

## ディベート型人工知能のための言語的証拠性推論問題の定式化

## Textual Evidence Inference for Debating AI

柳井 孝介\*<sup>1</sup>  
Kohsuke Yanai佐藤 美沙\*<sup>1</sup>  
Misa Sato柳瀬 利彦\*<sup>1</sup>  
Toshihiko Yanase是枝 祐太\*<sup>1</sup>  
Yuta Koreeda\*<sup>1</sup>日立製作所 研究開発グループ  
Research & Development Group, Hitachi Ltd.

This paper defines textual evidence inference problem, a new NLP task for decision-making application. The objective of the textual evidence inference problem is to classify a given text into evidence categories based on a given decision-making question and a given set of sentences representing values and preferences of decision-makers. We report the technical difficulties of the problem by using newly developed dataset.

## 1. はじめに

著者らの研究グループでは、人と議論ができるディベート型人工知能の研究を進めている [Sato 15]. ディベート型人工知能は、与えられた議題に対して、その賛否の根拠を提示することで、政策決定者や経営者の判断を支援する. たとえば、「カジノを建設すべきか」という議題が与えられると、「地域経済の活性化につながるので禁止すべき」や「ギャンブル依存症の問題を引き起こすので建設すべきでない」などの賛否の理由と、その根拠となる事例をニュース記事や白書の中から抽出して提示する.

本稿では、ディベート型人工知能のコア部分の1つとなる証拠性推論を定式化し、その問題の性質を議論する. 具体例として、以下の2文の関係性を考える.

$q$  : カジノを建設すべきか

$t$  : ミシシッピ州では 31 箇所あるカジノにより、およそ 35,000 人の雇用が創出されている

意思決定上の質問  $q$  に対し、テキスト  $t$  は肯定的な証拠となっていると考えることができる. これを抽象化して、意思決定上の質問  $q$  に対して、テキスト  $t$  が肯定的あるいは否定的な証拠になっているかどうかを判定する問題として、証拠性推論を定義することができる.

証拠性推論は、テキスト含意関係認識 (Recognizing Textual Entailment) と形式的には似ている. テキスト含意関係認識とは、2つのテキスト  $t_1$  と  $t_2$  が与えられたときに、 $t_2$  が  $t_1$  に意味的に含まれるかどうかを判定する問題である. たとえば、

$t_1$  : ミシシッピ州では 31 箇所あるカジノにより、およそ 35,000 人の雇用が創出されている

$t_2$  : 雇用を生み出すカジノもある

$t_3$  : カジノは雇用を生み出す

の2文があったとき、 $t_1$  から  $t_2$  を厳密に導くことができる. 一方、 $t_3$  は一般論として述べているのだとすると、 $t_1$  は特殊な事例について述べているかもしれないため、 $t_1$  からだけでは導くことができず、含意関係は成り立たない. あきらかに上

連絡先: 柳井 孝介, 日立製作所 研究開発グループ,  
kohsuke.yanai.cs@hitachi.com

記の  $q$  と  $t$  は含意関係ではなく、証拠性推論はテキスト含意認識とは別の種類の問題といえる.

再度、 $q$  と  $t$  の関係に注目すると、本当に  $t$  が肯定的な証拠になっているか、に関しても考察を要する. 仮に、カジノを建設しようとしている地域が、労働者不足だったとすれば、 $t$  は意思決定において肯定的な証拠にはならないと思われる. より簡単な例では、「商品 X を購入すべきか」という意思決定上の質問があったときに、「商品 X の色はグレー」というテキストが、肯定的な証拠になっているか、というものがある. 購入しようとしている人がグレーが好きであれば、肯定的な証拠であるし、そうでなければ、肯定的な証拠にはならない. したがって、証拠性は、意思決定をしようとしている人の価値観のようなものに依存すると考えることができる. この価値観とは、意思決定をしようとしている人の願望、目的、好みなどとみなすことができる. たとえば、下記のもので、意思決定を行うときに価値観の一種と考えることができ、これらは、意思決定する人や組織のコンテキストによって変わると考えられる.

雇用を生み出す  
色がグレー系である  
アジアで新事業を立ち上げる  
貧困をなくす  
持続可能な生産消費形態を確保する

本稿では、2章で関連研究を述べ、3章で証拠性推論問題を定式化し、作成したデータに関して説明する. 4章で作成したデータに基づいて証拠性推論がどのような性質の問題なのかを分析し、5章で今後の方向性を述べる.

