抽象議論理論的な同盟形成意味論

蟻坂 竜大　佐藤 健

抽象議論理論 (abstract argumentation theory) 的な同盟形成の意味論を提案する。例えば政党連合では、連合の形成には利点があるはずだが、同時に連合形成によって個々の政党の行動が制限されることが多い。こういった性質を表現するにあたり、典型的な抽象議論理論においての、容認されうる議論の集合は内部対立を起こさない (conflict-freeness)、という制約はふさわしくない。そのため、容認されうる議論の集合は内部対立を起こしていてもよいが、完全批判される議論を含まない (conflict-eliminability) という弱い制約が適切である。本論文では、conflict-eliminability を満たす集合間の同盟利益形成の意味論を抽象議論的に形成する。

We contemplate on abstract-argumentation-theoretic coalition formation semantics. Such semantics are interesting in several respects. They require a weaker notion than the typical conflict-freeness among sets of arguments, and they are semantics relativised to a certain set of arguments.

1　はじめに

エージェントの同盟形成の研究は様々な分野で行われているが、同盟を、その同盟に関わるエージェントの主張の集合と捉えることもできる。そのため、AI の知識表現分野の抽象議論理論を用いて議論主体の同盟形成意味論を考察することに意義はある。関連研究 [1] [6] [17] では、Dung の容認意味論 (acceptability semantics) を応用して容認できるエージェントの同盟を考察しており、焦点は利己的な同盟構成員の利益が最大となる集合を導くことにある。これらの研究では主に協力的な同盟が考察されており、同盟内には内部対立はなく、エージェントが同盟に属することで何かを失う必要も無い。

しかし、全ての同盟が協力的なものであるというわけではなく、政治の同盟では、例えば自民党に見られるように、派閥間で政策の食い違いがおこることは起こりうる。ここでは我々は政党の類の同盟を考察する。以下はそのような同盟のもつ興味深い特徴である。

- 個人的というより組織的：党内外を問わず気の合うエージェント同士で自由に協力しあうことはできない。
- 内部対立：　政党員は政党の政策を支持すべきで、それに関して他の政党員と争うことはない、と考えられる。しかし、政策の詳細に関しては政党の派閥同士で対立することは起こる。その結果、政党所属員が対外的に党員として発言する場合には自己の主張をある程度抑えなくてはならないことがなる。
- 政党間での非対称的な対立： 政党としての主張は対外的には矛盾を含むものであってはならないので、内部対立が派閥間で起こる点は政党としての主張に含むべきでない。しかし、他政党はそういった政党の事情に束縛されないので、その政党内の派閥の主張を批判することをもってその政党に対する批判とすることはしばしば見られる。

本論文では、我々は抽象議論的にこのような特徴を持つ同盟を扱い、同盟利益また同盟形成の意味論を考える。具体的には、（１）議論集合が利益を得る判断基

On Argumentation-Theoretic Coalition Formation Semantics

Ryuta Arisaka and Ken Satoh, 国立情報学研究所, National Institute of Informatics.

準が与えられた際、内部対立を含みうる議論集合 A はどの他の内部対立を含み得る議論集合と同盟を形成することで利益を得るか、（２）また、そのような利益意味論、さらに同盟のユーティリティー公準が与えられた際、A はいずれの他の議論集合と実際に同盟を組得るか、という２つの問いを考察する。

テクニカルな部分でこれらの意味論が面白い点は、まず第一にこれらは従来の議論容認意味論のような、与えられた議論集合のうちのどの集合が容認されるか、というグローバルな意味論ではなく、意味論自体がある議論集合に相対化されていることである。第二に、上述の内部対立また非対称性に対応するため、(1) Dung の容認される議論集合は内部対立を含まない、という制約より弱い、議論集合の構成議論が完全に批判されない限り内部対立は許容される、という制約を設定すること、（２）また内部対立を含む議論集合（政党構成員の視点）から内部対立を仮に除いた際に得られる内部対立を含まない議論集合（政党の視点）をともに考えること、（３）そして、それらを用いて、ある議論集合がどの集合外の議論から批判を受けるか、またその議論集合はそれらの議論を批判できるか、を考察すること、が必要となる。第一の点と同様、ここでも単に議論集合が前述の弱い制約を満たしているかが問題となっているわけではなく、議論集合内で議論が完全に批判されないが部分的な批判を受ける場合、その議論集合が内部対立を含まないためにはどの程度その議論を弱めるのか、を知らなければならない。対処として、我々は全ての議論に数的な容量を与え、また全ての批判にも批判数を与えることとする。

1.1 関連研究

Dung の容認意味論をもって同盟形成を考察する研究はすでにあげた [1] [6] [17]。これらの研究では利己的なエージェントによる内部対立を含まない種類の同盟を扱っている。[1] では、同盟に目標というメタな情報を与え、また、エージェントがどの目標を遂行できるか、ということも設定している。これに比較して、我々はより抽象議論的に同盟形成を考え、数値以上にメタな情報は取り入れない。代わりに、我々は同盟利益を、（１）集合の大きさ、（２）ある議論集合 A が議論集合 B と同盟を形成するとした際、A を批判していた議論の数は同盟形成後に減少するか、（３）議論集合がどのくらい弁護されているか、によって判断することにする。

