

Web上の意見に対する論理的解析手法の提案

Proposal of Logical Analysis of Opinions on The Web

十松和生 *¹ 岡田将悟 *² 新田克己 *³
 Kazuo Jumatsu Shogo Okada Katsumi Nitta

*¹東京工業大学大学院 総合理工学研究科

Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

The study propose logical analysis approach to opinion on the Web. There is a lot of information on the Web, however we are susceptible to being influenced by biased information. Therefore We think you can have the appropriate judgment for the information by evenly gathering and organizing for information. To this end, we propose a method of presenting the organized information by Argumentation Framework.

1. はじめに

世の中には多くの情報が溢れており、様々な話題に関する議論が行われているが、情報が分散しており、全体としてどのような結論になっているのかを判断することは容易ではない。Web上の情報にはサイトによって偏りがみられるかもしれないし、新聞、テレビのようなメディアは、同じ情報であっても各社の異なる解釈が含まれているかもしれない。

これらの状況から、満遍なく情報を集め、整理して提示することで、情報に対して適切な判断が出来ると考えられる。現在、MEDIA WATCH JAPAN のように、ある特定の話題に関して、Web上に存在する意見を収集して、意見別にまとめて提示しているページが存在しており (*¹)、興味ある話題のWebページに効率よくアクセスしたいユーザーに利用されている。

しかし、このような意見まとめページは、個々のWebページにはアクセスできるものの、それらのWebページの意見を統合し全体としてどのような結論が得られるのか判定しにくい。

そこで、本研究では、対立する争点の間の関係を整理して争点の構造化を行い、それぞれの争点ごとに関連する情報を収集し、数理議論学を利用して議論全体の結論を求める手法を提案することを目的とする。

2. 関連研究

2.1 意見を整理したサイト

先述した、MEDIA WATCH JAPAN では、賛否が分かれるような政治的テーマの論点整理を行っており、対立した主張を把握した上で自分のスタンスを持つのが良いという観点で、政党や主要新聞のスタンス比較を行っている。テーマ毎に論点が与えられており、わかりやすく各政党、各社の立場を列挙している。図1は 原発政策をテーマに、賛成派と反対派による意見を列挙したものであり、それぞれの争点ごとに、関連記事を引用してその分析を行っている。

2.2 情報の信頼性

従来より情報の信頼性に関する研究は行われており、例えば Inui ら [1] は、Web上に存在するテキスト情報は、その全てが

連絡先: 十松和生, 東京工業大学大学院 総合理工学部 知能システム科学専攻, 神奈川県横浜市緑区長津田町 4259 J2-53, 045-924-5205, jumatsu.k@ntt.dis.titech.ac.jp

*¹ <http://mediawatchjapan.com/>

脱原発派の主な主張	原発活用派の主な主張
• 原発なしでも電力はまかなえる • 廃棄物処理や事故リスクを考えれば原発は高コスト • 「核のゴミ」処分が確立しておらず無責任 • 脱原発は国民的議論で確認した民意である • 地震国日本では安全性の確保が難しい • 実効性のある非難計画なしでは無責任 • 安全審査には火山噴火リスクも考慮すべき • 周辺自治体の同意を得るべき	• 電力の安定供給には原発が必要 • 原発は燃料費が安く、安価な電力を得るために欠かせない • 火力発電の燃料費がかさみ国富が流出 • 「原発ゼロ」目標は無責任で撤回は妥当 • エネルギー安全保障上、多様な電源が必要 • 原子力関連の人材・技術継承の必要がある • 温室効果ガス削減のためにも重要 • 電力会社と安全協力を結ぶ立地自治会の合意で十分

図 1: テーマに関係した意見の列挙

真実というわけではなく、不正確な記述、偏りのある意見、陳腐化した情報などが混在している可能性が高いという観点から、ユーザが着目したある言明に関するトピックの文章集合から、そのトピックに対する多種多様な言明を抽出し、それらの間の意味的関係を解析して、俯瞰図をユーザに提示する事を目的とするシステムを開発した。言明を俯瞰的に概観するため、各言明を区別し、整理して言論マップとして提示した。

