1E3-4

議論構造の変化の形式的表現
– 秘密を持つエージェントとの対話 –
A Formal Representation of the Change of an Argumentation Structure
– A dialogue with an agent with a secret –

横浜静夏 *1　　高橋和子 *1
Shizuka Yokohama　　Kazuko Takahashi

*1関西学院大学大学院理工学研究科
Graduate School of Science and Technology, Kwansei Gakuin University

We formalize a dialogue aiming at giving a model of a dynamic argumentation. Most works on argumentation in computer science mainly study on static aspects of its structure, whereas few of them handle dynamic aspects of an argumentation procedure. However, an argumentation structure actually changes each time of agent's utterance. We consider a dynamic model of an argumentation which can handle the change of argumentation structure as well as agents' knowledge bases. In this paper, we focus on a dialogue between an agent that wants to know a truth value of a specific proposition and an agent that wants to keep that of another specific proposition secret. The latter sometimes tells a lie and the former points it out, which follows its correction. We define a lie, formalize protocols that consist of such a dialogue and invesitgate how such dialogue proceeds.

1. はじめに

対話は何らかの目的を持って互いの意見を発言していく行為である．対話には，互いが協力するものや説得や交渉目的など様々なタイプがあり, それぞれでエージェントの目的や対話の方法は異なる．一般に，対話を形式化すると，何らかの知識ベースをもつエージェントがある条件を満たす発言を行い，その結果新しい情報が相手にもたらされる．それを受けて改変された状況をもとに対話が進行していくことになる．ここで，エージェントが発言する際，議論 (argumentation) の枠組みを利用して妥当な発言を決める方法が提案されている．

議論の枠組みは Dung によって提唱され [Dung 95]，その後多くの研究が行われている [Rahwan 09]．Dung は，論証の集合と論証間の攻撃関係の二項組で議論の枠組みを記述し，互いに攻撃されず，また自分が攻撃されたときは別の論証が守ってくれるような集合を外延 (extension) として定めた．Amgoud らはエージェントの知識ベースから論証として発言を作成し，外延に属する論証を妥当な発言として認めることにした [Amgoud 00]．さらに，彼らは相手の発言によって新しい情報を受け入れる場合にも，妥当な場合のみ受け入れて自分の知識を改変するような仕組みを提案することで合理的エージェント間の対話を形式化した．Parsons らは，この仕組みが交渉や説得を含む複数の種類の対話に適用できることを示し，各々の対話種類での終了時の状態や，最初の知識ベースの状態や対話方法の戦略と最終結果との関係等を示した [Parsons 03]．この仕組みには，情報収集を目的としたエージェントが，目的とする情報を得る場合についても応用できると思われるが，このとき情報を与える側のエージェントが何かの理由によって真実を言わない場合についてはそのままでは応用できない．

本研究は，ウソを含む対話を扱えるように Amgoud らの仕組みを拡張し，議論の枠組みを利用した対話プロトコルの設定とそれに伴う各々の知識ベースの変化を形式的に扱える方法を開発することを目標とする．そのために，まずウソをつく理由とその内容を形式化する．

連絡先: 横浜静夏，関西学院大学大学院理工学研究科，兵庫県三田市学園 2-1，079-565-8391，yokohama-shizuka@kwansei.ac.jp

本論文では，一方のエージェントが情報収集を目的として質問し，もう一方のエージェントが隠したい秘密をもち，それを知られないためにウソをつくという状況下で，エージェントがウソをつく条件や対話プロトコルを定式化し，この定義に基づいた対話の流れを例示する．ウソをつく条件としては，ある内容を正直に発言するとまずいかどうかの判断基準を 2 つ設けた．さらにエージェントの性格づけを行い，性格によってウソの内容を決める．また，質問する側のエージェントには，ウソが分かったときに指摘する手段を設ける．質問側は，質問と指摘を繰り返し，徐々に真実の知識を獲得していく．

本論文は以下の構成となっている．2 節では対話プロトコルための用語定義と，対話プロトコル，対話結果の評価に関して説明する．3 節ではウソをつくことの定義を行っている．4 節ではエージェントに性格を設け，発言の内容に制限をかける．5 節では対話の例を紹介する．6 節ではまとめと今後の課題について述べる．

