

# 議論に基づくパレート最適性

\* 木藤浩之<sup>(1)</sup>, 新田克己<sup>(1)</sup>

(1) 東京工業大学大学院総合理工学研究科

〒226-8502 神奈川県横浜市緑区長津田町 4259 J2-53

E-mail : {kido,nitta}@ntt.dis.titech.ac.jp

**Abstract:** 本稿は、異なる願望を持つ主体が論理を駆使して行為を決定する状況で、集団の合理的選択がいかに対話的に達成され得るかを形式的側面から明らかにすることを旨とする。この目的のため、パレート最適性を議論の概念を用いて定式化し、対話的かつ手続き的にパレート最適性を評価する方法を提案する。

## 1. はじめに

我々は日々、対話や議論を行うことで現実の色々な問題を解決あるいは問題に対処している。我々の対話や議論を観察するとそれは多様な性質を持つことに気づく。(i)知識や信念などの相違によって対立や矛盾が生じるという性質、(ii)個々の事例の一般化による誇張や自身の目的からその手段を考察するなど、色々な種類の考え方が現れるという性質、(iii)ある意見はより説得力のある意見によって覆るという性質などである。(iii)を扱う推論は特に非単調推論と呼ばれる。議論 (Argumentation) は非単調推論の実現手法として開発されたという背景を持つが、他の多くの非単調推論の実現手法とは異なり、これら3つの性質すべての扱いを許容する推論の仕組みを与えるという特長を持つ。これに起因して、これまで議論に基づく説得、交渉、熟考の形式化が数多くなされてきた。ここで熟考とはエージェントの集団が選択肢の中からどの行動方針をとるかを決定するために行う対話とされる。他方、近年、利己的なエージェント間の戦略的相互作用を対象としたゲーム理論と議論の関係も深まりつつある。事実、自動交渉に対する古典的ゲーム理論に基づく手法の限界を議論に基づく手法が克服するという指摘や、ゲーム理論の知見を議論に導入する試みが既に始まっている。

他方、議論に基づいて単一エージェントの熟考を形式化する試みはいくつか存在するものの、それらは多数のエージェントが熟考を行うときに何が集団全体にとって合理的な行為であるか、そしてそれはいかに対話的に推論されるかということは対象とはしていない。それを扱うには集団全体にとっての合理性を与えるパレート最適性の概念が不可欠である。本稿では、各自異なる願望を持つエージェントが自身の知識を用いて熟考する状況に焦点を当て、集団にとっての合理的な行為の決定がいかにして対話的に形式化されうるかを明らかにすることを目的とする。

**例 (動機となる例)** エージェント  $i, j$  が二人の貯金を使ってマンションの購入を計画している状況を考える。 $i$  は治安の良さ、交通の便の良さ、日当たりの良さを望んでおり、 $j$  は治安の良さと閑静さを望んでいる。図 1の左側は購入の行為を、右側は実現される願望を表す。行為とそれによって実現される願望間には矢印が引かれている。今、マンション  $d$  の購入の是非について二人は議論する。

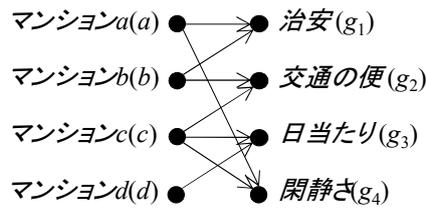


図 1 行為と実現される願望の関係

- j: 私にとってはマンション d よりマンション c の方がいい。だってマンション c は閑静さという私の願望を満たすけど、マンション d は私の願望を何も満たさないから。
- i: そうだね。私にとってもマンション c の方がいい。だってマンション c は私の願望のうち日当たりのよさと交通の便の良さを満たすけど、マンション d は日当たりの良さしか満たさないから。この場合、エージェント i, j の両者にとってマンション d よりもマンション c が好ましいことが対話から判明している。ゆえに合理的なエージェント i, j は集団の選択としてマンション d を候補に入れるべきでない。次はマンション a の購入についての議論である。
- j: 私にとってはマンション a よりマンション b の方がいい。だってマンション b は私の願望のうち治安と交通の便の良さを満たすけど、マンション a は治安の良さしか満たさないから。
- i: いや、私にとってはマンション a の方がいい。だってマンション a は私の願望のうち治安の良さと閑静さを満たすけど、マンション b は治安の良さしか満たさないから。
- i: ところで、私にとってはマンション c もいい。だってマンション c は私の願望のうち交通の便と日当たりの良さを満たすけど、マンション a はそれを満たさないから。
- j: いや、私にとってはマンション a の方がいい。だってマンション a は私の願望のうち治安の良さと閑静さを満たすけど、マンション c は閑静さしか満たさないから。
- i: ところで、私にとってはマンション d もいい。だってマンション d は私の願望のうち日当たりのよさを満たすけど、マンション a はそれを満たさないから。
- j: いや、私にとってはマンション a がいい。だってマンション a は私の願望のうち治安の良さと閑静さを満たすけど、マンション d は私の願望を何も満たさないから。

