3M3-2in

ディベート人工知能における意見生成

Argumentation Generation for Debating AI

柳井 孝介 *1 Kohsue Yanai
三好 利昇 *1 Toshinori Miyoshi
柳瀬 利彦 *1 Toshihiko Yanase
佐藤 美沙 *1 Misa Sato
丹羽 芳樹 *1 Yoshiki Niwa
Paul Reisert*2
乾 健太郎 *2 Kentaro Inui

*1日立製作所中央研究所 Central Research Laboratory, Hitachi Ltd.
*2東北大学大学院情報科学研究科 Graduate School of Information Sciences, Tohoku University

This paper discusses "value ontology", which is used to generate argumentation scripts in a debating AI system. The value ontology formulates a set of values that represents what is important in human's life, what is harmful in communities, etc. In this paper, we describe a methodology for creating the value ontology, and how it is used in automatic argumentation generation.

1. はじめに

著者らの研究グループでは，人と議論ができるシステムの研究を進めている。このシステムは，与えられた議題に対し意見を述べたり，人の意見に対して根拠や反論を提示したりすることで，人と議論をしながら問題の本質や解決策を見つけるためのものである。このようなシステムにより，組織における意思決定をより合理的なものに変えることができると考えている。本稿ではこのシステムをディベート人工知能と呼ぶ。これは，システムの達成度を測るためのわかりやすい指標として，競技ディベートで人に勝つことを目標としていることによる。またビジネス，司法，政治，医療などの多くの場面でなされる意思決定という行為のエッセンスが，競技ディベートに集約されているとの考えにもよる。

ディベートと意思決定の類似性を見るため，例として，「カジノを禁止すべきか」に関して意思決定することを想定する。ここで以下の 2 つの言明のみが与えられたと仮定する。

A Yes, ギャンブル中毒とコカイン中毒の脳の状態が似ているという研究結果が示されており，中毒性が高い。

B Yes, カジノがもらたす利益より，地域社会に与える悪影響の方が大きい。

この場合，A は明確にカジノがもたらす影響（ギャンブル中毒）を述べており，かつ関連する学術的証拠まで提示している。一方，B は明確な根拠が述べられておらず，主観的な印象とも受け取れる。そのため A の方が納得性が高く，もし他の情報や前提条件が提供されなければ，組織においては A の意見に基づいて意思決定がなされると思われる。競技ディベートにおいても証拠が重視される。また A に対して，

C 日本では成人の 5%がギャンブル依存症で世界でも突出している。

D ギャンブル依存症を防止する方法は様々あり，実施している国もある。

連絡先: 柳井 孝介，日立製作所中央研究所，
kohsuke.yanai.cs@hitachi.com

表 1: 主観評価の結果:
研究チームのメンバ 5 名で 50 論題に対する合計 150 の出力された意見を評価。0: 意味不明，1: 主張が理解できる，2: 主張が明確，3: 主張が明確で納得性がある。

評価値	ver2	ver3	ver4	ver5	ver6
3 以上	0	1	0	12	10
2 以上	2	17	15	20	26
1 以上	18	61	57	50	64
0	130	89	93	100	86

の 2 つの意見が与えられた場合，C は A の意見の納得性を強めており，一方で D は納得性を弱めていると感じると思われる。競技ディベートでは，C は重大性，D は内因性否定と呼ばれており，ディベート理論としてまとめられている [松本 09] 。「カジノを禁止すべきか」はディベートでの典型的な議題であるが，上記の例からもわかるように，競技ディベートと組織での意思決定で共通する部分がある。

ディベート人工知能に関連して，これまでに，心の社会に着想を得た人工知能アーキテクチャ [柳井 14] ，意見の材料となる文の収集および分類手法 [三好 14] ，材料文を並び替えて意見を生成する手法 [柳瀬 14] ，ナレッジベースと Toulmin Model を用いて意見を生成する手法 [Reisert 15] ，生成された意見を抽象化して可読性を高める手法 [佐藤 15] を提案してきた。これらの手法を統合して，与えられた議題に対して，7 センテンス程度の英語の意見を生成する人工知能システムを構築している *1。表 1 にシステムが出力した意見に対する主観評価の結果を示す。本稿では，最近の精度向上の主要因となった価値オントロジについて議論する。価値オントロジとは，人工知能システムに人の価値観の違いを持たせるためのものであり，本研究は，コンピュータにディベートをさせるために，この価値オントロジに注目したという点に新しさがあると考えている。価値オントロジを中核に据えているところが，ディベート人工知能が，他の検索や複数文書要約システムなどと大きく異なる点だと思われる。

