

F-006

## 範囲付き真理値と信頼度の概念を用いた議論フレームワーク An Argumentation Framework With Truth Range and Certainty Factor

加藤 健<sup>†</sup>      鈴木 輝彦<sup>‡</sup>      太原 育夫<sup>‡</sup>  
Takeru Kato   Teruhiko Suzuki   Ikuo Tahara

### 1 はじめに

議論とは日常において複数の話者により、ある議題に関する結論を導出したり、合意を形成したりするためのものである。議論に関する研究は、対立した状況の解消や合意形成を行うこと、社会的に通用するような結論を導出することを目的としている [1]。

梅田ら [2] は合意形成を行うことを目的とし、完全否定 (… ではない) と部分否定 (… とは限らない) の二種類の否定を取り扱うことで、根拠のない不確実な情報を取り扱うことを可能とした。また、確実なことを表す絶対規則と個人の主張・意見などを表す撤回可能規則の二つの規則により、表現の幅を広げている。このように多様な表現を可能にした上で、複数エージェントによる議論を通じて合意を形成し、合意した意見に対し補強を行いより良い結論を導出している。しかし、このフレームワークのもとでは料金交渉などの定量的な問題などに対応するのは困難だとされている。また、4 値論理により是非を一義に定められない問題の表現を行い、さらに導出した結論が社会的に通用するものとするために社会通念の導入を行った方法もある [3]。

本稿では、主観的で曖昧な情報を多く含んだ議論を数理的に表現する方法を検討する。

### 2 知識の表現

#### 2.1 事象と知識

命題定数で表されるものを事象とよび、含意形式  $\varphi \leftarrow \psi$  ( $\varphi$  を結論部、 $\psi$  を前提部とよび、 $\varphi$  はリテラルの選言、 $\psi$  はリテラルの連言である) を知識とよぶことにする。

たとえば、 $p, q, r$  を事象としたとき

$$p \leftarrow \neg q \wedge r$$

は知識である。

#### 2.2 事象の真偽値

事象の主観的な真偽性を 0(偽)~1(真) における実数値の範囲で表し、これを真偽値とよぶことにする。たとえば

$$p(\text{真偽値} : 0.5 \sim 0.8)$$

<sup>†</sup> 東京理科大学大学院理工学研究科情報科学専攻

<sup>‡</sup> 東京理科大学理工学部情報科学科

と表す。このとき、その否定  $\neg p$  は

$$\neg p(\text{真偽値} : 0.2 \sim 0.5)$$

となる。

### 2.3 信頼度

知識は必ずしも正しいとは言えないため、その正しさを表す指標として信頼度を導入する。信頼度は 0~1 の実数値で表され、1 に近づくほど正しいと言える。

たとえば、「彼は泣いている」ならば「彼は悲しい」という知識があったとして、そこに信頼度の概念を導入し信頼度を 0.7 と与えることによって、そうでない場合についても考慮することが可能となる。

前提の存在しない知識 (以下、根拠と呼ぶ) の信頼度は他の知識とは異なり任意には定められず、その根拠の結論部の真偽値から以下の式に従い導出されるものとする。

すなわち、根拠の結論部の真偽値の最大値を  $ctfmax$ 、その範囲を  $ctfrange$ 、その知識集合におけるその時の各知識のもつ真偽値の範囲の平均値を  $average$  としたとき、

$$ctfmax \times (1 - |ctfrange - average|) \quad (1)$$

とする。

ここで、知識集合における各知識の真偽値の範囲の平均値を用いているのは、真偽値の範囲は過剰に広くても狭くてもよいものではなく、また 0.5 などのような具体的に全てにおいて適用できる妥当な広さを定義できるものでもないためである。これにより、適切な範囲の大きさを定義し、その大きさから過剰に離れたものの評価を下げることができる。