## 2. 関連研究

著者らの研究グループでは、類似の自然言語処理のタスクとして、サポート性推論を定式化し、それを解く手法を提案している [佐藤 16]. サポート性推論では、議論の対象  $o$  と注目する価値  $v$  とテキスト  $t$  が与えられたときに、テキスト  $t$  が  $o$  と  $v$  の間の促進作用または抑制作用をサポートしているかどうかを推定する. サポート性推論は、本稿で議論する証拠性推論を解く際の部分問題の1つと考えることができる. 証拠性推論は、サポート性推論で扱っていた問題を再整理し、より一般化された問題として定義したものと捉えることができる.

また著者らの研究グループでは、人の価値観を体系化した価値観オントロジを構築し、それに基づいてテキストデータを分類したり要約したりするシステムを提案している [柳井 15]。本稿では、価値観をオントロジで表すには限界があると考え、オントロジではなく、自然言語文の集合で価値観を表現する。

Rinott らは Context Dependent Evidence Detection というタスクを提案している [Rinott 15]。たとえば「ドーピングはアスリートの健康に悪い」という主張が与えられたときに、文書の中から、定量的な証拠、専門家の見解、事例的な証拠を分類して抽出する。このタスクは、証拠に特化したテキスト含意関係認識に近いと考えられる。

### 3. 証拠性推論問題

#### 3.1 定義

意思決定上の質問  $q$ 、価値観を表す文集合  $\{v_1, v_2, \dots, v_n\}$ 、テキスト  $t$  が与えられたときに、 $t$  が各  $\{v_i\}$  に対して、以下のいずれかであることを推定する。

- PE: 肯定する証拠 (positive evidence)
- NE: 否定する証拠 (negative evidence)
- N-PE: 肯定していない証拠 (non positive evidence)
- N-NE: 否定していない証拠 (non negative evidence)
- NA: 証拠ではないもの (not evidence)

肯定していない証拠とは、たとえば、「X州では、カジノにより失業率は改善しなかった」のように、意思決定上の質問  $q$  に対する肯定を否定するような証拠である。このテキストは「カジノを建設すべきか」を否定もしていないことに注意する必要がある。同様に、否定していない証拠もありうる。N-PE や N-NE は、数は少ないと思われるが、意思決定において反例を見つけるときに重要となる。

証拠性推論問題の具体例は、以下のようなものである。

- $q$  : カジノを建設すべきか  
 $v_1$  : 雇用を生み出す  
 $v_2$  : 持続可能な生産消費形態を確保する  
 $v_3$  : 陸域生態系の保護、回復、持続可能な利用を推進する  
 $t$  : ミシシッピ州では 31 箇所あるカジノにより、およそ 35,000 人の雇用が創出されている

このときに、 $t$  が、 $v_1$ 、 $v_2$ 、 $v_3$  のそれぞれに対して、上記の PE、NE、N-PE、N-NE のどの証拠か、あるいは証拠になっていないかを推測する問題である。 $\{v_i\}$  は 1 つでも、自然言語処理のタスクとしては成立するが、実用的なユースケースでは複数の価値を考慮する必要があることが多いため \*1、あえて複数の価値があることを問題の定義に含めた。

通常、 $t$  は新聞や白書の 1 文やパラグラフを使うことが考えられるため、その長さを制限するのは難しい。一方で、意思決定上の質問  $q$  と価値観を表す文集合  $\{v_i\}$  は、ある程度、ユースケースで制御できる。あきらかに  $q$  や  $\{v_i\}$  の文長が長くなれば、自然言語処理のタスクとして難しくなる。そこで、テキ

\*1 ある商品を購入するケースであれば、色以外にも様々な観点を考慮するものと思われる。

表 1: 分析用のデータセット

項目	数
意思決定上の質問 $q$ の種類	12
$q$ あたりの平均の価値観数 $n$	4.2
$(q, v_i)$ の数	52
$(q, v_i, t)$ の数	101
肯定する証拠 (PE) の数	54
否定する証拠 (NE) の数	37
肯定しない証拠 (N-PE) の数	6
否定しない証拠 (N-NE) の数	4