1.1.1 対立許容抽象議論

内部無対立の公準を見直す研究はなされてきている。２価意味論 [16] は reductio ad absurdum を用いて矛盾を解決する。Arieli は容認、排斥、無関心、容認か排斥か判断しかねる、という４つのラベルを使うことによって内部無対立を緩和している [4]。しかし、これらのアプローチでは数値を直接取り扱わないため、批判の強度を表すことが難しい。重み付き抽象議論 (Weighted Argumentation) [12] では数値が批判関係に付与される。また、グローバルな inconsistency budget という数値も与えられる。この理論では、議論集合内で起こる批判の数値を足しあわせたものが矛盾の度合いで、それが inconsistency budget より小さければその集合の矛盾は許容できる、としている。しかし、我々のセッティングでは内部対立を含む議論集合から内部対立を除いた集合を得ることが必要であるため、このアプローチは十分でない。また、実質的には、というのも、対外的には、我々の扱う議論集合は矛盾を含んでいてはいけないので、矛盾許容は内部対立許容とはやや趣旨の異なるものである。これらと比較すると [14] [13] は我々のアプローチに近い。そのため、これらを応用することは可能であると考えるが、抽象議論理論の批判を論理的に考察することも重要である [2] [3] [7] ことも踏まえて、ここではより axiomatic なアプローチをとることにする。

1.1.2 動的抽象議論

我々のフレームワークはある種動的である。これまでの研究では構造的議論理論から反論不能な議論を変更したり [18]、新たな議論を足したり [8]、批判を修正したり [10]、命題論理を通して議論フレームワークを修正したり [9]、議論フレームワークを議論フレームワークで修正したり [5] することが行われてきた。これらは主にインプットがどのように議論フレームワークを変更すべきか、に焦点をあてている。我々のセッティングでは、対立の非対称性を表現するにあたって

内部対立解消前と解消後の両方が必要となるため、ある意味異なる視点のコーディネーションを行うことになる。我々の知りうる限りではそのような研究は上記の関連文献ではなされていない。

2　抽象議論理論

Dung が提唱する抽象議論理論 [11] は様々な方法で拡張されているが、本論文では Nielsen-Parsons のより一般的な議論フレームワーク [15] を用いる。定義を以下に挙げる。

議論（主張）は抽象的な存在とし、$\mathcal{A}$ を全ての議論のクラスとする。$\mathcal{A} \subseteq_{\text{fin}} \mathcal{A}$ また $G : (2^{\mathcal{A}} \backslash \emptyset \times \mathcal{A})$ とした場合、(A, G) を議論フレームワークと呼ぶ。$A_1 \subseteq A$ また $A' \subseteq A$ かつ $a \in A$ とした場合、$(A', a) \in G$ の場合に限り A_1 は a を批判するとみなす。$(A', a) \in G$ であるが、$A'' \subset A'$ となる全ての A'' に関して $(A'', a) \notin G$ となる場合に限り (A', a) を最小とみなす。A_1 に批判される $a \in A_1$ が無い場合に限り A_1 を無対立議論集合 (conflict-free arguments set) とする。$a \in A$ の時、全ての最小の $(A_x, a) \in G$ に対して A が少なくともひとつの $a_x \in A_x$ を批判する場合に限り、A は a を弁護するとみなす。A_1 が無対立議論集合かつ全ての $a \in A_1$ を弁護する場合に限り、A_1 は容認されるとみなす。容認される議論集合のうち、自身以外に上位集合を持たないものを、特に理想的な議論集合 (preferred extension もしくは preferred set) とする。

3　対立許容議論フレームワーク

我々のフレームワークの定義は以下のとおりである。$\mathbb{N}$ を０を含む自然数とする。また $\mathcal{S}$ を $\mathcal{A} \times \mathbb{N}$ とする。以下、$\mathcal{A}$ ではなく $\mathcal{S}$ を全ての議論のクラスとし、s（サブスクリプトありまたはなし）は S の構成員を指すものとする。

$S_1 \subseteq \mathcal{S}$ のうち、以下の３つの制約を満たすものを一貫性のある議論集合と呼ぶ。

- S_1 は有限集合。
- S_1 の全ての構成員 (a, n) について $n > 0$。
- $(a, n) \in S_1$ ならば S は $n = m$ でない (a, m) を含まない。

１つめの制約は通常の議論フレームワークで仮定されていることが多い。２つめの制約の説明として、我々は議論は容量を持たせるとしたことを思い返してもらいたい。容量が具体的に何であるかは解釈によるが、大きいほど議論としての情報量が多いとする。容量０というのはその議論に議論としての情報がないとみなすため、議論を扱う上で n が０以下となる (a, n) は考慮しない。３つめの制約に関して、議論 (a, n) の a は (a, n) の ID と仮定している。これはその議論の発言者のユニークな ID と捉えていただいてよい。一貫性のある議論集合では同じ ID は２度使われない。これ以降、特に断らない限り、我々は一貫性のある議論集合を考察しているものとする。

例 1　以下の４つの議論を考えてみる。

a_1　中国韓国との国際的な協力を支持する。高速道路の建設を支持する。小農家の保護を支持する。ガソリン税の導入を支持する。

a_2　中国韓国との国際的な協力を支持する。自由貿易を支持する。消費税の増税を支持する。

a_3　靖国神社参拝を支持する。高速道路の建設に反対する。アメリカとの軍事関係の強化を支持する。車に関しては自由貿易を支持する。小農家の保護に反対する。

a_4　靖国神社参拝に反対する。

簡単な容量としては、各議論を構成する詳細な議論の数を考えることもできる。その場合、上記の議論は我々のフレームワークでは $(a_1, 4), (a_2, 3), (a_3, 5), (a_4, 1)$ となる。

さて、批判関係に移る。R を $2^{\mathcal{S}} \times \mathcal{S} \rightharpoonup$ とし、以下の制約を満たすものとする。$R(S, s)$ は S の s に対する批判の強度を表す。また、(S, s) が R で定義されている、と書くこともあるし、$R(S, s)$ が定義されている、と書くこともある。両者に区別はない。

1. $R(\emptyset, s)$ は定義されない。**[Consistency]**
2. 全ての $S_1 \subseteq S$ また、全ての $s \in S$ において、(S_1, s) が R で定義される時、$\emptyset \subset S_2 \subseteq S_1$ ならば (S_2, s) は R で定義される。**[Quasi-closure by subset relation]**
3. 全ての $S_1, S_2 \subseteq S \subseteq_{\text{fin}} \mathcal{S}$ また全ての $s \in S$ にお

