3O1-3

不誠実な論証を扱う抽象議論フレームワークに基づく対話モデル

A Dialogue Model for Untrusted Arguments Based on an Abstract Argumentation Framework

國生　一幸　　高橋　和子
Kazuyuki Kokusho　Kazuko Takahashi

関西学院大学大学院　理工学研究科
School of Science & Technology, Kwansei Gakuin University

This report proposes a dialogue model based on an abstract Argumentation Framework (AF). Each agent has her own AF and the prediction of her opponent's AF, and gives arguments based on these AFs. When an argument is given, not only the argument itself, but also arguments and/or attacks that are derived by this argument are added to these AFs. The mechanism in which untrusted arguments are provided and suspected using derivation rules in this model is proposed.

1. はじめに

議論は議題に関しての賛成意見と反対意見を互いに繰り返すことによって構築される。Dung は議論の意見とその反論をグラフにおけるノードとエッジに対応させることにより、議論の構造を抽象化する議論フレームワークを定義した [Dung 95]。議論フレームワークでは、論証の受理性を攻撃関係から構成されるラベリングによって判断する方法がある [Baroni 11]。これらの定義を基に、多くの研究が行われている。

議論フレームワークに基づく対話モデルとしては、各エージェントが独立した知識ベースをもち、対話の進行とともにそれが更新されると考えるのが自然である。この場合、提議された論証が別の論証を誘発する場合がある。Satoh らは、誘発を derivation rule として記述し、動的な対話の仕組みを論理プログラミングで実現した [Satoh 11]。

[Satoh 11] で扱われた例を一部変更した以下の対話例を考える。ジョンが駅の駐車場に止めてある車から財布を盗んだのかについて、**P** と **C** が議論している。

P「ジョンは駅の駐車場に止めてある車から財布を盗んだ」
C「ジョンが駅の駐車場に止めてある車から財布を盗んだという証拠はない」
P「証拠ならある、被害にあった車の中にジョンの学生証が落ちていた」
C「ジョンは学生証を他の人に貸していたから、持っていない」
P「証拠ならまだある、ジョンを駅の駐車場で目撃した人がいる」
C「ジョンは図書館にいたよ、駅の駐車場にはいない」
P「図書館は学生証がないと入れないよ、本当は図書館にいなかったんじゃないのか」

これは、「学生証を持っていない」、「図書館にいた」という発言によって、エージェントの知識が動的に増えていき、それを使って推論した結果、矛盾をつくという例である。Satoh らはこの仕組みを反論の繰り返しとして扱っており、相手の知識は予測していない。本研究では、相手の発言に対して単なる反論だけでなく、相手の知識を予測した嘘指摘をプロトコルとして取り入れ、嘘をついたりそれを指摘する仕組みを扱うモデルを提案する。

連絡先: 國生　一幸，関西学院大学大学院 理工学研究科 高橋和子研究室，兵庫県三田市学園 2 丁目 1 番地，079-565-8391，dbq68680@kwansei.ac.jp

Sakama は議論フレームワークにおける受理集合の考え方を使って、正直な論証と不正直な論証を定義し、嘘を含む論争ゲームを考案した [Sakama 12]。しかし、Sakama の研究では矛盾する発言を指摘する方法を提案しているが、嘘を指摘する手法は提案していない。また、論証の誘発についても考慮していない。嘘を指摘するためには相手が発言していない事実を知っており、その事実が相手の発言を攻撃していることを指摘しなければならない。Sakama の論争ゲームでは相手が発言していない事実を予測する手段がないため嘘指摘ができない。

Yokohama は相手の予測知識を与えることにより、嘘を指摘する手法を提案した [Yokohama 16]。この手法は誘発についても考慮しているが、予測知識には様々な条件を設けなければならない。

嘘を指摘するためには相手が発言していない事実を知っており、その事実が過去の発言を攻撃していることを指摘しなければならない。本研究では誘発規則を用いることにより、相手が発言していない事実を予測し、その事実を用いて嘘指摘する手法を提案する。

本発表は以下のように構成される。2 節では、論争ゲームにおける基礎知識と論争ゲームについて記述する。3 節では、誘発規則の定義、嘘と矛盾についての定義、誘発規則を用いた嘘指摘をする手法について記述する。4 節では、誘発規則を用いて嘘を指摘する対話例を記述する。5 節では、まとめと今後の課題について述べる。