たとえば、 $p(\text{真偽値} : 0.5 \sim 0.8)$  であり、 $average$  が 0.2 だとすると、 $p \leftarrow$  の信頼度は

$$0.8 \times (1 - |0.3 - 0.2|) = 7.2$$

となる。

### 3 真偽値の一貫性

#### 3.1 知識集合と真偽値集合

各知識集合は、知識の集合である知識集合に対して、その知識集合に含まれる各事象の真偽値を集めた真偽値集合をもつ。これは、その知識集合の中で各事象の真偽値が共通であるためである。

### 3.2 伝播による真偽値の一貫性の向上

含意形式で表現される知識においては、前提部の真偽値が結論部の真偽値と一貫性をもつ真偽値であるべきである。

そこで、前提部の真偽値と信頼度を用いて、結論部に与えられるであろう真偽値を計算し結論部の値を修正することにする。これを本稿では伝播として定義する。

まず、前提部がもつ真偽値を以下のように定める。前提部が単一のリテラルからなる場合、そのリテラルの真偽値をそのまま前提部の真偽値とする。前提部がリテラルの連言である場合、各リテラルについて(1)式を用いて評価しその値が最小のリテラルの真偽値を前提部の真偽値とする。

前提部がもつ真偽値の最小値を  $atfmin$ 、最大値を  $atfmax$ 、知識の持つ信頼度を  $trustv$ 、結論部の各リテラルの真偽値の最小値を  $ctfmin$ 、最大値を  $ctfmax$  としたとき、その結論部のリテラルが複数の結論部に存在するときその重複度を  $n$  とし、新しい真偽値の最小値  $nctfmin$  を

$$nctfmin = \frac{ctfmin \times n + affmin \times trustv}{n + 1} \quad (2)$$

とし、最大値  $nctfmax$  を

$$nctfmax = \frac{ctfmax \times n + affmax \times trustv}{n + 1} \quad (3)$$

とする。これにより、その知識集合の中での全ての伝播を考えた真偽値を付与することが可能となる。ただし、 $ctfmin$ 、 $ctfmax$  が存在しなかった場合は0として扱う。

なお、結論部が選言により表現されている場合、その選言に含まれる各リテラルに対して同様に行う。

## 4 議論

### 4.1 議論ルール

議論は、提出された主張に対して肯定的な立場をとる知識集合と否定的な立場をとる知識集合の二つにより構成される。

知識を  $A = \varphi_A \leftarrow \psi_A$ 、 $B = \varphi_B \leftarrow \psi_B$  とし  $\mu_A$  を  $\psi_A$  に含まれるリテラルとしたとき、知識の提出の仕方として以下のように攻撃と補強の二種類を考え、その条件をみたすことにより提出が可能であるとする。

#### 1. 攻撃

$\mu_A = \neg\varphi_B$ 、であるか  $\psi_A = \phi$  かつ  $\varphi_A = \neg\varphi_B$  であるとき  $B$  は  $A$  を攻撃すると言う。

#### 2. 補強

$\psi_A \neq \phi$  かつ  $\varphi_B = \mu_A$  であるとき  $B$  は  $A$  を補強すると言う。

たとえば、以下のような場合、知識  $B$  は知識  $A$  を攻撃している。

$$\begin{array}{l} A = p \leftarrow q \wedge \neg r \\ \uparrow \\ B = r \leftarrow s \end{array}$$

$$\begin{array}{l} A = p \leftarrow \\ \uparrow \\ B = \neg p \leftarrow r \end{array}$$

また、以下のような場合、知識  $B$  は知識  $A$  を補強している。

$$\begin{array}{l} A = p \leftarrow r \wedge q \\ \uparrow \\ B = q \leftarrow s \end{array}$$

攻撃においては知識  $A$  と知識  $B$  は異なる知識集合に存在する知識であり、補強においては知識  $A$  と知識  $B$  は同一の知識集合に属する知識でなければならない。また、攻撃は知識の前提部もしくは前提が知識のない知識の結論部に対してのみ行うことが可能である。