いて、(S_1, s) また (S_2, s) が R で定義されるなら, $(S_1 \cup S_2, s)$ もまた R でていぎされる。**[Closure by set union]**

4. 全ての $S_1 \subseteq S \subseteq_{\text{fin}} \mathcal{S}$ また全ての $s \in S$ において、(S_1, s) が R で定義されるなら, $R(S_1, s) > 0$ が成り立つ。**[Attack with a positive strength]**
5. $n \leq m$ となる全ての $(a, n), (a, m) \in \mathcal{S}$ において以下が成り立つ。$s \in \mathcal{S}$ また $(a, n) \in S_1$ となる $S_1 \subseteq_{\text{fin}} \mathcal{S}$ において $R(S_1, s)$ が定義されるならば、$S_2 = (S_1 \backslash (a, n)) \cup (a, m)$ において $R(S_2, s)$ は定義され、また $R(S_1, s) \leq R(S_2, s)$ が成り立つ。**[Attack monotonicity 1 (source)]**
6. 全ての $S_1, S_2 \subseteq S \subseteq_{\text{fin}} \mathcal{S}$ また全ての $s \in S$ において, (S_1, s)、(S_2, s) また $(S_1 \cap S_2, s)$, が R で定義されるなら, $i = 1$ また $i = 2$ において $R(S_1 \cap S_2, s) \leq R(S_i, s)$ が成り立つ。**[Attack monotonicity 2 (source)]**
7. $n \leq m$ となる全ての $(a, n), (a, m) \in \mathcal{S}$ において、 $S_1 \cap \bigcup_{l \in \mathbb{N}} \{(a, l)\} = \emptyset$ となる $S_1 \subseteq_{\text{fin}} \mathcal{S}$ があり、$(S_1, (a, n))$ が R で定義されるなら, $(S_1, (a, m))$ も R で定義され, また $R(S_1, (a, n)) \leq R(S_1, (a, m))$ が成り立つ。**[Attack monotonicity 3 (target)]**
8. 全ての $S_1 \subseteq_{\text{fin}} \mathcal{S}$ また全ての $s \in \mathcal{S}$ において、$s \in S_1$ ならば $R(S_1, s)$ は定義されない。**[No self attacks]**

[Coherence] により、批判は無から起こらない。**[Quasi-closure by subset relation]** により、集合からある議論に対して R が定義される場合、空集合を除いたその集合の全ての下位集合からその議論に対してい R が定義される。これは Nielsen-Parson の議論フレームワークと比較すると違う点ではあるが、我々は批判と完全批判を区別するので本質的な差とはならない。**[Closure by set union]** は **[Quasi-closure by subset relation]** の逆の制約である。**[Attack with a positive strength]** の目的は以下の通りである。前述のように、$R(S_1, s)$ は S_1 から s への批判の強度を表す。仮にその値が 0 だとすると、要するに S_1 は s を批判していない、と考えることができるので、そのような関係は考慮する必要がない。**[Attack monotonicity 1]** は以下のような性質を表す。R は $S_1 \subseteq S \subseteq_{\text{fin}} \mathcal{S}$ から $s \in S$ に定義されているとした場合、他は変えずにある $s_1 \in S_1$ に関して $R(\{s_1\}, s)$ を増加させたとする。その場合、その変更後の S_1、 S_1'、からの s への批判は依然として起こるべきである。**[Attack monotonicity 2]** は、仮に集合がある議論を批判する場合、その集合の上位集合はその批判の強度を減少させない、ということを表す。**[Attack monotonicity 3]** の説明として, とある集合がとある強度である議論を批判しているとする。ここで、その批判を受ける議論の容量を増加させた場合、我々が示した議論容量の理解によれば、容量が増加した議論は増加前の情報を含み、さらにそれ以上の情報を含み得るわけであるから、その集合は当然変更後の議論を変更前と同程度またはさらに強く批判する。**[No self attacks]** は矛盾的な議論が現れることを許さない。

[Attack monotonicity 1] の一般化も成り立つ。π を自然数と何かしらの順序対をとる関数とし、$\pi(n, \Gamma) = \{n$ 番目の Γ の構成員 $\}$ とする。

公理 1 $S_1 \subseteq_{fin} \mathcal{S}$ また $s \in \mathcal{S}$ において $R(S_1, s)$ が定義されるものとする。$S_2 \subseteq_{fin} \mathcal{S}$ が *(1)* $\bigcup_{s_x \in S_1} \pi(1, s_x) = \bigcup_{s_x \in S_2} \pi(1, s_x)$ また *(2)* $(a, n) \in S_1$ ならば $n \leq m$ において $(a, m) \in S_2$、 を満たすならば、$R(S_2, s)$ は定義される。

我々の議論フレームワークは $\mathcal{S}$ を一貫性のある議論集合として、$(\mathcal{S}, R)$ と定める。

3.1 批判

R を用いて批判と完全批判を以下に定義する。

定義 1 $R(S_2, s)$ が定義される $S_2 \subseteq S_1$ が存在する場合に限り、$S_1 \subseteq S$ は $s \in S$ を批判する、とする。S_1 が s を批判し、なおかつ（1）$R(S_2, s)$ が定義され、（2）$S_x \subseteq S_1$ となる S_x において $R(S_x, s)$ が定義されるならば $\pi(2, s) \subseteq R(S_2, s)$、となる $S_2 \subseteq S_1$ が存在する場合に限り、$S_1 \subseteq S$ は $s \in S$ を完全批判