2. 論争ゲーム

2.1 基礎知識

議論フレームワークは $AF=(AR, AT)$ のように定義される。ここで、AR は論証の集合、AT は攻撃関係である。攻撃関係は論証の集合の二項関係であり、攻撃関係 (A, B) は論証 A が論証 B を攻撃していることを表している。

AR から $\{in, out, undec\}$ への関数 L が、$A \in AR$ に対して下記の条件を満たすとき、完全ラベリングであるという。

- $L_{AF}(A) = in \Leftrightarrow \forall B \in AR, (B, A) \in AT$ ならば $L_{AF}(B) = out$
- $L_{AF}(A) = out \Leftrightarrow \exists B \in AR, (B, A) \in AT$ かつ $L_{AF}(B) = in$
- $L_{AF}(A) = undec \Leftrightarrow L_{AF}(A) \neq in$ かつ $L_{AF}(A) \neq out$

ラベリングは議論フレームワークにおける各論証の受理性を判定している。ラベリングが in であるならば論証は受理されており、out なら論証は受理されない。また、$undec$ である場合は論証が受理されているかどうか分からない。

2.2 論争ゲーム

論争ゲームは提議エージェント $\mathbf{P}$ と対立エージェント $\mathbf{C}$ により行われる。論争ゲームの目的は $\mathbf{P}$ が初めに主張する議題について、$\mathbf{C}$ に受理可能性を判定してもらうことである。

論争ゲームにおけるすべての知識を表す議論フレームワークを $UAF = (AR, AT)$ と記述する。UAF の部分議論フレームワーク $AF_X = (AR_X, AT_X)$ を $AR_X \subseteq AR, AT_X \subseteq AT$ と定義する。AF_X が UAF の部分議論フレームワークであるとき、$AF_X \subseteq UAF$ と記述する。

論争ゲームにおいて、初期状態で $\mathbf{P}$,$\mathbf{C}$ がそれぞれ AF_P,AF_C を AF として持つとする。ただし、$AF_P, AF_C \subseteq UAF$ とする。$\mathbf{P}$ が最初に主張する論証を Ag とする。論争ゲームは $\mathbf{P}$ によって選択された Ag を基にして、対話が進行する。

エージェントは論争ゲームにおいて主張を発言するかパスをすることができる。主張とは $\mathbf{P}$ による議題の発言または、エージェントが相手の主張を否定する発言である。パスとはエージェントが何も発言しないことである。議題を発言するには、議題に用いる論証が $\mathbf{P}$ の AF に存在していなければならない。相手の主張の否定を発言するには、エージェントのもつ AF に相手が主張に用いた論証に対して、攻撃関係がある論証が存在していなければならない。同じ議題を主張することはできず、主張できる議題がなければパスをする。

論争ゲームでは、まず、$\mathbf{P}$ が主張により議題を発言する。この議題に対して $\mathbf{C}$ は主張を発言するかパスをする。$\mathbf{P}$ は $\mathbf{C}$ の主張に対してまた、新たな主張を発言する。このようにして、論争ゲームは進行していく。

主張に用いられた論証がエージェントの持つ AF に存在しない場合、$AF = (AR, AT)$ は更新される。AR には主張に用いられた論証が新たに加えられる。AT にはエージェントが持つ論証と加えられた論証から成り立つ攻撃関係が新たに加えられる。

論争ゲームが終了するのは各エージェントがパスを続けて主張したときである。このとき、$\mathbf{C}$ が知識ベースとして持つ AF のラベリングにおいて、Ag が in であるならば、$\mathbf{P}$ が論争ゲームに勝利し、in でないならば、$\mathbf{P}$ が論争ゲームに敗北する。

3. 誘発規則を用いた戦略の考案

3.1 誘発規則

誘発規則は $A \Leftarrow B$ という導出規則で表す。これは論証 B が論証 A を引き起こすことを意味し、A を誘発された論証と呼ぶ。B の部分は一般に複数の論証が記述できるが、ここでは説明の簡単化のため一つにしている。誘発規則を知っていれば、相手の発言 B から誘発された A も自分の AF に追加し更新する。また、誘発された論証からさらに別の論証が誘発されることもあり、論証の誘発が終わるまで AR は更新される。AR の更新が終わると、AT にはプレイヤーが持つ論証と誘発により加えられた論証から成り立つ攻撃関係が新たに加えられる。