スト  $x$  に含まれる節の数  $L_{ph}(x)$  を用いて、タスクの難しさを表す近似として、

$$TEI(L_{ph}(q), \frac{1}{n} \sum_{i=1}^n L_{ph}(v_i)) \quad (1)$$

と書くことにする。これは単に、意思決定上の質問  $q$  に含まれる節の数と、価値を表す文集合の節の数の平均を表している。たとえば、上記の証拠性推論問題の例の場合は、係り受け解析ツールである CaboCha [Kudo 02] のチャンク数で節の数を数えることにすると、TEI(2.3, 4.1) と表せる。意思決定上の質問  $q$  が複数あるときは、それぞれの数を  $q$  に対して平均を取るものとする。

### 4. 問題の性質の分析

証拠性推定の問題の性質を分析するための小規模のデータセットを作成した。日本語の新聞記事を文で区切ったものを  $t$  の対象とした。表 1 にデータセットの概要を示す。データセットを証拠性推論の評価用データとみなした場合のタスクとしての難しさは、TEI(2.3, 4.1) であった。以下では CD-毎日新聞データ集 \*2 のテキストから引用し、例を示しながら問題の分析をする。

[例 1]

- $q$  : カジノを建設すべきか  
 $v_1$  : 財政収入を増やす  
 $t$  : 政府の説明では、カジノの解禁で年間一億五千万スイスフラン (約百二十億円) の収入が見込まれるとしている。

肯定する証拠と人手で分類された例である。厳密に考えれば、「カジノの解禁」が「カジノを建設」することにつながり、「カジノを建設」することで、収入が増えることを、 $t$  が暗に示している。厳密に意味を考えれば、「カジノを建設」が間に挟まることを推論する必要がある。しかしこの例に限れば、そこまで深い意味解析や推論をせずに、単に「カジノ」や「収入」の文字列を追うだけでも、自動で分類ができる可能性がある。この例では、 $v_1$  を「財政収入が増える」と読み替えれば、 $t$  と  $v_1$  は含意の関係にある。もし  $t$  が「カジノを建設すれば年間一億五千万スイスフランの...」であれば、 $q$  の「カジノを建設(する)」が、 $t$  の仮定部分に含意されていると考えられる。一般的には、肯定する証拠 (PE) の場合は、 $t$  が  $q$  と  $v_i$  を含意していることもであると予想される。ただしそれだけでは十分条件ではなく、 $t$  のうち  $q$  を含意している部分と、 $v_i$  を含意し

\*2 <http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html>

ている部分で、因果関係等のなんらかの関係が成立していることが必要である。

[例 2]

- $q$  : カジノを建設すべきか  
 $v_1$  : 財政収入を増やす  
 $t$  : ただ先の「リスク」にも、カジノを解禁した米国のある州でカジノ収入 1 ドルあたり州政府が防犯対策などで 3 ドルのコストを負担したとの推計が紹介されている。

否定する証拠と人手で分類された例である。  $t$  の構造を簡潔に示すと下記のようになり、階層化された句構造を持っていることがわかる。

ただ先の「リスク」にも、  
 カジノを  
 解禁した  
 米国のある州で  
 カジノ収入 1 ドルあたり  
 州政府が防犯対策などで  
 3 ドルのコストを  
 負担したとの  
 推計が  
 紹介されている。

$t$  が階層化された句構造を持っているため、厳密に意味を追うには、階層構造を辿るやや複雑な処理が必要となる。「財政収入が増える」ことは、 $t$  の「カジノ収入 1 ドル」の部分に含意されているが、収入とコストを合計するとマイナスであると解釈される例である。ただ、「カジノ」「解禁」「コスト」「負担」などのキーワードが出現しているため、Bag-of-Wordsなどで意味を理解せずに、ラベルを当てただけなら正解できる可能性がある。この例からも分かるように、 $t$  と  $v_i$  の間に含意関係があっても、逆の証拠となる場合がある。

[例 3]

- $q$  : クラブ活動を廃止すべきか  
 $v_1$  : 健康的な生活を確保する  
 $t$  : 中学校の教員は「部活の過熱は過労死予備軍の育成」と訴えてきた。

肯定する証拠と人手で分類された例である。厳密に言えば、「部活動は過熱しやすい」という前提があれば、推論が成立する例と考えられる。データを作成する際には、どの程度、このような前提を許容するかを取り決める必要があると思われる。細かい点では、クラブ活動と部活動が、このコンテキストでは同一の意味を持つことを判定する必要がある。さらにクラブ活動とサークル活動は、似た意味を持つと解釈されるコンテキストもあるが、上記のコンテキストだとサークル活動では成立しないように思われる。一方で、「教員が訴えてきた」ということでもって、意思決定の証拠となるかどうかは、人によっても判断が分かれるところと考えられる。[Rinott 15] のように、一般的に人が証拠とみなすテキストの種類を分類し、どの種類まで証拠としてデータに含めるかを定める必要がある。