する、とする。

よって批判が完全批判となるのは批判を行う集合のある下位集合から対象の議論への批判強度が、その議論の容量より少なくない場合である。

定義 2 (最大批判強度) $V^{\max}(S_1, s)$ を以下のように定義する。S_1 が s を批判しない時０である。さもなければ、（１）$R(S_2, s)$ が定義され、（２）$S_x \subseteq S_1$ となる (S_x, s) において R が定義されるならば $R(S_x, s) \leq R(S_2, s)$、を満たす $S_2 \subseteq S_1$ による批判強度 $R(S_2, s)$ である。

例 2 例１に戻ろう。$(a_1, 4)$ は $(a_3, 5)$ と３つの点（靖国神社参拝が中国韓国との国際協力に支障をきたすもの、とする）において対立する。このため、批判強度を対立する主張の数、と単純に解釈した場合、$R(\{(a_1, 4)\}, \{(a_3, 5)\}) = R(\{(a_3, 5)\}, (a_1, 4)) = 3$, となる。

3.2　対立許容議論集合

定義 3 (対立許容議論集合) S_1 を構成する議論が S_1 により完全批判されることのない場合に限り、S_1 を対立許容議論集合とする。

対立許容は無対立よりも弱い制約である。対立許容を定める動機はすでに述べた。

定義 4 (実質議論) S_1 が対立許容議論集合である場合に限り、$\alpha : 2^{\mathcal{S}} \rightharpoonup 2^{\mathcal{S}}$ は S_1 において定義されるものとする。$\alpha(S_1)$ が定義される場合、$\alpha(S_1) = \{(\pi(1, s), n) \mid s \in S_1 \text{ and } n = \pi(2, s) - V^{\max}(S_1, s)\}$ と定める。また、$\alpha(S_1)$ を S_1 の実質議論と呼ぶ。

例２において、$\{(a_1, 4), (a_3, 5)\}$ の実質議論は $\{(a_1, 1), (a_3, 2)\}$ となる。

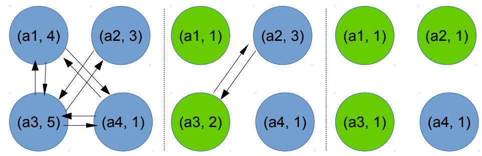
公理 2 全ての $S_1 \subseteq S$ において, α が S_1 で定義されているならば, 全ての $\alpha(S_1)$ の構成員は $\mathcal{S}$ の構成員である。特に、$(a, -n) \in \alpha(S_1)$ となる $a \in \mathcal{A}$ $n \in \mathbb{N}$ は存在しない。

実質議論は無対立である。

定義 5 (実質議論の視点) $\mathsf{Del}_R(S, S_x)$ を $\{(S_y, s) \mid s \in S_x \subseteq S \text{ and } S_y \subseteq S_x \text{ and } R(S_y, s)$ が定義される。$\}$ とおく。また、S_1 を S の下位集合とする。α が S_1 において定義されるならば、$((S \backslash S_1) \cup \alpha(S_1), R \backslash \mathsf{Del}(S, S_1))$ を S_1 が S に関してもつ視点とする。また、それを $\mathsf{View}_R(S, S_1)$ と表す。

公理 3 $\alpha(S_1)$ が定義されるならば $\alpha(S_1)$ は $\mathit{View}_R(S, S_1)$ で無対立である。

例 3



左の図は例１の議論を、議論間の批判を踏まえて図式化したものである。真ん中の図、また右の図はそれぞれ $\{(a_1, 4), (a_3, 5)\}$ と $\{(a_1, 4), (a_2, 3), (a_3, 5)\}$ の左図に対する視点である。[†1] 中図の $(a_1, 4)$ は a_1 のガソリン税支持を表す。また、$(a_3, 2)$ はアメリカとの軍事関係の強化、車の自由貿易を表す。実質議論の視点の定義、また R の定義により、$(a_1, 1)$ と $(a_3, 2)$ の間には批判関係はない。（右図の $(a_1, 1), (a_2, 1), (a_3, 1)$ の間でも同様である。）一方で車の自由貿易と $(a_2, 3)$ の自由貿易は詳細な点については対立するため、批判関係がある。$(a_1, 1)$, $(a_2, 3)$, $(a_3, 2)$ と $(a_4, 1)$ の間では対立は起こらない。右図の中図と同様に考える。

3.3　同盟批判、同盟容認議論集合また同盟理想議

†1　一言のことわりとして、これらは左図から唯一に決定されるものではない。これらは例１の具体的な議論またその批判関係より得ている。

論集合

対立許容議論集合は外部の議論を実質議論によってのみ批判可能である。

定義 6 (同盟批判また同盟完全批判) $\alpha(S_1)$ が定義され、(S_2, s) において $\pi(2, \mathsf{View}_R(S, S_1))$ が定義されるような $S_2 \subseteq \alpha(S_1)$ が存在する場合に限り、S_1 は s を同盟批判する、と定める。S_1 が s を同盟批判し、$\pi(2, \mathsf{View}_R(S, S_1)))(\alpha(S_1), s) \geq \pi(2, s)$ が成り立つ場合に限り、S_1 は s を同盟完全批判する、と定める。

我々の例では、$\{(a_3, 5)\}$ から $(a_4, 1)$ が同盟完全批判となる。$\{(a_1, 4), (a_3, 5)\}$ から $(a_2, 3)$ は同盟批判となる。明らかではあるが、対立許容議論集合は自身の構成員を同盟批判しない。

公理 4 以下は同義である。

1. $S_1 \subseteq S$ は $s \in S$ を同盟批判する。
2. $S_1 \subseteq S$ は $s \in \pi(1, \mathit{View}_R(S, S_1))$ を同盟批判する。

以下も同義である。

1. $S_1 \subseteq S$ は $s \in S$ を同盟完全批判する。
2. $S_1 \subseteq S$ は $s \in \pi(1, \mathit{View}_R(S, S_1))$ を同盟完全批判する。