攻撃と補強はその議論において、その攻撃またはその補強の知識を提出する知識集合ではない知識集合が提出した知識により前提部のリテラルが全て攻撃を受ける場合には、提出を行うことができない。前提部が存在しない知識の場合には、結論部のリテラルが攻撃を受ける場合、提出をすることができない。

補強は特に知識の提出を促し、議論を進めるために用いられる。

### 4.2 攻撃による信頼度の変化

攻撃により攻撃された知識の信頼度は低下する。すなわち、攻撃された知識の信頼度を  $atedknt$ 、攻撃した知識の個数を  $i$ 、 $n$  個目の攻撃知識の信頼度を  $atknt_n$  としたとき、攻撃された知識の信頼度を

$$atedknt \times \prod_{i=1}^n (1 - atknt_i) \quad (4)$$

に更新する。

### 4.3 重みの計算

議論において提出された、各知識について真偽値と信頼度から知識の重みを計算し、それに基づいて議論の結論を導出する。重みとはその知識がその知識集合の意見を正当化するのにどれだけ寄与しているかを表す値である。

まず、知識に含まれるリテラルの重さを計算する。リテラルの重さは、その真偽値の最大値を  $eventmax$ 、

範囲の大きさを  $eventrange$ , その知識が属する知識集合における真偽値の範囲の平均値を  $average$  としたとき,

$$eventmax \times (1 - |eventrange - average|) \quad (5)$$

とする. そして, 前提部と結論部の重みを以下のように計算する. リテラル  $p, q$  の重みをそれぞれ  $V_p, V_q$  としたとき,

$$V_{p \vee q} = \min(V_p, V_q) \quad (6)$$

$$V_{p \wedge q} = \max(V_p, V_q) \quad (7)$$

と定める. このようにして前提部と結論部の重みが得られることになる.

決定された前提部の重みを  $V_{Assump}$ , 結論部の重みを  $V_{Conc}$  とすると, その知識が持つ重みは知識の信頼度を  $trust$  として,

$$(V_{Assump} + V_{Conc}) \times trust \quad (8)$$

となる.

#### 4.4 議論の結論導出

提出された知識から各知識の重みを計算し, 提出された知識集合ごとに和を取り, その合計値の高いものが結論として正しいということになる.

肯定側の重みの合計を  $V_{Aff}$ , 否定側の重みの合計を  $V_{Den}$  とし,  $V_{Aff}$  と  $V_{Den}$  のうち値の大きい方を  $V_{Truth}$  とすると, 正しい結論は

$$\frac{V_{Truth}}{V_{Aff} + V_{Den}} \quad (9)$$

だけ正しいとして結論が出される.

### 5 伝播と議論の例

#### 5.1 例題 1

以下の2つの知識集合とその真偽値を用いて, 「今日の食事をどうするか」という議論を行う.

##### 肯定側知識集合

作るのが面倒 ← (信頼度:なし)  
 ご飯を外食にする ← 作るのが面倒 (信頼度:0.9)

##### 肯定側真偽値集合

作るのが面倒 (真偽値: 0.8~0.9)  
 ご飯を外食にする (真偽値:なし)

##### 否定側知識集合

¬お金がある ← (信頼度:なし)  
 ¬ご飯を外食にする  
 ← ¬お金がある (信頼度:0.8)  
 作るのが楽しい ← (信頼度:なし)  
 ¬作るのが面倒 ← 作るのが楽しい (信頼度:0.7)

##### 否定側真偽値集合

お金がある (真偽値: 0.0~0.2)  
 ご飯を外食にする (真偽値:なし)  
 作るのが楽しい (真偽値: 0.65~0.8)  
 作るのが面倒 (真偽値:なし)

この真偽値集合に対し伝播と信頼度の確定を行うと, 次のように根拠の信頼度と真偽値が定まる.