*1 システムは日々更新されており，その時のバージョンによっては，提案手法がすべて統合されていないこともある。

表 2: 価値オントロジ

分野	価値	極性	表現語	文脈語	対立価値
health	disease	-1	disease:36.6, complication:40.1	AIDS, Alzheimer, ...	fortune, family, ...
economy	investment	1	investment:27.8, development_aid:48.2	asset, bank, capital, ...	income, environment, ...
finance	income	1	revenue:35.4, wage:39.8	budget, company, ...	cost, corruption, ...
finance	cost	-1	expense:35.9, expenditure:55.7	fund, fuel, lower, ...	technology, health, ...
...	...	...	...	...	...

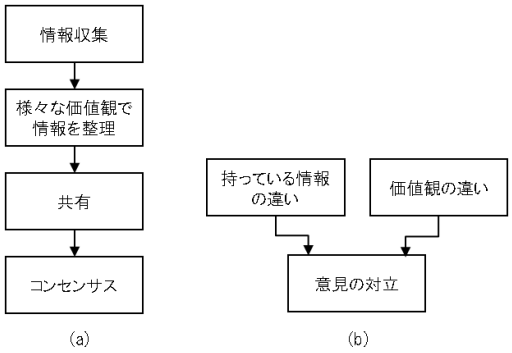

図 1: 意思決定と意見の対立

以下では 2 章で価値オントロジの構造について述べ，3 章で価値オントロジがどのように使われるかを説明する。

2. 価値オントロジの構造

価値オントロジの具体的な説明の前に，価値オントロジの必要性について論じたい。前述の意見 A（ギャンブルは中毒を引き起こす）に対立する意見として，以下の意見 E を考える。

E No, トルコでは隣国から毎年数十万人の観光客がカジノを訪れ，地域経済の活性化に貢献している。

この意見は,「地域経済」という別の価値から意見を述べており，A と E が与えられた場合，意思決定者は,「住民が精神的に健康であること」と「地域経済が活性化されること」の 2 つの価値のうち，どちらが重要かという点で最終的な意思決定をする。「精神的な健康」に価値をおいている人と,「地域経済」に価値をおいている人で，意思決定は異なるであろう。

この具体例を念頭に，図 1 を使って意思決定と意見の対立について一般的に論じたい。図 1 の (a) は企業での代表的な意思決定である稟議の流れを示したものである。まず判例や根拠など意思決定に必要な情報を収集し，続いて様々な価値で情報を整理する。続いてこれらをステークホルダーで共有し（根回し），最終的に会議にてコンセンサスをとる。一人の場合であっても，重要な意思決定する場合には，複数の価値でメリット/デメリットを洗い出していると思われる。図 1 の (b) は，意見の対立が起こる構図を示している。意見の対立は，持っている情報の違いと価値観の違いから生じる。例えば前述の意見 A と E では,「精神的な健康」に重きを置いている人と,「地域経済」に重きを置いている人では，意見の対立が生じ得る。(a)(b) の双方において，価値観が重要な役割を持つ。

以上の議論から，ディベートや意思決定の場での，意見の選択/コンセンサス/対立において，値観の違いがその根本にあることが推察される。この考察から， ディベート人工知能とは，人の価値観を中心に据えて，それぞれの価値を高めたりあるいは価値を落としたりする事象を，人の価値観を軸にアラインすることと捉えることができる。そこで人の価値観をオントロジとしてまとめ，機械的に利用可能とする必要性が出てくる。本研究では,「精神的な健康」「地域経済」などの個々の価値の集合として形成されるものを価値観と考え，人の価値観の違いとは，それぞれの価値の重みの違いと捉えている。価値オントロジはこの価値観の違いを表現するためのものと位置付けている。言論マップ [村上 10] [Watanabe 12] では，Web 上に発信された言明を，同意，矛盾，根拠などの意味的関係で整理している。価値オントロジは言明ではなく，価値感の間の関係を記述することを目指している。