3.2 嘘と矛盾

論争ゲームにおける正直と嘘について定義する。エージェントが正直な主張をするということは、自分の持つ AF の論証のラベルが in である論証を主張するということである。また、エージェントが嘘の主張をするということは、自分の持つ AF の論証のラベルが in でない論証を主張するということである。

論争ゲームにおける矛盾について定義する。矛盾した主張とは、あるエージェントが自分の過去の発言と現在の主張における論証の間で攻撃関係が存在することである。

3.3 誘発規則を用いた嘘指摘

では、誘発規則を用いて嘘を見破る方法について説明する。任意のエージェントを $\mathbf{X}$,$\mathbf{Y}$ とする。$\mathbf{X}$ が誘発規則により誘発された A を主張したとき、$\mathbf{X}$ は誘発に用いられてる B を自分の AF に所持していることになる。このとき、$\mathbf{Y}$ がその誘発規則を知っていたならば、$\mathbf{X}$ は誘発規則を用いて A を主張したことを $\mathbf{Y}$ も知ることができるため、$\mathbf{Y}$ は主張した A だけでなく、誘発に用いられた B も知ることができる。

ここで、$\mathbf{Y}$ の予測 AF を EAF_X と定義する。EAF_X は $\mathbf{X}$ の AF に対する $\mathbf{Y}$ の予測であり、$\mathbf{Y}$ が所持する。EAF_X は $\mathbf{X}$ と $\mathbf{Y}$ が主張した論証と誘発規則によって誘発された論証で構成される。EAF_X に存在する論証のうち、$\mathbf{X}$ が主張した論証に対して、誘発された論証が攻撃しているときに、$\mathbf{X}$ が嘘をついていると $\mathbf{Y}$ は判断する。

4. 例

ジョンが駅の駐車場に止めてある車から財布を盗んだのかについて、$\mathbf{P}$ と $\mathbf{C}$ が議論している。

論証:

Ag, A, B, C, D, E, F, G

攻撃関係:

$(A, Ag), (B, A), (C, B), (D, C), (C, D), (E, A), (F, E), (G, F), (F, G)$

Ag:ジョンは駅の駐車場に止めてある車から財布を盗んだ

A:ジョンが駅の駐車場に止めてある車から財布を盗んだという証拠はない

B:被害にあった車の中にジョンの学生証が落ちていた

C:ジョンは学生証を他の人に貸していたから、持っていない

D:ジョンは学生証を貸していない

E:ジョンを駅の駐車場で目撃した人がいる

F:ジョンは図書館にいた

G:ジョンは図書館にいない

誘発規則:$G \Leftarrow C$

$G \Leftarrow C$:ジョンは学生証を持っていないため、学生証が必要な図書館に入れない

初期状態の $\mathbf{P}$ の AF: $AF_P = (\{Ag, B, E\}, \emptyset)$

$\mathbf{P}$ が予想する $\mathbf{C}$ の知識:$EAF_C = (\emptyset, \emptyset)$

$\mathbf{P}$ が持つ誘発規則:$G \Leftarrow C$

初期状態の $\mathbf{C}$ の AF:

$AF_C = (\{A, C, D, F, G\}, \{(D, C), (C, D), (G, F), (F, G)\})$

論争ゲーム

$\mathbf{P}$ は Ag:「ジョンは駅の駐車場に止めてある車から財布を盗んだ」と主張する。$\mathbf{C}$ は Ag を知ったことにより、AF_C に Ag が加えられる。EAF_C にも同様に Ag が加えられる。

$\mathbf{C}$ は Ag に対して、A:「ジョンが駅の駐車場に止めてある車から財布を盗んだという証拠はない」と主張する。$\mathbf{P}$ は A を知ったことにより、AF_P に A が加えられる。EAF_C にも同様に A が加えられる。