##### 肯定側知識集合

作るのが面倒 ← (信頼度:0.8955)  
 ご飯を外食にする ← 作るのが面倒 (信頼度:0.9)

##### 肯定側真偽値集合

作るのが面倒 (真偽値: 0.8~0.9)  
 ご飯を外食にする (真偽値: 0.72~0.81)

##### 否定側知識集合

¬お金がある ← (信頼度:0.95375)  
 ¬ご飯を外食にする  
 ← ¬お金がある (信頼度:0.8)  
 作るのが楽しい ← (信頼度:0.797)  
 ¬作るのが面倒 ← 作るのが楽しい (信頼度:0.7)

##### 否定側真偽値集合

お金がある (真偽値: 0.0~0.2)  
 ご飯を外食にする (真偽値: 0.2~0.36)  
 作るのが楽しい (真偽値: 0.65~0.8)  
 作るのが面倒 (真偽値: 0.44~0.545)

この2つの知識集合に議題「ご飯を外食にする ←」を与えて議論を行ったときの結果を図1に示す. なお, これ以降, 議論における知識の表記を

提出知識集合:攻撃か補強か:≪結論部 ← 前提部:信頼度≫ とし, 議題に関しては提出先知識集合と攻撃か補強かの項を省略する.

さて, 図1において「肯定側:補強:≪ご飯を外食にする ← 作るのが面倒:0.9≫」の知識が「否定側:反論:≪¬作るのが面倒 ← 作るのが楽しい:0.7≫」により攻撃を受けているため, 「肯定側:補強:≪ご飯を外食にする ← 作るのが面倒:0.9≫」の信頼度が変

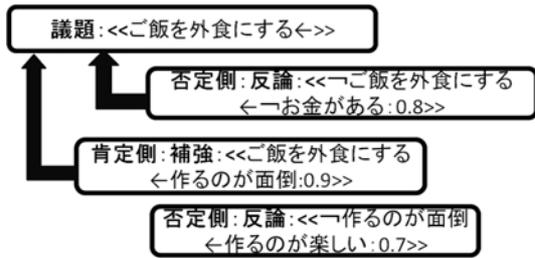


図 1: 例題 1 における議論

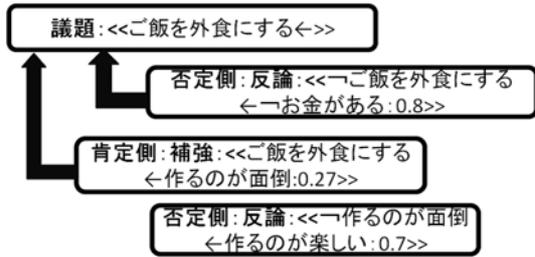
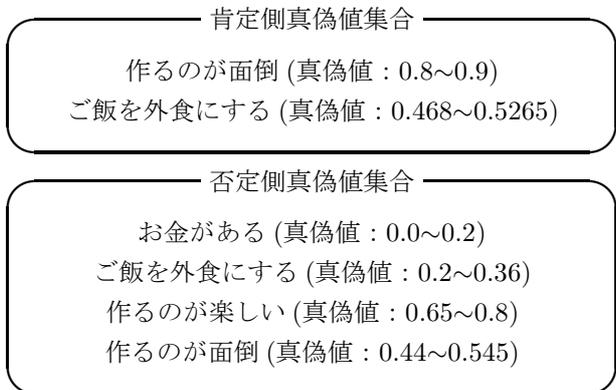


図 2: 攻撃による信頼度の変化

化し, 図 2 のようになる.

この議論結果を受け, 再度伝播が行われる. その伝播により真偽値集合は次のように変化する. ここでは, 信頼度の変化によって真偽値が変化した肯定側知識集合のみを記述する.



これにより, 肯定側, 否定側により提出された知識の重みはそれぞれ次のようになる.