同盟容認議論集合、また同盟理想議論集合も2章で取り上げた容認議論集合、理想議論集合を対立許容議論集合にアレンジしたものとなる。

定義 7 (同盟容認（理想）議論集合)
S_1 が対立許容議論集合であり、なおかつ $S_2 \subseteq \pi(1, \mathsf{View}_R(S, S_1))$ が $s \in S_1$ を批判する場合に $R(S_x, s)$ が定義される全ての $S_x \subseteq S_2$ において $s_x \in S_x$ を同盟完全批判する集合が $\alpha(S_1)$ の下位集合としてある場合に限り、S_1 を同盟容認議論集合とする。S_1 が同盟容認議論集合であり同盟容認集合となる $S_1 \subset S_y \subseteq S$ を満たす S_y がない場合に限り、S_1 を同盟理想議論集合とする。

上記の定義において $S_2 \subseteq S$ ではなく $S_2 \subseteq \pi(1, \mathsf{View}_R(S, S_1))$ となっているが、これは S_1 が必ずしも無対立とならないからである。仮に $S_2 \subseteq S$ としてしまうと S_1 内での対立が批判とされてしまう。しかし、同盟容認性はそれが考察される対立許容議論集合 S の視点からみた S の容認性であるので、S 内の対立は考慮されるべきでない。対して、 $s \in \alpha(S_1)$ ではなく $s \in S_1$ となっている理由は、外部からの対立許容議論集合への批判はその構成員を批判するのみで十分であるからである。そのため、例3において $\{(a_1, 4), (a_3, 5)\}$ (また $\{(a_1, 4), (a_2, 3), (a_3, 5)\}$) は同盟容認議論集合ではない。

4 同盟利益意味論また同盟形成意味論

本章では前章で定義した対立許容議論集合を同盟の抽象議論理論的なモデルと捉え、同盟形成意味論を考察する。まず、同盟を組むというからには少なくとも2つ以上の議論集合があるはずであるためどの議論集合が容認されるかやどの容認される議論集合が他の容認される議論集合よりも理想であるのか、が問題とされているとは思わない。与えられた対立許容議論集合が他の独立の対立許容議論集合と同盟を組みうるか、がより的を得た問いであろうと考える。そのため、我々の意味論は対立許容議論集合に相対化されたもの、つまり対立許容議論集合をパラメーターとしてとるもの、とする。我々は初めに同盟利益意味論を考え、それから同盟利益意味論を基にする同盟形成意味論を考察する。以下の定義を仮定する。

定義 8 (一方批判) S_1 を対立許容議論集合とする。S_x が $s \in S_1$ を批判し、なおかつ S_1 がどの $s_x \in S_x$ も同盟批判しないような $S_x \subseteq \pi(1, \mathsf{View}_R(S, S_1))$ が存在する場合に限り、S_1 は一方批判を受ける、とする。

例3において、$\{(a_1, 4), (a_2, 3), (a_3, 5)\}$ は $(a_4, 1)$ より一方批判を受ける。

定義 9 (対立許容議論集合の状態) $\preceq : 2^{\mathsf{S}} \times 2^{\mathsf{S}}$ という二項関係を定め、S_1、S_2 が共に対立許容議論集合であり、また次のいずれかを満たす場合に限り $(S_1, S_2) \in \preceq$ が満たされるとする。なお、$S_1 \preceq S_2$ とも書く。

1. S_2 は同盟容認議論集合。
2. S_1 は一方批判を受ける。
3. S_1、S_2 のどちらも同盟容認議論集合でないし、一方批判も受けない。

$(S_1, S_2) \in \preceq$ であるとき、S_2 は少なくとも S_1 ほどには良い状態である、とする。

大まかには、$S_x \subseteq S$ が同盟容認議論集合であれば外部からの批判から完全に弁護されているわけであるから良い状態にある。S_x が一方批判を受けるのであれば、外部批判に対する答えがないわけであるから悪い状態にある。そのどちらでもない場合は前者よりは悪い状態であるが後者よりは良い状態となる。

定義 10 (同盟許可) $S_1 \cap S_2 = \emptyset$ また $S_1 \cup S_2$ が対立許容議論集合である場合に限り S_1 と S_2 の同盟が許可される、とする。

以下の結果により、この定義は十分である。

補題 1 $S_1 \subseteq S$ また $s \in S_1$ とする。S_1 が s を完全批判するならば, $S_1 \subseteq S_2 \subseteq S$ を満たす全ての S_2 において S_2 は s を完全批判する.

公理 5 S_1 と S_2 の同盟が許可されるならば, S_1 と S_2 は必然的に同盟容認議論集合である。

証明 1 仮にそうでないとすると、対立容認性の定義により、S_i （i は１か２）が $s \in S_i$ を完全批判するような S_i が存在する。補題 1 を適用。

次の定義を便宜的に定め、その上で同盟利益を定義する。

定義 11 $\mathit{Attacker} : 2^{\mathsf{S}} \rightarrow 2^{\mathsf{S}}$ を、$\mathit{Attacker}(S_1) = \{s \in S \mid s$ に批判される s_1 が存在する $\}$. を満たすものとする。$\mathit{Attacker}(S_1)$ を S_1 の批判者とする。

定義 12 (同盟利益) $\trianglelefteq : 2^{\mathsf{S}} \times 2^{\mathsf{S}}$ を、$S_1 \trianglelefteq S_2$ (または $(S_1, S_2) \in \trianglelefteq$) ならば次の３つの制約を全て満たすものとする。