この場合、エージェント i, j の両方において b よりも好ましいマンションも、また一方のエージェントにおいて b より好ましく、他方のエージェントにおいて b と比較できないようなマンションも対話から判明していない。ゆえに合理的なエージェント i, j は集団の選択としてマンション a を候補に入れるべきである。これらの例は、マンション d は二人にとってよりよいマンション c が存在するという意味でパレート最適ではなく、一方マンション a は二人にとってパレート最適であることを非形式的に意味する。本稿の目的はここで見たパレート最適性の判別のため対話を形式化することである。

## 2. 議論に基づくパレート最適性

### 2.1. 議論フレームワーク

議論フレームワークの記述には、論証を記述する言語、論証、論証間の攻撃関係を定義する必要がある。各定義の要点を具体的事例に沿って概説する。

**例 1 (論理言語)** エージェント i, j が共有する知識 K, 及びエージェントの願望の集合  $G_i, G_j$  を次式で与えられるものとする。

- $K = \{a \Rightarrow g_1, a \Rightarrow g_4, b \Rightarrow g_1, b \Rightarrow g_2, c \Rightarrow g_2, c \Rightarrow g_3, c \Rightarrow g_4, d \Rightarrow g_3\}$

- $G_i = \{Dg_1, Dg_2, Dg_3\}$ ,  $G_j = \{Dg_1, Dg_4\}$

a, b, c, d,  $g_l (1 \leq l \leq 4)$  は命題言語のリテラルであり, 非形式的にエージェントによって真偽を制御することができるものとする. 「action  $\Rightarrow$  goal」の形式を持つ規則をデフォルトと呼び, デフォルトは action が実現されるならば典型的に goal が実現されることを意味する.  $D$  は願望を表す様相演算子であり  $Dg_2 \in G_i$  は, エージェント  $i$  は  $g_2$  が実現されることを望んでいることを意味する.

Walton によると熟考では, 所与の願望からそれを実現する行為を望ましい行為として, それを否定する行為を望ましくない行為として推論する実践的推論が主に用いられる. 本稿では, 「実現されると証明されない願望は否定される」という閉世界仮説の立場に問題を限定し, 前者のみを扱うことにし, その一形式化手法である既知の肯定的実践的三段論法 (Positive Practical Syllogism) を用いる. また, 同一の撤回可能 (defeasible) な結論を一つにまとめる累積的推論規則 (ACCrual inference rule) を用いる. ACC の結論の添字は前提の集合である.

例 2 (実践的論証) 例 1 の行為と願望の関係は次の論証として記述され得る.

$$\begin{array}{l}
 \text{A: } \frac{\frac{Dg_1, a \Rightarrow g^1 (PPS)}{Da^{i1}} \quad \frac{Dg_4, a \Rightarrow g^4 (PPS)}{Da^{i2}}}{Da} (ACC) \qquad \text{B: } \frac{\frac{Dg_1, b \Rightarrow g^1 (PPS)}{Db^{i3}} \quad \frac{Dg_2, b \Rightarrow g^2 (PPS)}{Db^{i4}}}{Db} (ACC) \\
 \\
 \text{C: } \frac{\frac{Dg_2, c \Rightarrow g^2 (PPS)}{Dc^{i5}} \quad \frac{Dg_3, c \Rightarrow g^3 (PPS)}{Dc^{i6}} \quad \frac{Dg_4, c \Rightarrow g^4 (PPS)}{Dc^{i7}}}{Dc} (ACC) \qquad \text{D: } \frac{Dg_3, d \Rightarrow g^3 (PPS)}{Dd^{i8}} (ACC)
 \end{array}$$

図 2 実践的論証 A,B,C,D

熟考はその目的上, 代替案の存在に起因して対立が生じる. 本稿では, 一方の論証 A で支持される行為 a によって実現される i の願望が, 他方の論証 B で支持される行為 b によって実現できないとき, i の立場から見ると b は a の代替案として不十分であるという意味で, A は B を打破する. 形式的には,  $\text{conc}(A) \neq \text{conc}(B)$  (A と B の結論が異なる) かつ  $\text{prem}(A) \cap G_i$  (A の前提にある i の願望集合) が  $\text{prem}(B) \cap G_i$  の部分集合でないとき, i に関して A は B を打破する. また, i に関して A が B を打破し, i に関して B が A を打破しないとき, i に関して A は B を完全に打破するという.

例 3 (打破関係) 例 2 における,  $i, j$  に関する打破関係  $\text{defeat}_i, \text{defeat}_j$  を示す.

- $\text{defeat}_i = \{(A, C), (A, D), (B, A), (B, C), (B, D), (C, A), (C, B), (C, D), (D, A), (D, B)\}$
- $\text{defeat}_j = \{(A, B), (A, C), (A, D), (B, C), (B, D), (C, B), (C, D), (D, B)\}$

論証と打破関係の定義に基づいて議論フレームワークを与える.  $\text{Args}$  を実践的論証の集合,  $\text{defeat}_i$  を  $i$  に関する打破関係とする. 議論フレームワーク  $AF$  を  $AF = \langle \text{Args}, \text{defeat}_1, \dots, \text{defeat}_n \rangle$  と定義する.