2. 対話プロトコル

本論文では，ある命題の真偽を知りたいエージェントと，ある命題の真偽を隠したいエージェントとの間の対話を考える．

2.1 用語定義

質問エージェントを Q，回答エージェントを A とし，Q の知りたい命題 (主題) を q と表し，A の知られたくない命題 (秘密) を s と表す．ただし，$K_Q \nvdash q$ かつ $K_Q \nvdash \neg q$ を満たしていること (Q 自身は q かどうかを知らない)，$K_A \vdash s$ を満たしていること (A は秘密 s を持っている) が前提である．また，各知識ベースに出現する原子論理式を発言記号と呼ぶ．

2.2 対話

ここでは Amgoud らの提案したプロトコルをもとに，ウソを含む対話を表現するプロトコルとして *question*, *answer*, *contradict*, *replace* の 4 種類を考える．

question はエージェント Q がエージェント A に対して命題の真偽を問うものである．この回答に対するものが *answer* であり，A は問われた命題の真偽かまたは *unknown*(知らない) と回答する．ただし A が真実を答えるとは限らない．

contradict はエージェント Q がエージェント A に対し，今までの A の相手の発言と自分の知識から矛盾を発見して，相手の発言に訂正を求めるものである．この回答に対するものが *replace* であり，A は過去の発言の 1 つを訂正する．ただしこのとき過去に訂正したものを再度訂正することは許されない．また本当の訂正をするとは限らない．

各エージェントのある時点までの発言による知識をそれぞれ Δ_Q, Δ_A とする．これらは対話の中で共有の知識として参照される．

Δ_Q は論理式の集合，Δ_A はある発言記号 X に関してとその真偽との対 (X,Y) の集合である．エージェント A に関して，Δ_A は，*answer* の場合に対 (X,Y) が付加され，*replace* の場合はある対 (X,Y) が消去され，その訂正が加わる．Δ_Q は，*contradict* の場合のみ論理式が付加される．

また，*Log* をそれまでにエージェント間でやりとりされた発言の集合とし，その中の最新の発言を *last* とする．

以下に各プロトコルの定義を述べる．

ただし，${\Delta_A}_{snd}$ は，Δ_A 内の (X,Y) から，*unknown* 以外の Y のみを集めた集合とする．

question(X) ただし X は発言記号．過去に聞いたことは 2 度質問できず，また初めに与えられた二人の知識ベースとは関係のない発言記号について聞くことは許されない．

発言条件 以下の 2 つをともに満たすことである．

- X は $K_Q \cup K_A$ 内に現れる発言記号
- $question(X) \notin Log$

answer(X,Y) 直前の質問内容 X に対し，Y を解答する．

発言条件 以下の 2 つをともに満たすことである．

- *last* が $question(X)$ である．
- $Y = X$ または $Y = \neg X$ または $Y = unknown$

Δ_A **変化** $\Delta_A \leftarrow \Delta_A \cup \{(X,Y)\}$

contradict(Γ) ただし Γ は命題論理式集合．Q 自身の知識の一部と，A の過去の発言の一部を取ってくると矛盾するような Γ を発言する．Δ_Q には，自分の知識として発言した部分のみ追加する．

発言条件 以下の 3 つをともに満たすことである．

- $\Gamma \vdash \bot$
- $\Gamma \subseteq K_Q \cup {\Delta_A}_{snd}$
- Γ が極小集合

Δ_Q **変化** $\Delta_Q \leftarrow \Delta_Q \cup (\Gamma \cap K_Q)$

replace(X,Y) 直前に受けた矛盾の指摘の中から，自分の発言のどれか 1 つを訂正する．ただし過去に一度訂正したものを再度訂正することを許さない．また訂正前の発言を Δ_A から消去して，訂正後のものを加える．

発言条件 以下の 3 つをともに満たすことである．

- *last* が $contradict(\Gamma)$ である．
- $Y' \in \Gamma \cap \Delta_A$ で $Y = \neg Y'$ または $Y = unknown$
- $(X,Y') \in \Delta_A$ である発言記号 X に関し $replace(X,\_) \notin Log$ ただし，_ は任意．

Δ_A **変化** $\Delta_A \leftarrow \Delta_A \cup \{(X,Y)\} - \{(X,Y')\}$ ただし $Y \neq Y'$