1. $S_1 \subseteq S_2$ （より大きい集合)。
2. $S_1 \preceq S_2$ (より良い状態)。
3. $|\{s \in \mathsf{Attacker}(S_1) \mid S_1$ は s を同盟完全批判しないかつ $s \notin S_1\}| \geq |\{s \in \mathsf{Attacker}(S_1) \mid S_2$ は s を同盟完全批判しないかつ $s \notin S_2\}|$ (より少数の批判者)。

$S_1 \trianglelefteq S_2$ である場合に限り、S_2 は S_1 にとって利益のある同盟であるとする。

2 番目の制約により、$S_1 \trianglelefteq S_2$ ならば S_1 と S_2 のどちらも対立許容議論集合となる。3 番目の制約では不等号の右側の項においても $s \in \mathsf{Attacker}(S_1)$ となっているが、これは誤りではなく、この同盟利益が、S_1 に対しての利益を図っているためである。例 3 では $\{(a_1, 4), (a_3, 5)\} \trianglelefteq \{(a_1, 4), (a_2, 3), (a_3, 5)\}$ が成り立つ。

公理 6 上記の制約の一つのみを満たすように (S, R) を選ぶことができる。

4.1　同盟利益に関する結果

$S_1 \trianglelefteq S_1 \cup S_2$ となるような (S, R), $S_1, S_2 \subseteq S$ が存在することは容易にわかるので、他の結果を示す。

定理 1 S_1 を対立許容議論集合とし、S_x を同盟容認議論集合とする。$S_1 \subset S_x$ ならば、以下が満たされる。

1. $S_2 = S_x \backslash S_1$ は対立許容議論集合である。
2. S_1 と S_2 の同盟は許可される。
3. $S_1 \trianglelefteq S_x$。

定理 2 (存在定理) 各 $S_1 \subseteq S$ に関して、S_1 と同盟が許可され、$S_1 \trianglelefteq S_1 \cup S_2$ であり、なおかつ $S_1 \cup S_2$ が同盟容認議論集合となる S_2 が存在するならば、S_1 と同盟が許可され、$S_1 \trianglelefteq S_1 \cup S_3$ であり、$S_1 \cup S_3$ が同盟理想議論集合であり、$S_2 \subseteq S_3$ となる S_3 が存在する。

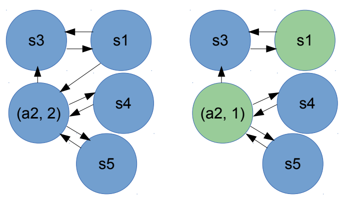
定理 3 (相互最大同盟) S_1 を対立許容議論集合とし、$\mathit{Pref}(S_1)$ を S_1 を下位集合として含む全ての同盟理想議論集合の集合とする。$\mathit{Pref}(S_1) \neq \emptyset$ ならば以下が

成り立つ。全ての $S_x \in \mathit{Pref}(S_1)$ において、$S_1 \trianglelefteq S_x$、また $(S_x \backslash S_1) \trianglelefteq S_x$、また $S_1 \trianglelefteq S_y$ や $(S_x \backslash S_1) \trianglelefteq S_y$ となる $S_x \subset S_y \subseteq S$ は存在しない。

しかし、一般的にはそのような相互利益は確約されない。

定理 4 (同盟利益の非対称性) $S_1 \trianglelefteq S_1 \cup S_2$ が満たされるが $S_2 \trianglelefteq S_1 \cup S_2$ が満たされないような S_1 と S_2 を有する (S, R) が存在する。

証明 2



S を $\{s_1, (a_2, 2), s_3, s_4, s_5\}$ とし、R は上図批判関係に対応するものとし、以下を満たすとする。
$R(\{s_1\}, (a_2, 2)) = 1$;
$R(\{s_1, (a_2, 1)\}, s_3) \geq \pi(2, s_3)$; $R(\{s_1\}, s_3) < \pi(2, s_3)$;
$R(\{(a_2, 2)\}, s_3) < \pi(2, s_3)$; $R(\{(a_2, 1)\}, s_4) < \pi(2, s_4)$;
$R(\{(a_2, 1)\}, s_5) < \pi(2, s_5)$; $R(\{(a_2, 2)\}, s_4) \geq \pi(2, s_4)$;
$R(\{(a_2, 2)\}, s_5) \geq \pi(2, s_5)$; $R(\{s_3\}, s_1) \geq 2$;
$R(\{s_4\}, (a_2, 1)) \geq 2$; and $R(\{s_5\}, (a_2, 1)) \geq 2$。

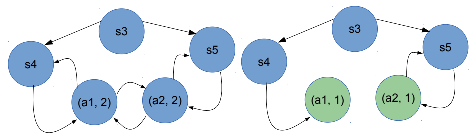
ここで、S_1 を $\{s_1\}$ とし、また S_2 を $\{(a_2, 2)\}$ とすると、S_1 と $S_1 \cup S_2$ のどちらも中間の状態にあるので $S_1 \trianglelefteq S_1 \cup S_2$ である。(より少数の批判者) も満たされる。しかし、$S_2 \trianglelefteq S_1 \cup S_2$ は、(より少数の批判者) を満たさず、成立しない。

更に、$\trianglelefteq$ は次のような一見成立しそうな特性も満たさない。

定義 13 ($\trianglelefteq$ の連続性) S_1 を対立許容議論集合とし、$\mathsf{Max}(S_1)$ を（1）$S_1 \trianglelefteq S_x$ また（2）$S_x \subset S_y \subseteq S$ ならば $S_1 \trianglelefteq S_y$ ではない、という条件を満たす全ての S_x の集合とする。ここで、各 $S_w \subseteq S_z$ において、S_1 と $S_w \backslash S_1$ の同盟が許可されるならば $S_1 \trianglelefteq S_w$ となるような $S_z \in \mathsf{Max}(S_1)$ が存在する場合に限り、$\trianglelefteq$ は弱い連続性を持つものとする。全ての $S_z \in \mathsf{Max}(S_1)$ において $\trianglelefteq$ が弱い連続性を持つとき、$\trianglelefteq$ は連続性を持つものとする。