3. 議論フレームワーク (Argumentation Framework)

議論中の様々な意見から、「確実に成り立つことは何か」、「確実ではないが、成り立ちうる事は何か」、「絶対に成り立たない事は何か」を識別し、どの意見 (論証) が結論として言えるかを判断するには数理議論学が有効である。数理議論学においては、議論構造を「議論集合」と、それらの間に存在する「攻撃関係」により、議論フレームワーク (Argumentation Framework: AF) として表現し、その議論構造から導き出せる論証の成否を定義している [2]。

また、Bench-Capon は、AF を拡張して、論証の価値を考慮して議論の結論を判定する Value based Argumentation Framework (VAF) を提案した [3]。

AF と VAF は以下のように定義されている。

$$AF = \langle A, R \rangle$$

$$VAF = \langle A, R, V, val, valpref \rangle$$

A : 論証集合, R : 論証間の攻撃関係

V : 価値の集合

val : A の要素から V の要素へと写像する関数,

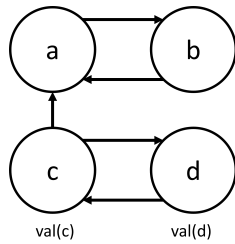


図 2: Argumentation Framework

$valpref$: 価値間の優先関係,

図 2 の例では,

$A = \{a, b, c, d\}$, $R = \{(a, b), (b, a), (c, a), (c, d), (d, c)\}$ である。

この例では、絶対に成立する論証（基礎拡大）は空集合（存在しない）であるが、成立する可能性のある論証集合（選好拡大）は、 $\{\{a, d\}, \{b, a\}, \{b, c\}\}$ である。

ここで、論証 c と論証 d の価値がそれぞれ $val(c)$, $val(d)$ であり、 $valpref$ で $val(d) > val(c)$ であるならば、図 2 において、c から d への攻撃は無効になり、基礎拡大は $\{d\}$ 、選好拡大は $\{\{a, d\}, \{b, d\}\}$ となる。

4. 提案手法

4.1 概要

図 3 で提案手法の流れを説明する。まず、興味のある問題に関する議論の争点構造を整理し、AF の形式で表現する (4.2 節)。図 4 は原発再稼働問題における AF の例である。次に AF の各論証に関連する Web ページや文献をインターネットや文献データベースなどから検索、収集する (4.3 節)。これらを AF の論証に結びつけ、その論証の価値判断を行う (4.4 節)。最後に、その価値を利用して、VAF として議論全体の結論を求める (4.5 節)。