- 肯定側:補強:◀ ご飯を外食にする ← 作るのが面倒:0.9>>:0.435
- 否定側:反論:◀ ◯ ご飯を外食にする ← ◯ お金がある:0.8>>:1.314
- 否定側:反論:◀ ◯ 作るのが面倒 ← 作るのが楽しい:0.7>>:0.917

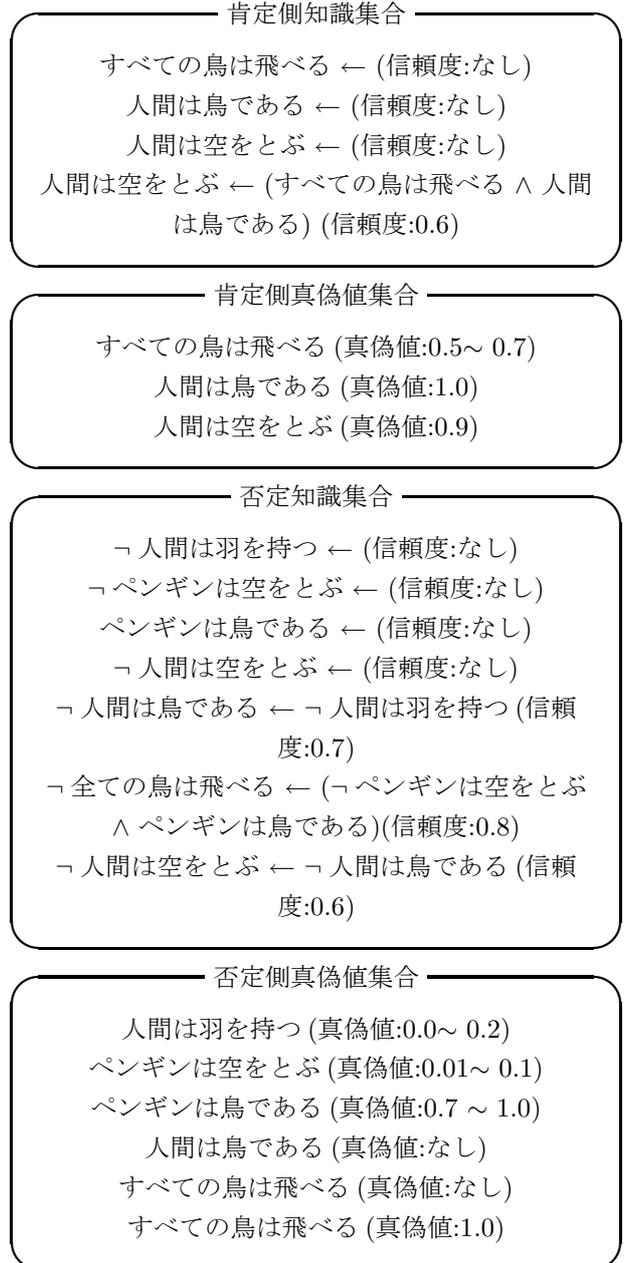
このことから, 肯定側の提出知識による重みの合計

は 0.435, 否定側の提出知識による重みの合計は 2.231 となり, 重み合計の比較により否定側の意見が 8 割 3 分ほど正しいという結果になる.