定理 5 (利益の非連続性) *(1)* S_1, S_2 また S_x が対立許容議論集合であり、*(2)* $S_x \in \mathit{Max}(S_1)$ であり、また *(3)* $S_2 \subseteq S_x$ であるが、$S_1 \trianglelefteq S_2$ でないような S_1, S_2, S_x を含む議論フレームワークが存在する。この結果は $\mathit{Max}(S_1) = \mathit{Pref}(S_1)$ としても変わらない。

証明 3 $\mathsf{Max}(S_1) = \mathsf{Pref}(S_1)$ を考えるのみでよい。



議論フレームワーク (S, R) において、 S を $\{(a_1, 2), (a_2, 2), s_3, s_4, s_5\}$ とし, R を上図の批判関係を反映するものとしまた以下の条件を満たすものとする。
$R(\{(a_1, 2)\}, s_4) \geq \pi(2, s_4)$; $R(\{(a_2, 2)\}, s_5) \geq \pi(2, s_5)$;
$R(\{(a_1, 2)\}, (a_2, 2)) = 1$;
$R(\{(a_2, 2)\}, (a_1, 2)) = 1$; $R(\{s_3\}, s_4) \geq \pi(2, s_4)$;
$R(\{(a_2, 1), s_3\}, s_5) \geq \pi(2, s_5)$; $R(\{s_4\}, (a_1, 1)) \geq 2$;
$R(\{s_5\}, (a_2, 1)) \geq 2$。ここで、明らかに α is $S_1 = \{(a_1, 2)\}$, $S_a = \{(a_2, 2)\}$ $S_2 = S_1 \cup S_a$, $S_x = \{(a_1, 2), (a_2, 2), s_3\}$ は対立許容議論集合である。また、S_x は同盟理想議論集合である。しかし、$S_1 \trianglelefteq S_2$ は満たされない。また、S_2 は、右図のように一方批判を受け、対して S_1 と S_a は中間の状態にあるため、$S_a \trianglelefteq S_2$ でもない。

ではあるが、特別な条件下では連続性は成り立つ。

定理 6 (利益の連続性) $S_x \subseteq S$ を同盟理想議論集合とする。全ての独立の S_x の下位集合 S_y, S_z について $S_y \trianglelefteq S_y \cup S_z$ ならば $\trianglelefteq$ は全ての $S_1 \subset S_x$ において弱い連続性を持つ。

4.2 同盟形成意味論

同盟利益の関係を用いて同盟形成意味論を定める。以下のユーティリティーを仮定する。

I 同盟が少なくとも一方の議論集合にとって利益となるならばその同盟は形成されるべきである。

II 同盟が両方の議論集合にとって利益となるならばその同盟は形成されるべきである。

III 同盟から将来的に潜在的に最大限の利益が期待されるならばその同盟は形成されるべきである。

これらのうち初めの２つは $\trianglelefteq$ から即座に理解できる。３つ目の公準に関しての我々の解釈は以下の通りである。政党、つまり我々のセッティングではある対立許容議論集合 S_1、が他の対立許容議論集合 S_2 との同盟を考察する場合、S_1 が S_2 と同盟を形成する前の状態では $\mathsf{Max}(S_1)$ が S_1 にとって最大利益をもたらす同盟の集合である。であるわけだが、一旦同盟が形成されると $\mathsf{Max}(S_1 \cup S_2)$ がその同盟にとって最大利益をもたらす同盟の集合となり、ここで明らかに $\mathsf{Max}(S_1 \cup S_2) \subseteq \mathsf{Max}(S_1)$ である。これが意味することは、S_2 の選択が $\mathsf{Max}(S_1) \backslash \mathsf{Max}(S_1 \cup S_2)$ の全てを $S_1 \cup S_2$ から到達不可能にすることで、そのため S_1 は、$\mathsf{Max}(S_1 \cup S_2)$ の全ての構成員が $\mathsf{Max}(S_1)$ のある構成員よりも厳密に少利益となる場合にはそのような S_2 と同盟を形成しないというインセンティブを持つ。３つ目の公準はこのような直感を反映している。

定義 14 (最大利益関係)

$\leq_l, \leq_b, \leq_f \colon 2^{\mathcal{S}} \times 2^{\mathcal{S}}$ が以下を満たすとする。

1. $S_1 \leq_l S_2$ の場合に限り $|S_1| \leq |S_2|$.
2. $S_1 \leq_b S_2$ の場合に限り S_2 は少なくとも　S_1 と同程度に良い状態にある。
3. $S_1 \leq_f S_2$ の場合に限り S_2 は少なくとも S_1 と同じく少数の批判者、もしくはそれ以下の批判者を持つ。

各 $\beta \in \{l, b, f\}$ について、$S_1 \leq_\beta S_2$ であるが $S_2 \leq_\beta S_1$ でない場合に限り、$S_1 <_\beta S_2$ と書くことにする。この時、$\trianglelefteq_{\mathrm{m}} : 2^{\mathcal{S}} \times 2^{\mathcal{S}}$ を、$S_1 \trianglelefteq_{\mathrm{m}} S_2$ ならば以下の制約を満たすとする。

1. $S_1 \trianglelefteq S_2$。
2. 少なくともある $S_x \in \mathsf{Max}(S_2)$ において以下が満たされる。全ての $S_y \in \mathsf{Max}(S_1)$ について,
   $S_x <_\beta S_y$ となる $\beta \in \{l, b, f\}$ が存在するならば
   $S_y <_\gamma S_x$ となる $\gamma \in (\{l, b, f\} \backslash \beta)$ が存在する。