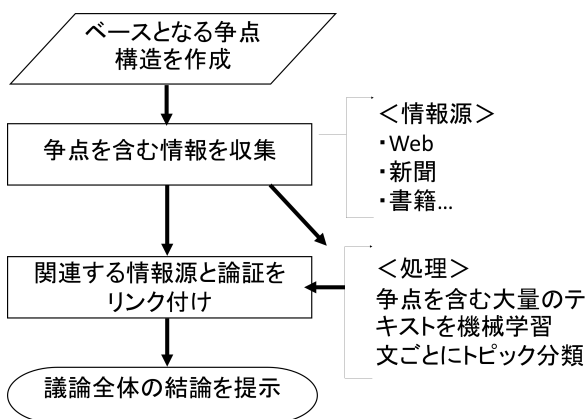


図 3: 提案手法

4.2 ベースとなる争点構造の作成

基本となる争点構造を作成するには、まず図 1 のようにテーマに関する争点を列挙し、それらの間の攻撃関係、支持関係を

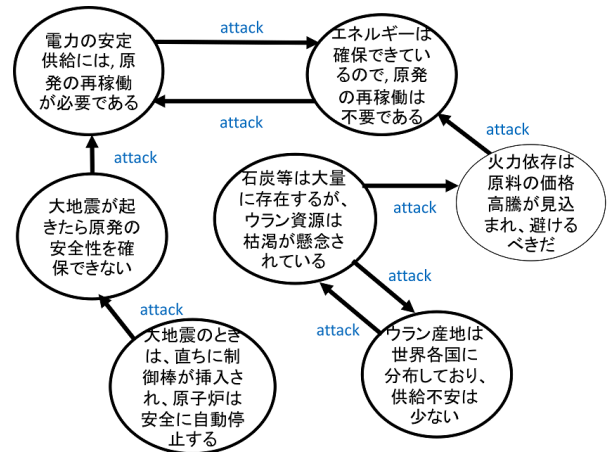


図 4: 議論の争点構造

判断する。2つの争点間に支持関係があるときは、それらを1つの論証としてまとめることにより、AF が構築される。

さらに、各論証に関連するキーワードを設定しておく。後述するように、それぞれの論証に関する Web ページや文献情報などがある程度収集されている場合には、呂らによる ssLDA 手法によるトピック推定手法を利用し、それぞれの論証に該当するかどうかを判定するためのキーワードを抽出する [4]。

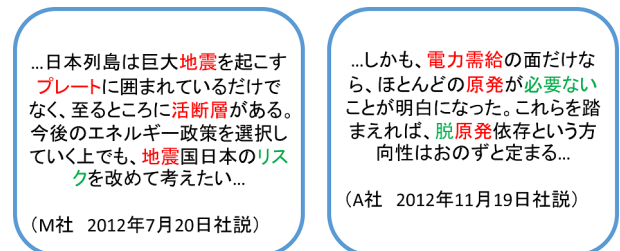


図 5: 論証を含む社説の例

4.3 意見収集方法

争点構造にある論証を含む情報を探すために、Web ページや文献情報のテキストに対して、各論証のキーワード比較をして、トピック推定を行う。その際に各論証に対しても、生成されたトピックの中から最も適しているトピックを割り当てる。その後、各論証と同じトピックの文に対して、予め設定しているキーワードで検索をかけて絞り込み、その文に対して構文解析を行い、否定の判定をすることで、論証推定の精度を上げる。

6 新聞社の 2008 年 3 月から 2015 年 3 月まで約 7 年分の社説を集めたデータベースを対象に意見収集を行うと、図 5 の例に示すような社説が検索される。

4.4 論証へのリンク付け

図 6 は 関連ページを論証にリンク付けしている図であるが、図 4 の「大地震が起きたら原発の安全性を確保できない」という論証について、M 社の 2012 年 7 月 20 日の記事に含まれていると判断できたため、その論証に関連しているものとしている。記事に含まれている赤い文字はキーワード、リスクという単語がネガティブな評価を表している。図 7 では、各論証ごとに関連する情報源の数を論証に添えている。

