to \neg a, \neg a, s\}$$
$$\Pi_P = \{s, s \to \neg a\}$$

この時，同じくアリスが初手において，戦略 $\mathcal{S}_{\mathcal{NF}}$ に基づき $assert(\{g, g \to a\}, a))$ を行うとする．すると対話が以下のように進行する可能性がある．

$$m_0 = (P, assert(\{g, g \to a\}, a))$$
$$m_1 = (C, lie?(\{s, s \to \neg a\}, \neg a))$$

m_1 において C が嘘指摘 $lie?(\{s, s \to \neg a\}, \neg a)$ をできる理由は，C の予測する P の議論システム $\mathcal{PAF}_P^{d_1}$ において，論証 $(\{s, s \to \neg a\}, \neg a)$ のラベルが in であり，かつ論証 $(\{g, g \to a\}, a)$ を攻撃しているからである．

これは，アリスの最初の発言「チャーリー先生は面倒見がいいよ．面倒見がいい研究室に入りたいなら一緒に入ろうよ」に対し，「君はチャーリー先生が厳しいことも，僕が厳しい研究室を嫌うことも知ってたはずなのに，それを隠して説得しにきてないかい？」とボブが指摘するという対話である．ここでアリスは，m_1 への弁解（チャーリーは厳しくない，または，それでも研究室には入るべきであるという強い根拠を言う等）ができず，対話が終了しアリスは嘘つきであると判定される．

このようにアリスの嘘があばかれないどうかは，ボブの持つアリスの予測知識ベースに左右される．

4.4 嘘指摘に関する性質

互いが正直な発言だけをしていれば，少なくとも嘘指摘をされ弁解できずに対話が終了することがない，ということが保証できるか考察する．以下では，正直な発言に対して嘘指摘をされた場合，それに弁解できるための条件を示す．

［補題 1］ 対話 $d_{k+1} = [m_0, ..., m_{k-1}, m_k]$ において，$m_{k-1} = (X, assert/S(S', p'))$，$m_k = (Y, lie?(S, p))$ であるとする．この時，もしも，$m_{k-1} = (X, assert/S(S', p'))$ が正直な発言であるならば，$\mathcal{AF}_X^{d_{k+1}}$ 内に論証 (S, p) を攻撃しラベルが in である論証 (S'', p'') が必ず存在する．

［証明 1］ 仮定より，$m_{k-1} = (X, assert/S(S', p'))$ は正直な発言であり，またこの手を出した後の X 自身の議論システムは変わらない $(\mathcal{AF}_X^{d_{k-1}} = \mathcal{AF}_X^{d_k})$ ため，$\mathcal{AF}_X^{d_k}$ においても，論証 (S', p') は in にラベリングされている...(a).

また，lie?のプロトコルの定義より，d_k の時点で，$P\mathcal{AF}_X^{d_k}$ 内に，ラベルが in または undec である論証 (S, p) が存在し，論証 (S', p') を攻撃している．

ここで，$PAR_X^{d_k} \subseteqq AR_X^{d_k}$ なので，$\mathcal{AF}_X^{d_k}$ 内には論証 (S, p) が存在し，論証 (S', p') を攻撃している．ラベリングの定義と (a) より，この論証 (S, p) は out にラベリングされている．$\mathcal{AF}_X^{d_k}$ 内で論証 (S, p) が out にラベリングされているということはすなわち，論証 (S, p) を攻撃し，かつラベルが in であるような論証 (S'', p'') が少なくとも 1 つ存在するということである...(b).

ここで，$AF(\Sigma_X \cup CS_Y^{d_k}) = AF(\Sigma_X \cup CS_Y^{d_k} \cup S \cup \{p\}) = AF(\Sigma_X \cup CS_Y^{d_{k+1}})$ ゆえに $\mathcal{AF}_X^{d_k} = \mathcal{AF}_X^{d_{k+1}}$．この事実と (b) より，$\mathcal{AF}_X^{d_{k+1}}$ 内にも，論証 (S, p) を攻撃し，ラベルが in である論証 (S'', p'') が存在する． □

この補題より，以下の定理が成り立つ．

［定理 1］ 補題 1 において，$(X, assert/S(S'', p'')) \nmid m_i (0 \leqq i \leqq k)$ の時，X は論証 (S'', p'') を嘘指摘 m_k に対しての弁解 $(m_{k+1} = (X, assert(S'', p'')))$ として正直に発言できる．

また，常に嘘指摘ができることを保証されていないならば，

相手の嘘を指摘できずに見逃してしまうかもしれない．以下では，Y が正直か疑わしい論証を発言したときに，X がそれを指摘できるための条件を示す．

［補題 2］ 対話 $d_k = [m_0, ..., m_{k-1}]$ において，$m_{k-1} = (Y, assert/S(S', p'))$ であるとする．このとき，X が論証 (S', p') を正直な発言か疑うならば，$PA\mathcal{F}_Y^{d_k}$ 内において，(S', p') を攻撃し，ラベルが in または undec であるような論証 (S, p) が必ず存在する．