5.2 例題 2

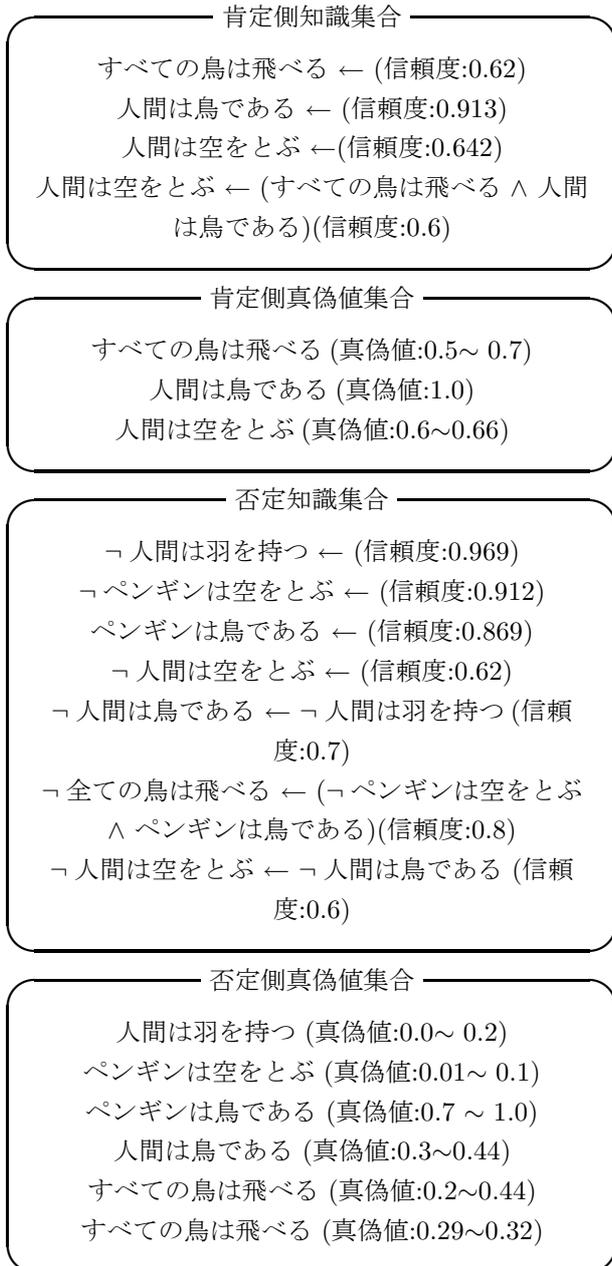
前節では問題なく回答を導出できた例について紹介した. 次に明らかな誤りを含んだ知識集合との議論を行った場合について示す.

以下の知識集合により, 「人間は飛ぶのか」という議題について議論を行う.



この知識集合それぞれについて伝播を行う. 例題 1 とは異なり, 結論部に関しても真偽値が定められているものがある. これに関しても伝播が発生し, 前提部の真偽値との一貫性が高まる.

伝播の結果は次のようになる.



この2つの知識集合に議題「人間は空をとぶ ←」を与えて議論を行ったときの結果を図3に示す。

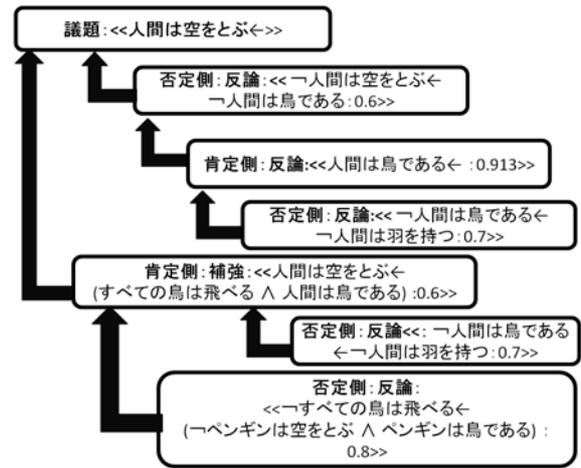


図3: 例題2における議論

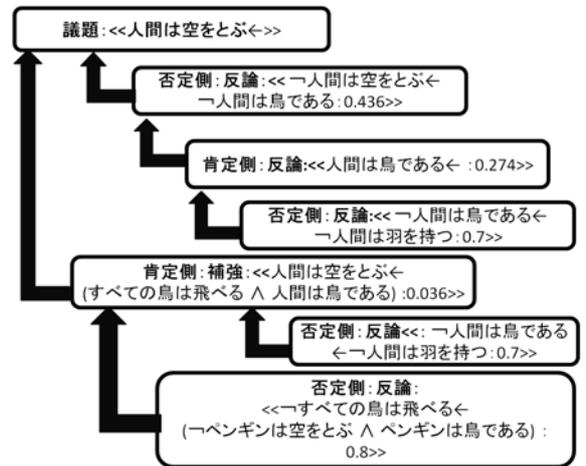


図4: 攻撃による信頼度の変化

いくつかの知識が反論を受けているため、それにに応じて信頼度が変化する。その結果を図4に示す。これにより、肯定側、否定側により提出された知識の重みはそれぞれ次のようになる。