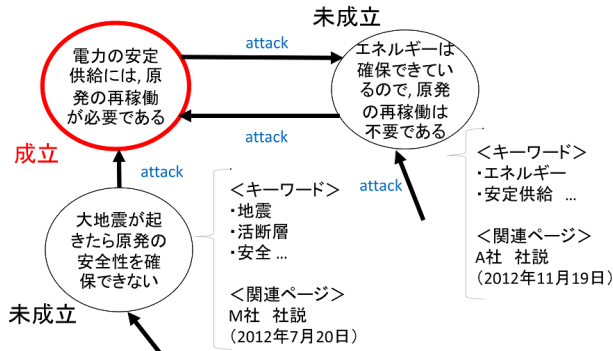


図 6: 論証へのリンク付け

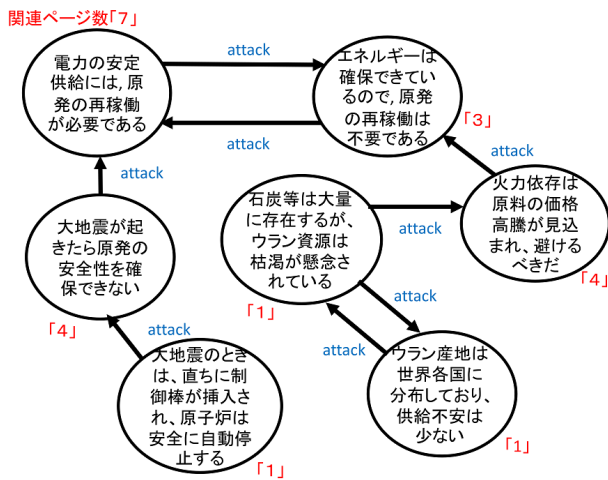


図 7: 各論証に関するページ数

4.5 VAF による結論

AF の各論証の価値を、その論証にリンクされた Web ページや文献情報で評価する。評価方法には、単純に論証にリンクされた Web ページや文献情報の数で価値を決める方法や、Web ページや文献の信頼度を考慮することも考えられる。

図 7 の例で、単純にリンクされた数で論証の価値評価を行うと、図 8 のように、価値が低い論証から価値が高い論証への攻撃は無効になる（点線で示された矢印は無効）。結果として、成立する論証集合は、「電力の安定供給には、原発の再稼働が必要である」や、「火力依存は原料の価格高騰が見込まれ、避けるべきだ」、「大地震が起きたら原発の安全性が確保できない」、「大地震のときは、直ちに制御棒が挿入され、原子炉は安全に自動停止する」が根拠拡大の要素となる。

5. おわりに

今回提案する手法では、争点構造を可視化し、さらにその議題の結論を表示することにより、適切な情報判断を行える可能性があると考えている。結論を出す事を重要視しておらず、このようなやり方で求めた場合はこのようになる、という意味であり、それを見てどのように判断するかは、使用者に委ねる部分であり、基本的にはできる限り客観性を持つものを提示したい。今回提案した手法では、VAF の論証に与えられた *val* の値の大

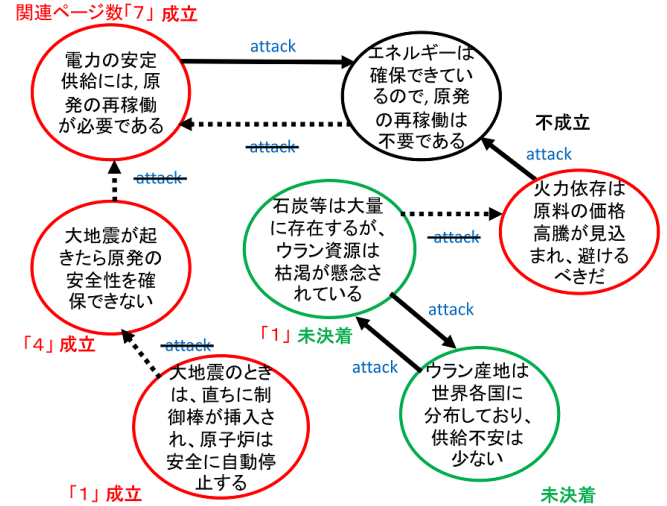


図 8: 議論の結論

小を単純に比較して、論証の強弱としたが、例えば、*valpref*s の値を、様々な政治的観点、個々の立場によって変更する事で、結論の提示方法を変更するといった事も考慮していく方法も考案していきたい。さらに、情報源に注目することで、Web サイトごとの意見の偏りや、意見の時系列変化等の提示も行いたいと考えている。

参考文献

- [1] Phan Minh Dung: "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games", Artificial Intelligence 77, ELSEVIER, 1995.
- [2] T. J. M. Bench-Capon: Persuasion in practical argument using value-based argumentation frameworks. Journal of Logic and Computation, 13(3):429-448, 2003.
- [3] Organizing Information on the Web through Agreement-Conflict Relation Classification. Junta Mizuno, Eric Nichols, Yotaro Watanabe and Kentaro Inui. In Proceedings of the Eighth Asia Information Retrieval Societies Conference (AIRS2012), 2012.
- [4] Lu, Youwei, Shogo Okada, and Katsumi Nitta. "Semi-supervised Latent Dirichlet Allocation for Multi-label Text Classification." Recent Trends in Applied Artificial Intelligence. Springer Berlin Heidelberg, 2013. 351-360.