表 2 に価値オントロジの一部を示す。文献 [三好 14] では，価値促進抑制テーブルと呼んでいたが，その後，データ構造や役割が変わり，価値オントロジと呼ぶ方が適切と考えている。表 2 に示すように，大分類として分野があり，その下に価値がある。各価値は，極性，表現語，文脈語，対立価値からなる。極性は，その価値が人生やコミュニティにとって良いものか，悪いものかを表す。表現語は，その価値が文書中に現れるときの自然言語上の表現であり，付記されている数値は重みを表す。文脈語は，その価値が文書中に現れるときに共起しやすい語であり，主に表現と意味の間の曖昧性を解決するために使われる。表 2 の例では，価値 disease の文脈語として AIDS と Alzheimer が紐づけられている。対立価値は，意見の対立が起こるときに，その価値と対立しやすい価値を表す。表 2 の例では，disease の対立価値として fortune と family が紐づけられている。これは「病気」を価値の軸として意見を述べているときには,「富」を軸とした反論がなされ得ることを意味している。

価値オントロジは人手で継続的に整備する。本研究では，ディベートのデータベースである Debatabase [deb] を使いながら，以下のような手順で価値オントロジを整備した。

1. 国の機関を手がかりとして分野を決める。例えば，education, economy, industory というように分野の候補をリストアップして人手で整理する。

2. Debatabase の議論を LDA（Latent Dirichlet Allocation）でクラスタリングし，ディベートで頻出する概念を調べ，分野または価値に加える。

3. Debatabase の各記事の価値を表現語のキーワードマッチなどで推定する。推定できた記事からは価値の表現語を新たに抽出し，推定できなかった記事からは価値オントロジに含まれていない新しい価値がないかを探す。またその記事で共起している語を人手で精査し，文脈語として抽出する。これをブートストラップ的に繰り返す。

4. Debatabase の意見とそれに対する反論のペアに対し，同様に議論されている価値を推定する。これにより対立が

生じやすい価値のペア（disease と fortune など）を算出し，対立価値として抽出する。

3. 価値オントロジの活用

価値オントロジはディベート人工知能の人格相当のものと考えることができる。価値オントロジの表現語に付されている重みを変更したり，ユースケースによって価値オントロジの内容を変えたりすることで，ディベート人工知能の振る舞いを変えることを想定している。例えば，企業のミドルマネージャがディベート人工知能を使って議論する場合，自分の価値観にあうように価値オントロジをカスタマイズすることが考えられる。例えば，部下の健康，チームのパフォーマンス，プレゼンスなどを細分化したものを中心に価値オントロジを構築するであろう。逆に，部下の考えを代弁させるため，部下の価値観にあうように価値オントロジをチューニングすることも考えられる。

このような役割を持たせるため，現状のディベート人工知能システムの様々なところで価値オントロジが使われている。以下に価値オントロジの有用性をまとめる。

意見の多様性 図 1 で説明したように意思決定する場合には，複数の価値で情報を整理することが重要である。そこで Gigaword（英文ニュース記事のデータ）[gig] から，促進/抑制を表す動詞（increase, promote, deccrease, suppress など）の目的語に価値オントロジの表現語が含まれている文を探す。例えば，”Casinos cause addiction” などが相当する。価値でクラスタリングすることで， 1 つの価値ではなく，多様な価値で意見をまとめることができる。

議論における価値の遷移のモデル化 Debatabase の記事を文単位で価値を推定し，その議論で議論されている価値がどのように変化するかを統計的にモデル化する。例えば，「精神的健康」の価値で論じた後には，「家庭」の価値で論じることが多いなどをモデル化できる。このモデルをディベート人工知能における意見生成に使うことができる。