$S_1 \trianglelefteq_{\mathrm{m}} S_2$ ならば集合の大きさ、状態、もしくは批判者の数という判断基準において最大の利益を持つ $\mathsf{Max}(S_1)$ の集合のうち、少なくとも一つが $\mathsf{Max}(S_2)$ に含まれることとなる。

以下、I を満たす W、II を満たす M、I また III を満たす WS、また II また III を満たす S という４つの意味論を示す。

定義 15 (同盟形成意味論)

$$\mathsf{W}(S_1) = \{S_2 \subseteq S \mid S_1 \trianglelefteq S_1 \cup S_2 \text{ or } S_2 \trianglelefteq S_1 \cup S_2\}.$$
$$\mathsf{M}(S_1) = \{S_2 \subseteq S \mid S_1 \trianglelefteq S_1 \cup S_2 \text{ and } S_2 \trianglelefteq S_1 \cup S_2\}.$$
$$\mathsf{WS}(S_1) = \{S_2 \subseteq S \mid S_1 \trianglelefteq_{\mathrm{m}} S_1 \cup S_2 \text{ or } S_2 \trianglelefteq_{\mathrm{m}} S_1 \cup S_2\}.$$
$$\mathsf{S}(S_1) = \{S_2 \subseteq S \mid S_1 \trianglelefteq_{\mathrm{m}} S_1 \cup S_2 \text{ and } S_2 \trianglelefteq_{\mathrm{m}} S_1 \cup S_2\}.$$

直感的には、各 $\rho \in \{\mathsf{W}, \mathsf{M}, \mathsf{WSS}\}$ について $\rho(S_1)$ ならば、S_1 は $S_2 \in \rho(S_1)$ と同盟を組むことを I、II、III の条件にそって承諾することとなる。

定理 7 S_1 を対立許容議論集合とすると、以下が満たされる。*(1)* $\mathsf{M}(S_1) \subseteq \mathsf{W}(S_1)$. *(2)* $\mathsf{WS}(S_1) \subseteq \mathsf{W}(S_1)$. *(3)* $\mathsf{S}(S_1) \subseteq \mathsf{M}(S_1)$. *(4)* $\mathsf{S}(S_1) \subseteq \mathsf{WS}(S_1)$. 一方で $\mathsf{WS}(S_1) \subseteq \mathsf{M}(S_1)$ また $\mathsf{M}(S_1) \subseteq \mathsf{WS}(S_1)$ は必然ではない。

5 終わりに

本論文では抽象議論理論的な同盟利益形成意味論を提案した。この研究成果は多くの抽象議論理論の研究分野にまたがっており、一層の研究が期待される。

参 考 文 献

[1] Amgoud, L.: An argumentation-based model for reasoning about coalition structures, *Argumentation in Multi-Agent Systems*, 2005, pp. 217–228.

[2] Amgoud, L. and Ben-Naim, J.: Axiomatic Foundations of Acceptability Semantics, *KR*, 2016, pp. 2–11.

[3] Amgoud, L. and Besnard, P.: Bridging the Gap between Abstract Argumentation Systems and Logic, *SUM*, 2009, pp. 12–27.

[4] Arieli, O.: Conflict-Tolerant Semantics for Argumentation Frameworks, *JELIA*, 2012, pp. 28–40.

[5] Baumann, R. and Brewka, G.: AGM Meets Abstract Argumentation: Expansion and Revision for Dung Frameworks, *IJCAI*, 2015, pp. 2734–2740.

[6] Boella, G., van der Torre, L., and Villata, S.: Social Viewpoints for Arguing about Coalitions, *PRIMA*, 2008, pp. 66–77.

[7] Caminada, M. and Amgoud, L.: An Axiomatic Account of Formal Argumentation, *AAAI*, 2005, pp. 608–613.

[8] Cayrol, C., de Saint-Cyr, F. D., and Lagasquie-Schiex, M.-C.: Change in Abstract Argumentation Frameworks: Adding an Argument, *Journal of Artificial Intelligence Research*, Vol. 38(2010), pp. 49–84.

[9] Coste-Marquis, S. and Konieczny, S.: A Translation-Based Approach for Revision of Argumentation Frameworks, *JELIA*, 2014, pp. 397–411.

[10] Coste-Marquis, S., Konieczny, S., Mailly, J.-G., and Marquis, P.: On the Revision of Argumentation Systems: Minimal Change of arguments Statuses, *KR*, 2014.

[11] 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(1995), pp. 321–357.

[12] Dunne, P. E., Hunter, A., McBurney, P., Parsons, S., and Wooldridge, M.: Weighted argument systems: Basic definitions, algorithms, and complexity results, *Artificial Intelligence*, Vol. 175, No. 2(2011), pp. 457–486.

[13] Gabbay, D. M. and Rodrigues, O.: An equational approach to the merging of argumentation networks, *Journal of Logic and Computation*, Vol. 24, No. 6(2014), pp. 1253–1277.

[14] Leite, J. and Martins, J.: Social Abstract Argumentation, *IJCAI*, 2011, pp. 2287–2292.

[15] Nielsen, S. H. and Parsons, S.: A generalization of Dung's Abstract Framework for Argumentation, *Argumentation in Multi-Agent Systems*, 2006, pp. 54–73.

[16] Pereira, L. M. and Pinto, A. M.: Approved Models for Normal Logic Programs, *LPAR*, 2007, pp. 454–468.

[17] Riley, L., Atkinson, K., and Payne, T.: A Dialogue Game for Coalition Structure Generation with Self-Interested Agents, *COMMA*, 2012, pp. 229–236.

[18] Rotstein, N., Moguillansky, M. O., García, A. J., and Simari, G. R.: An abstract Argumentation Framework for Handling Dynamics, *Proceedings of the Argument, Dialogue and Decision Workshop in NMR 2008*, 2008, pp. 131–139.