例 4 (議論フレームワーク) 例 2, 例 3 で与えたゴールと行為の関係に関して, 議論フレームワーク  $AF$  は  $AF = \langle \{A, B, C, D\}, \text{defeat}_i, \text{defeat}_j \rangle$  となる.

## 2.2. パレート最適性の議論意味論

議論意味論は, 議論の論理的帰結を意味論の側面から与える. 既存の熟考や実践的推論の議論に基づく形式化手法が Dung 意味論に基づく一方で, 本稿が対象とするパレート最適性は Dung の議論意味論では的確に捉えられない. そこで議論で扱われる概念を用いて論証を単位としてパレート最適性を定義する.  $\text{Args}$  を実践的論証集合とする. 次の条件を満たすとき,  $A \in \text{Args}$  はパレート最適である.

- $\forall B \in \text{Args} (\exists i (i \text{ に関して } B \text{ は } A \text{ を打破する}) \rightarrow \exists j (j \text{ に関して } A \text{ は } B \text{ を完全打破する}))$

**例 5 (パレート最適性)** 例 4の AF に関して、図 3は論証を節に打破関係を有向の辺として記述したものである。矢印の元から先へ矢印上に記述されたエージェントに関する打破が行われる。図 3の左上、右上、左下、右下の順に A, B, C, D それぞれと他の論証間で存在する打破関係を存在するものだけ記述したものである。上式より、A はパレート最適であり、B, C, D はパレート最適でない。

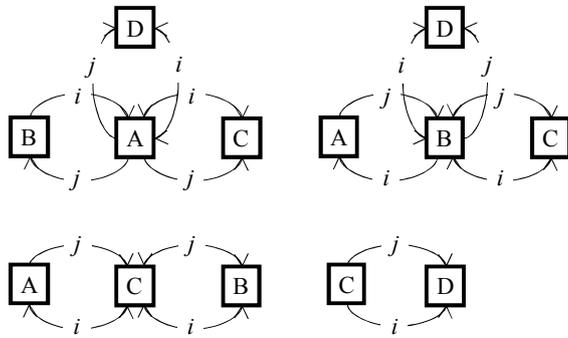


図 3 各論証に存在する打破関係

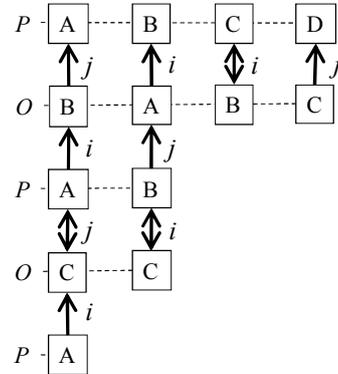


図 4 パレート最適性の対話的証明

### 2.3. パレート最適性の対話的証明

対話的証明論は、議論の論理的帰結の評価手続きを与える。証明は抽象的存在である提案者 P (Proponent)と反対者 O (Opponent)が交互に直前の論証を打破し合うことで進められる。ただし提案者、反対者はそれぞれ求める論理的結論に応じて異なるルールに従わなければならない。このルールの存在によって、どちらが対話の最後の発言者か（すなわち、どちらが対話の勝者か）ということだけに基いて議論の論理的帰結が評価される。パレート最適性の対話的証明を目指す本稿において提案者が従う主要なルールは、反対者の直前の論証を完全打破する論証を提示すること、最初に提示した論証のみを用いることである。一方、反対者が従う主要なルールは提案者の直前の論証を打破する論証を提示すること、同一論証の二度出しはしないことである。今後、論証 A が AF に関してパレート最適であるとき、かつそのときに限り A で始まる対話に提案者は勝利するような、ルールの記述を行う計画である。

**例 6 (パレート最適性の対話的証明)** 図 4は各論証のパレート最適性の対話的証明を示したものである。A を提案者の最初の論証とする対話（すなわち左端の一系列）は提案者によって対話が終了しており、提案者は勝利している。一方、B, C, D のいずれの対話も反対者によって対話が終了しており提案者は敗北している。これより A はパレート最適であることが証明され、B, C, D はパレート最適でないことが証明される。

### 3. おわりに

本稿では、パレート最適性を議論の概念を用いて定式化し、対話的かつ手続き的にパレート最適を評価する方法を提案した。本手法は、各自異なる願望を持つエージェント間の熟考において、集団にとって合理的な意思決定を行うための基礎となるものである。やるべきことは多く残されているが、著者が最も有望であると考えるのは、理論的 (theoretical) 理論との結合である。古典的ゲーム理論の限界としてエージェントの持つ効用が不変であることが指摘されている。例えば、サンルーフ付きの車の購入を望むエージェントがサンルーフに雨漏りの危険があることを知れば、その効用は変化する。古典的ゲーム理論ではこの現象を扱うことができない。これに対して、本手法に多くの議論に基づく形式化が採用している論理的矛盾に基づく打破を加えることで、先の例を扱うことが可能となる。