- 肯定側:反論: <<人間は鳥である< :0.274>>:0.228
- 肯定側:補強: <<人間は空をとぶ< (すべての鳥は飛べる ∧ 人間は鳥である):0.036>>:0.046
- 否定側:反論: <<¬人間は空をとぶ< ¬人間は鳥である:0.436>>:0.584
- 否定側:反論: <<¬人間は鳥である< ¬人間は羽を持つ :0.7>>:1.085
- 否定側:反論: <<¬全ての鳥は飛べる< (¬ペンギンは空をとぶ ∧ ペンギンは鳥である):0.8>>:1.17

このことから、肯定側の提出知識による重みの合計は0.273、否定側の提出知識による重みの合計は3.924となり、重み合計の比較により否定側の意見が9割3分ほど正しいということとなる。

### 5.3 考察

#### 5.3.1 例題1に関する考察

この議論フレームワークでは、例題1のような議論では結論として完全に否定側が勝つわけではなく、肯定側の意見、この場合で言えば「作るのが面倒だから、ご飯を外食にする」という意見が完全に反対されずに結論に影響を与えていると考えられる。すなわち、結論を「ご飯を外食にしない」ではなく、「ご飯を外食にしないが、一部の食事を出来合いにする」などという結論を出すことも可能ではないかと考えられる。

#### 5.3.2 例題2に関する考察

例題2の場合、明らかに誤っていると判断できる知識が存在している。肯定側知識集合はそれに立脚して成立した知識集合であり、否定側知識集合はその全てに対し反論を行っている。

この2つの知識集合を2値の真理値で議論を行った場合、結論は「人は飛ばない」と導出される。しかし、本稿における真偽値を使用した場合、誤った知識を完全に否定できず、幾つかの意見が正しい部分もあると受け止めることができる結果となっている。

これは、否定側知識集合における信頼度が1.0でない、すなわち必ずしも正しいものではないとされている知識が提出されたために起きた現象である。したがって、否定側知識集合が全て信頼度が1.0であるような知識によって構成されていれば肯定側の意見を完全に否定することができる。

しかし、日常においては必ず正しいと言えるような知識は数少ないと考えられるため、日常における議論表現を行うという観点から考えると信頼度が極めて高い知識集合が存在することは議論の目的として相応しくないとも言える。

## 6 おわりに

本稿では、リテラルの真理性を実数値の範囲を持つ真偽値として定義し、含意式の信頼性を信頼度という概念で表現することにより、曖昧な事柄についての表現を可能とし、そうした知識表現に基づく議論フレームワークを提案した。

この方法により、断定的ではないある程度の幅をもった結論を得ることが可能となったが、反面明らかに誤った知識に対して断定的に反論することは難しくなった。

しかし、一度議論を行うことで完全ではないにしろ

誤った知識の信頼度の低下に成功している。これを繰り返し、様々な議題に対し誤りから立脚した知識集合と誤りのない知識集合を議論させることで最終的に誤った知識の信頼度を0にし、知識の誤りを修正することが可能ではないかと考えられる。

今後多くの議論例を用いて本方法の評価をしていきたい。

### 参考文献

- [1] P.Besnard and A.Hunter, "Elements of Argumentation," The MIT Press(2008)
- [2] 梅田勇一, 沢村一, "議論を計算とコミュニケーションの基本メカニズムとするエージェントシステム," 情報処理学会論文誌, vol.43, No.5, pp.1518-1527(2002)
- [3] 梅田勇一, 高橋武久, 沢村一, "準無矛盾論理に基づく議論フレームワーク," 人工知能学会論文誌, vol.19, No.2B, pp.83-94(2004)