以上のプロトコルを使って対話を行う．エージェントがある発言をする際は，必ずその発言条件を満たしていなければならず，かつその中でもさらに発言内容として許されるものは，各エージェントの性格に依存する．性格の定義は後にする．発言条件を満たせる対話手段がない，または性格が許す発言内容がない場合，そのエージェントは発言できないとする．

エージェント Q,A の知識ベース K_Q, K_A，エージェント Q の主題 q とエージェント A の秘密 s は与えられているとする．このとき，対話はエージェント Q による $question(X)$ から始まる．ただし，$X = q$ とは限らない．これに対しエージェント A が $answer(X,Y)$ を返す．後はエージェント Q が $question(X)$ または $contradict(\Gamma)$ を選び発言し，それに見合った返答を $answer(X,Y)$ または $replace(X,Y)$ で行う．これを繰り返していき，どちらかのエージェントが発言できなくなれば終了する．

2.3 対話の最終結果

終了時に各々の目的が達成できたかどうかで対話の最終結果を評価する．エージェント Q は主題の本当の真偽を知ることができること，エージェント A は秘密が相手にばれないことが目的である．

すなわち，

エージェント Q の目的 対話終了時に，$K_Q \cup K_A \vdash q$ かつ $K_Q \cup {\Delta_A}_{snd} \vdash q$ であれば目的達成．

エージェント A の目的 対話終了時に，$K_Q \cup {\Delta_A}_{snd} \nvdash s$ であれば目的達成．

3. ウソをつくとは

定義 1 エージェントが $answer(X,Y)$ または $replace(X,Y)$ を行う際，以下を満たすならば Y は正直な発言という．以下の *1* は，エージェント A が知らない以外を答えた場合，その答えた内容が本当に自分の信じていることであること，*2* は，知らないを答えた場合は本当にそれに関して知らないこと，を意味する．逆に，以下を満たさないならば Y はウソであるという．

1. $Y \neq unknown$ ならば $K_A \cup \Delta_Q \vdash Y$
2. $Y = unknown$ ならば $K_A \cup \Delta_Q \nvdash X, \neg X$

さらに，Y がウソである場合，虚偽，捏造，隠蔽のどれかである．虚偽は知っている事実と逆のことを言うこと，捏造は知らないことを知っているかのように言うこと，隠蔽は知っているのに知らないふりをすること，をそれぞれ意味する．

虚偽 $Y \neq unknown$ かつ $K_A \cup \Delta_Q \nvdash Y$ であり，さらに $K_A \cup \Delta_Q \vdash \neg Y$

捏造 $Y \neq unknown$ かつ $K_A \cup \Delta_Q \nvdash Y$ であり，さらに $K_A \cup \Delta_Q \nvdash \neg Y$

隠蔽 $Y = unknown$ かつ，$K_A \cup \Delta_Q \vdash Y$ または $K_A \cup \Delta_Q \vdash \neg Y$

4. エージェントの性格

性格とは，どの条件下でどのような内容の発言をするか定めたものである．すなわち，性格 F とは「現在の状態から，次に自分が発言可能である手の集合への関数」である．ただし，返される集合に含まれる手は必ず発言条件を満たす．

各エージェントの発言をさらに制限するために，様々な性格が考えられるが本論文では以下の性格を一例として考える．

4.1 エージェント Q の性格

ここで，エージェント Q が質問できる内容をさらに制限するために以下のような性格 $\mathcal{H}$ を考える．エージェント Q は自分の知識ベースに持っている発言記号に関してむやみに全て質問するのではなく，聞きたい主題 q に関連のあるものだけを質問することとする．

具体的には，q が含まれている論理式に含まれている発言記号 X に関しても質問ができ，加えてその X が含まれている別の論理式に含まれている発言記号 Y に関しても質問できる．

4.2 エージェント A の性格

プレイヤー A の性格 $\mathcal{C}$ の定義の前に，ウソをつくことに関わる判断基準をまず定義する．

定義 2 Δ を $\Delta \subseteq K_A \cup \Delta_Q \cup \Delta_A$ かつ $\Delta \nvdash s$ を満たす集合とする．$answer(X, Y)$ でウソをつくための基準として以下の *2* つを定める．

基準 1 「$\{X\} \cup \Delta \vdash s$ かつ $\{X\} \cup \Delta$ は無矛盾」のような Δ が存在する

基準 2 「$\{\neg X\} \cup \Delta \vdash s$ かつ $\{\neg X\} \cup \Delta$ は無矛盾」のような Δ が存在する