$\mathbf{P}$ は A に対して、B:「被害にあった車の中にジョンの学生証が落ちていた」と主張する。$\mathbf{C}$ は B を知ったことにより、AF_C に B が加えられる。EAF_C にも同様に B が加えられる。

C は B に対して、C:「ジョンは学生証を他の人に貸していたから、持っていない」と主張する。このとき、**C** は D:「ジョンは学生証を貸していない」ということを知っており、かつ、攻撃関係:$(D,C),(C,D)$ も知っているため、嘘の主張をしたのである。**P** は C を知ったことにより、AF_P に C が加えられる。このとき、**P** は $G \Leftarrow C$ により G も知ることができるので、AF_P に G も加えられる。この時点で、各エージェントは図 1 のような AF を持つ。

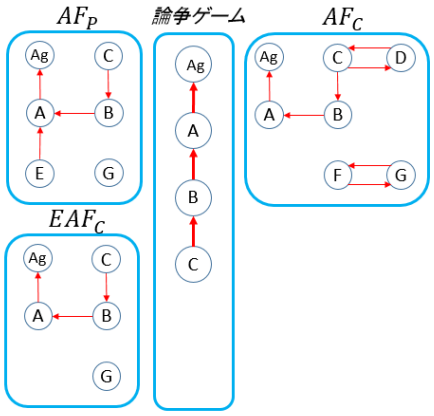

図 1: **C** が C を主張した時点での各エージェントの持つ AF

P は A に対して、E:「ジョンを駅の駐車場で目撃した人がいる」と主張する。**C** は E を知ったことにより、AF_C に E が加えられる。EAF_C にも同様に E が加えられる。

C は E に対して、F:「ジョンは図書館にいた」と主張する。このとき、**C** は G:「ジョンは図書館にいない」ということを知っており、かつ、攻撃関係:$(G,F),(F,G)$ も知っているため、嘘の主張をしたのである。**P** は F を知ったことにより、AF_P に F が加えられる。EAF_C にも同様に F が加えられる。この時点で、各エージェントは図 2 のような AF を持つ。

ここで、**P** は EAF_C の F,G に注目する。F は **C** が主張した論証であり、G は **C** が持っていると予測した論証である。G は F を攻撃しているため、F のラベルが $undec$ である。このことから、**C** は F に受理性がないと分かっていながら主張したと判断できるため、**P** は **C** が嘘をついたと指摘できるのである。

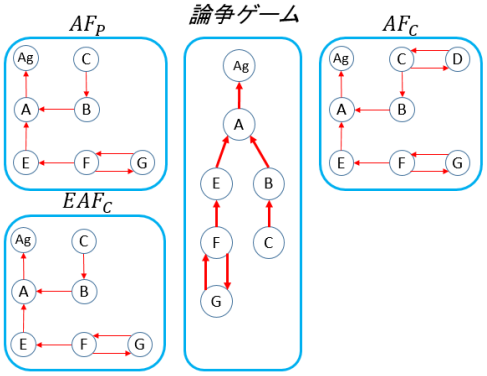

図 2: **P** が G の主張により F は嘘であると指摘した時点での各エージェントの持つ AF

5. まとめ

本研究では、誘発規則を用いることにより相手が発言していない事実を予測し、その事実を用いて嘘を指摘する手法を提案した。

今後は、誘発規則を用いた嘘指摘により、エージェントが論争ゲームで勝利する戦略を考案する予定である。

参考文献

[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, pp. 321-357 (1995).

[Baroni 11] Baroni, P., Caminada, M., and Giacomin, M.: An introduction to argumentation semantics, The Knowledge Engineering Review, Vol. 26: 4, pp. 365-410 (2011).

[Satoh 11] Satoh, K., and Takahashi, K.: A Semantics of Argumentation under Incomplete Information, JURISIN 2011 Program (2011).

[Sakama 12] Sakama, C.: Dishonest Arguments in Debate Games, Proceedings of the 4th International Conference on Computational Models of Argument (COMMA 2012), Frontiers in Artificial Intelligence and Applications, Vol. 245, pp. 177-184 (2012).

[Yokohama 16] Yokohama, S. and Takahashi, K.: What Should an Agent Know Not to Fail in Persuasion?, EUMAS AT 2015 2015: Multi-Agent Systems and Agreement Technologies pp. 219-233 (2015).