反論の生成 相手の意見の軸となっている価値を推定し，続いて，価値オントロジと照らし合わせて，相手の意見の中で抜け落ちている価値や，価値オントロジ上で同じ分野に属する他の価値を求める。あるいは，価値オントロジの対立価値を用いて相手の意見の価値と対立しがちな価値をを求める。この価値に基づいて Gigaword を使って意見を生成することで，相手の意見に対する反論意見を生成する。

4. おわりに

本稿では値オントロジに焦点を当てて，ディベート人工知能における意見生成について論じた。近年，Argumentation Mining という名称で，Web フォーラムや Twitter での議論を分析する研究が盛んになってきている。例えば，ユーザの投稿が議題に対して賛成か反対かを推定する手法 [Murakami 10] [Somasundaran 09] や，議題が与えられたときに文中から主張に相当する部分を抽出する手法 [Levy 14] が提案されている。しかしながら，人の価値観を体系化し，価値観に基づいて議論を解析したり，議論を生成したりする試みは本研究以外にはまだなされていないように思われる。

本稿では 紙面の制約もあり，価値オントロジの意義や必要となる背景など，定性的な議論に終始した。今後は価値オントロジの定量的な評価を行っていく予定である。

参考文献

[deb] DEBATABASE: A WORLD OF GREAT DEBATES: http://idebate.org/debatabase

[gig] LDC Annotated English Gigaword: https://catalog.ldc.upenn.edu/LDC2012T21

[Levy 14] Levy, R., Bilu, Y., Hershcovich, D., Aharoni, E., and Slonim, N.: Context Dependent Claim Detection, in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 1489–1500, Dublin City University and Association for Computational Linguistics (2014)

[Murakami 10] Murakami, A. and Raymond, R.: Support or Oppose? Classifying Positions in Online Debates from Reply Activities and Opinion Expressions., in *COLING (Posters)*, pp. 869–875 (2010)

[Reisert 15] Reisert, P., Inoue, N., Inui, K., Yanase, T., and Yanai, K.: Computationalizing a Toulmin Model for Argumentation Generation, 言語処理学会第 21 回年次大会 (2015)

[Somasundaran 09] Somasundaran, S. and Wiebe, J.: Recognizing Stances in Online Debates, in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing*, pp. 226–234 (2009)

[Watanabe 12] Watanabe, Y., Mizuno, J., Nichols, E., Narisawa, K., Nabeshima, K., Okazaki, N., and Inui, K.: Leveraging Diverse Lexical Resources for Textual Entailment Recognition, Vol. 11, No. 4, pp. 18:1–18:22 (2012)

[佐藤 15] 佐藤 美沙, 柳井 孝介, 三好 利昇, 柳瀬 利彦, 丹羽 芳樹：議論文生成における文抽象化のための固有表現抽象化, 言語処理学会第 21 回年次大会 (2015)

[三好 14] 三好 利昇, 佐藤 美沙, 柳井 孝介, 柳瀬 利彦, 丹羽 芳樹：ディベート立論のための材料文の収集と分類の検討, 第 7 回データ指向構成マイニングとシミュレーション研究会 (2014)

[松本 09] 松本 茂, 鈴木 健, 青沼 智：英語ディベート 理論と実践, 玉川大学出版部 (2009)

[村上 10] 村上 浩司, 水野 淳太, 後藤 隼人, 大木 環美, 松吉 俊, 乾 健太郎, 松本 裕治：文間意味的関係認識による言論マップ生成, 言語処理学会第 16 回年次大会, pp. 559–562 (2010)

[柳井 14] 柳井 孝介, 柳瀬 利彦, 三好 利昇, 丹羽 芳樹, 佐藤 美沙：ディベート人工知能のためのアーキテクチャ, 第 7 回データ指向構成マイニングとシミュレーション研究会 (2014)

[柳瀬 14] 柳瀬 利彦, 三好 利昇, 柳井 孝介, 岩山 真, 丹羽 芳樹：ディベートの意見文章生成のための分散表現を用いた文の並び替え, 第 7 回データ指向構成マイニングとシミュレーション研究会 (2014)