[例 4]

- $q$  : クラブ活動を廃止すべきか  
 $v_1$  : いじめをなくす  
 $t$  : 部活動は、いじめ問題に有効な方策になり得る。

否定する証拠と人手で分類された例である。ただし、 $t$  が事実ではなく一般論であるため、このような形式のテキストを証拠としてよいか、よく検討し、データセットを作成するときに定義しておく必要がある。

[例 5]

- $q$  : クラブ活動を廃止すべきか  
 $v_1$  : いじめをなくす  
 $t$  : 名門といわれる学校の運動部や地域の少年スポーツクラブの中では、しごきの名の下に日常的な暴力が正当化されている。

肯定する証拠と人手で分類された例である。同じ価値に対しても、肯定と否定の両方の証拠があることがわかる。従って矛盾のない公理系のようなもので推論するのは、難しいことがわかる。先の例と合わせて、実際の意思決定の際には、もう少し詳しい事例を知りたくなると思われる。ただし実用上は、 $t$  の出典元の  $t$  の前後に、具体的な例が書かれていると考えられるため、別の手法で、 $t$  の前後に具体例は事例が書かれているかどうかを判定し、その有無でフィルタリングする方法で対応できると考えられる。

[例 6]

- $q$  : 貿易自由化を推進すべきか  
 $v_1$  : 農林水産業・農山漁村の所得を向上させる  
 $t$  : フィリピン大学のサルバドル・カテロ教授は「貿易自由化による利益は、農民にまで届きにくい。

肯定しない証拠 (N-PE) として人手で分類された例である。 $t$  は、否定もしていないことに注意が必要である。

## 5. おわりに

本稿では、証拠性推論問題を定式化し、問題を分析するための第一弾のデータセットを作成した。今後は、アノテーションの定義を厳密化し、より精密で大規模なデータセットを整備する。特に実用性を重視して、TEI がより高いデータセットを作成する。またアノテータアグリーメントを測定し、自然言語処理の問題としてどの程度成立しているかを明らかにすることが必要である。続いて、ディープラーニング等を用いて、証拠性推定問題を解くための手法を構築する。例えば、[柳瀬 16] で提案されているように、 $q$  と  $v_i$  と  $t$  を入力し、アテンション機構を使って、 $q$  と  $t$  のアライメントをとり、同様に  $v_i$  と  $t$  アライメントをとって、最終的な識別をする方法が考えられる。一方で、高精度のアルゴリズムの開発は困難と思われるため、精度が不十分でも有用性を発揮できるユースケースを探し、本問題の実用性を実証していく。

## 参考文献

[Kudo 02] Kudo, T. and Matsumoto, Y.: Japanese Dependency Analysis using Cascaded Chunking, in *Proceedings*

of *CoNLL 2002 Post-Conference Workshops*, pp. 63–69 (2002)

[Rinott 15] Rinott, R., Dankin, L., Perez, C. A., Khapra, M. M., Aharoni, E., and Slonim, N.: Show Me Your Evidence - an Automatic Method for Context Dependent Evidence Detection., in *EMNLP*, pp. 440–450 (2015)

[Sato 15] Sato, M., Yanai, K., Miyoshi, T., Yanase, T., Iwayama, M., Sun, Q., and Niwa, Y.: End-to-end Argument Generation System in Debating, in *Proceedings of ACL-IJCNLP 2015 System Demonstrations* (2015)

[佐藤 16] 佐藤 美沙, 柳井 孝介, 柳瀬 利彦, 三好 利昇, 是枝 祐太, 丹羽 芳樹: 意見文章自動生成のための組合せ構文特徴を用いたサポート性推定, *人工知能学会論文誌*, Vol. 31, No. 6 (2016)

[柳井 15] 柳井 孝介, 三好 利昇, 柳瀬 利彦, 佐藤 美沙, 丹羽 芳樹, Reisert, P., 乾 健太郎: ディベート人工知能における意見生成, *人工知能学会全国大会論文集*, 第 29 巻, pp. 1–3 (2015)

[柳瀬 16] 柳瀬 利彦, 柳井 孝介, 佐藤 美沙, 三好 利昇, 丹羽 芳樹: Neural Attension Model を用いた観点付き評判分析, *人工知能学会全国大会論文集* (2016)