証明はラベリングの定義から明らかである．またこの補題より，定理 1 と同様に以下の定理が成り立つ．

［定理 2］ 定理 2 において，$(X, lie(S, p)) \neq m_i (0 \leqq i \leqq k)$ の時，m_k において X は論証 (S, p) を嘘指摘 $(lie?(S, p))$ として発言できる．

5. ディスカッション

前節で導入した act である $lie?(S, p)$ や，$assert$ の弁解における前提条件には，以下のような問題点がある．

問題 1　正直か疑わしい論証を相手が発言した場合に，必ずそれを嘘指摘できるとは限らない．

問題 2　X が嘘指摘をした後に Y が弁解をしても，X が納得できないことがある．

問題 3　Y からの嘘指摘に対して反論できる論証を X が持っているにも関わらず，弁解できないことがある

以上のような問題が起こる原因は，同じ論証を用いた嘘指摘，及び弁解が複数回できないことであると考えられる．

そこで，これらの問題解決のために，嘘指摘を含む可能手を「嘘指摘，及び弁解に関しては，同じ手を何度も出せる」ように変更することを提案する．例えば，弁解のための act として $excuse(S, p)$ を別途もうけ，弁解の前提条件はそのままに，$assert$ と違い何度でも発言できるようにする．さらに $lie?(S, p)$ は，直前の言明だけでなく弁解に対してもできるようにし，かつ何度でも出せるようにする．この改変により，定理 1，2 から，相手の発言が正直か疑うときはいつでも嘘指摘をすることができ，また自分が正直な発言をしたならば，それに対して嘘指摘をされても必ず弁解できることが保証される．

ところがこのように単純に嘘指摘・弁解を何度でも出せるように改変すると，対話の停止性が問題となる．そこで，「弁解のために用いることのできる論証は，嘘指摘に用いられた論証から攻撃を受けていないものに限る」と制限を加える．そのようにした場合，「嘘指摘とその弁解の繰り返しは，相手が納得するか，自分が弁解できなくなるかのいずれかで必ず終了する」等，停止性に関する問題について今後検討する必要がある．

6. 関 連 研 究

本研究のモデルは Amgoud らの研究に基づいている [1]．彼らの主な貢献は，各々の議論システムを用いることで現在のエージェントの信念を表現し，その対話モデルがいくつかの対話タイプに適用できることを示した点である．また戦略も提案し，その戦略に沿った場合の対話例も示している [5]．

彼らの研究と本研究の最も大きな違いは，予測知識ベースの有無である．彼らは相手の内部状態を一切考慮せずに戦略を構築していたのに対し，本研究では相手の知識状態を加味した戦略を構築した．また，彼らの研究には，嘘も発言可能にした拡張対話モデルがあるが [6]，嘘の指摘に関する表現や方法については考えられていない．それに対し本研究では，相手の予測議論システムを用いることで，嘘指摘の表現を可能にした．

また，Sakama は不正直な論争ゲームを提案した [7]．彼のモデルでも，ラベルが in でない論証の発言を虚偽と定義しており，勝つ機会を得るための条件等を調べている．しかし彼の論争ゲームの目的は相手を負かすことであり，説得を目的としていない．また相手の予測表現等は取り入れておらず，正直な発言を心がける等の戦略は提案しているが，具体的な戦略までは提案していない．また嘘の指摘に関する表現は行っていない．

7. 結　　論

本研究では，相手の予測知識ベースを使用した対話モデルを提案した．その上で，説得のための戦略の提案や嘘をあばくためのプロトコルを導入した．

本研究の主な成果は，相手の予測知識を使用した対話の形式化，その対話における戦略の提案，及び嘘指摘を含む対話への拡張表現の提案である．相手の持つ知識の予測を導入することで，戦略提案や嘘指摘等，説得対話のモデル化において表現をより豊かにすることができた．

今後の課題として，まず 5 節で議論したような問題点を解決できるように，提案した嘘指摘のプロトコルの性質を調べることがあげられる．また，どのような場合に嘘があばかれる・あばかれないのかの条件を見出し，最終的には嘘指摘を考慮した新たな戦略提案等も行いたい．

文　　献

[1] L. Amgoud, N. Maudet, and S. Parsons, “Modelling dialogues using argumentation,” AAMAS, pp.31–38, 2000.

[2] S. Yokohama and T. Kazuko, “What should an agent know not to fail in persuasion?,” EUMAS-AT2015, 2015.

[3] P.M. Dung, “On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games,” Artificial Intelligence, vol.77, no.2, pp.321–357, 1995.

[4] P. Baroni, M. Caminada, and M. Giacomin, “An introduction to argumentation semantics,” The Knowledge Engineering Review, vol.26, pp.365–410, Dec. 2011.

[5] L. Amgoud and N. Maudet, “Strategical considerations for argumentative agents (preliminary report).,” NMR, pp.399–407, 2002.

[6] E. Sklar, S. Parsons, and M. Davies, “When is it okay to lie? a simple model of contradiction in agent-based dialogues,” ArgMAS, pp.251–261, 2005.

[7] C. Sakama, “Dishonest arguments in debate games.,” COMMA, pp.177–184, 2012.