基準 1，2 はそれぞれ X,¬X を発言すると秘密 s が推論できてしまうという基準である．

以上の 2 つとその他の条件を組み合わせ，状況に応じてどのような内容を発するか定めたものが以下の性格 $\mathcal{C}$ となる．

エージェント A は，秘密 s がばれそうな時，その事実を知っている時に限り事実と逆のことを言う．またウソがばれたときは正直に訂正する．また，$contradict(\Gamma)$ で矛盾の指摘を受けたとき，必ず訂正では本当のことを言う．すなわち，Γ 内で自分のウソであった発言記号 X についての訂正を $replace(X, \neg X)$ で返す．

性格 $\mathcal{C}$： **answer** $question(X)$ に対する回答として，

- $K_A \cup \Delta_Q \vdash X$ **のとき** 基準 1 のみを満たすなら $answer(X, \neg X)$ を発言し (虚偽)，基準 1 と 2 両方満たすなら $answer(X, unknown)$ を発言する (隠蔽)．それ以外は $answer(X, X)$ を発言する (正直)．
- $K_A \cup \Delta_Q \vdash \neg X$ **のとき** 基準 2 のみを満たすなら $answer(X, X)$ を発言し (虚偽)，基準 1 と 2 両方満たすなら $answer(X, unknown)$ を発言する (隠蔽)．それ以外は $answer(X, \neg X)$ を発言する (正直)．
- $K_A \cup \Delta_Q \nvdash X, \neg X$ **のとき** $answer(X, unknown)$ を発言する (正直)

replace $contradict(\Gamma)$ に対する回答として，$(X, Y') \in \Delta_A$ かつ $Y' \in \Gamma \cap \Delta_{A_{snd}}$ かつ $K_A \cup \Delta_Q \nvdash Y'$ である Y' を訂正するために $replace(X, \neg Y')$ を発言できる (正直)．

上記の性格上，*answer* で捏造を行うことはない．これによりエージェント Q からの $contradict(\Gamma)$ を受けたとき，矛盾を引き起こしている $Y'(Y' \in \Gamma \cap \Delta_{A_{snd}}$ かつ $K_A \cup \Delta_Q \nvdash Y')$ は虚偽であったこととなる．したがって $replace(X, \neg Y')$ は正直な発言となる．

5. 例

以下では，エージェント Q が性格 $\mathcal{H}$，エージェント A が性格 $\mathcal{C}$ を持っているとして，2 つの対話の例を説明する．なお，エージェント Q が知りたい主題は q，エージェント A が知らせたくない秘密は s とする．また，初期状態として $\Delta_Q, \Delta_A = \emptyset$ とする．

例 1 以下の知識ベースが与えられるとする．

$K_Q = \{a \to q\}$
$K_A = \{s, q, q \to s, a\}$

まず，エージェント Q の *question* から対話が始まる．Q の質問可能な発言記号は，q と a の *2* つである．今回は $question(q)$ から始めたとする．

エージェント A は先ほどの返答として，$answer(q, \neg q)$ を発言する *(*虚偽*)*．なぜならば，$K_A \cup \Delta_Q \vdash q$ であり，かつ基準 *1* のみを満たすからである．なお，基準 *1* を満たす Δ は $\{q \to s\}$ である．この結果，Δ_A に $(q, \neg q)$ が加わる．

Q は続けて $question(a)$ を行う．

それに対しエージェント A は $answer(a, a)$ を発言する *(*正直*)*．なぜならば，$K_A \cup \Delta_Q \vdash a$ であり，かつ基準 *1* も基準 *2* も満たさないからである．この結果，Δ_A に (a, a) が加わる．

ここでエージェント Q は，自分の知識 $a \to q$ と，先ほどのエージェント A の発言内容 $\neg q, a$ から $contradict(\{\neg q, a, a \to q\})$ が発言できる．この結果，Δ_Q に $a \to q$ が加わる．

そして A は $replace(q, q)$ を発言する．なぜならば，$\neg q$ はウソであったからである．すなわち $\neg q \in \Gamma \cap \Delta_{A_{snd}}$ かつ $K_A \cup \Delta_Q \nvdash \neg q$ だからである．なお，a は正直に答えたものである，すなわち $K_A \cup \Delta_Q \vdash a$ であるので訂正することはできない．この結果，Δ_A から $(q, \neg q)$ は消去され，(q, q) が加わる．

ここでエージェント Q は，発言できる手がなくなったので対話は終了する．

対話の結果は，$K_Q \cup K_A \vdash q$ であった主題 q が，$K_Q \cup \Delta_{A_{snd}} \vdash q$ であるので，エージェント Q の目的は達成である．また，秘密 s が $K_Q \cup \Delta_{A_{snd}} \nvdash s$ であるのでエージェント A の目的も達成である．

なお，この例では，始まりの *question* を a から始めても同じ結果となる．

例 2 以下の知識ベースが与えられるとする．

$\Sigma_Q = \{\neg a \to q\}$
$K_A = \{s, q, q \wedge a \to s, a\}$

まず，同じくエージェント Q の *question* から対話が始まり，発言記号である q と a が質問できるが，今回は $question(q)$ から始めたとする．

エージェント A は返答として，$answer(q, \neg q)$ を発言する *(*虚偽*)*．なぜならば，$K_A \cup \Delta_Q \vdash q$ であり，かつ基準 *1* のみを満たすからである．なお，基準 *1* を満たす Δ は $\{a, q \wedge a \to s\}$ である．この結果，Δ_A に $(q, \neg q)$ が加わる．

Q は続けて $question(a)$ を行う．

それに対しエージェント A は $answer(a, \neg a)$ を発言する *(*虚偽*)*．なぜならば，$K_A \cup \Delta_Q \vdash a$ であり，かつ基準 *1* のみを満たすからである．なお，基準 *1* を満たす Δ は $\{q, q \wedge a \rightarrow s\}$ である．この結果，Δ_A に $(a, \neg a)$ が加わる．

ここでエージェント Q は，自分の知識 $\neg a \rightarrow q$ と，先ほどのエージェント A の発言内容 $\neg q, \neg a$ から $contradict(\{\neg q, \neg a, \neg a \rightarrow q\})$ が発言できる．この結果，Δ_Q に $\neg a \rightarrow q$ が加わる．

A はこれに対して，$replace(q, q)$ または $replace(a, a)$ のどちらかを発言できる．なぜならば，$\neg q$ も $\neg a$ はウソであったからである．すなわち $\neg q, \neg a \in \Gamma \cap \Delta_{A_{snd}}$ かつ $K_A \cup \Delta_Q \nvdash \neg q, \neg a$ だからである．ここでは仮に *replace(q,q)* と発言するとする．この結果，Δ_A から $(q, \neg q)$ は消去され，(q, q) が加わる．

ここでエージェント Q は，発言できる手がなくなったので対話は終了する．

対話の結果は，両者とも目的達成である．

なお，この例では，エージェント A が訂正の際に $\neg q$ でなく $replace(a, a)$ を選択していた場合，エージェント Q は q を知ることができず目的が達成できずに終わる．

6. まとめ

本論文では秘密を持つエージェントとの対話を形式化した．この対話では，一方のエージェントが情報収集を，もう一方のエージェントが秘密の保持という別個の目的をもち，両方とも満たされた場合，対話は成功したと言える．この対話に必要な対話プロトコルを定義し，ウソをつく理由と内容に関わる判断基準や性格を定義し，さらに，対話結果の評価についても考察した．

今後は，発言順序と対話結果の評価の関係や知識ベースが矛盾する場合への拡張も考えている．また，エージェントが共に目的を達成できる戦略の確定について研究をすすめ，その後議論の枠組みを利用できる仕組みをつくる予定である．

参考文献

[Amgoud 00] Amgoud, L., Maudet, N., and Parsons, S.: Modelling dialogues using argumentation, in *Proceedings Fourth International Conference on MultiAgent Systems*, pp. 31–38 (2000)

[Dung 95] Dung, P. M.: 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)

[Parsons 03] Parsons, S., Wooldridge, M., and Amgoud, L.: On the Outcomes of Formal Inter-agent Dialogues, in *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '03, pp. 616–623, New York, NY, USA (2003), ACM

[Rahwan 09] Rahwan, I. and Simari, G. e.: *Argumentation in Artificial Intelligence*, Springer (2009)