

# 国会会議録を用いたディベート人工知能による意見生成

## Debating AI: Argument Generation by Retrieving Sentences from the Minutes of the Diet

佐藤 美沙  
Misa Sato

柳井 孝介  
Kohsuke Yanai

柳瀬 利彦  
Toshihiko Yanase

是枝 祐太  
Yuta Koreeda

丹羽 芳樹  
Yoshiki Niwa

日立製作所 基礎研究センター  
Center for Exploratory Research, Hitachi, Ltd.

We will demonstrate an argument generation system in debating, which aims to utilize the Minutes of the National Diet for decision-making support. Users can specify a motion such as “This house should ban casinos” and then the system outputs argument scripts both for and against the motion. This paper describes an outline of the system and the method to adapt the system to utilizing text data in the minutes. We evaluated generated arguments which are composed of speeches by members of the Diet, by comparing with generated arguments composed of written sentences from newswire articles.

### 1. はじめに

筆者らの研究グループでは、電子化されたテキストデータを意思決定支援に活用することを目的としてディベートをするシステムを開発している[柳井 15]。このシステムを本稿ではディベート人工知能と呼ぶ。本発表ではディベート人工知能による意見生成のデモンストレーション展示を行う。聴講者が議題を与えると、このシステムは賛成・反対両方の立場から最大 6 つの意見を述べ、それぞれの理由や根拠を提示する。たとえば、「カジノを禁止すべきか」という議題が与えられると、賛成側の場合は「ギャンブル依存症の問題を引き起こす」、反対側の場合は「地域経済の活性化につながる」などを理由として挙げ、続いてテキストデータの中からその理由や根拠となる事例を抽出して提示する。

国会会議録は、日本の国会で行われたすべての本会議、委員会等の会議録であり、国会会議録検索システム\*1を通して容易にダウンロードすることができる。日本の国会での議論が記録されている上に入手も容易な電子化されたテキストという有望な性質を持ちながら、会議発言の書き起こしであるために分量が非常に多く読み物としての整理はなされていないため、そのまま読むには労力を要する。その労力を軽減させる支援ツールを開発することで活用を促進できると考えられる。特に、国会が意思決定の場であり、賛否両論ある議題について国民の多様な価値観に訴えかけるような情報が述べられていると考えられる点に着目すると、賛否や観点を考慮した整理を行う意思決定支援が一つの有望な活用であると考えられる。そこで本稿では、これまで開発してきたディベート人工知能を用いて国会会議録から意見文章を生成することにより、国会会議録の活用を行う。これまでディベート人工知能では、新聞記事や白書といった書き言葉をデータソースとして用いてきたため、新聞記事を利用した意見生成結果と比較を行い、データソースの違いによる変化を確認する。

### 2. 関連研究：国会会議録の活用

国会会議録から有益な情報を抽出し、人間に扱いやすい形で提示する研究がこれまでも行われてきている。まず、会議の内容を要約する研究がある。[山本 05]は原文からできるだけ情報を欠落させない報知的な要約を行うことで、[川端 07]は全文

を読むべきか否かの判断を助ける指示的な要約を行うことで、会議ごとの内容把握を助けている。次に、会議録には発話者の氏名も記載されているため、発話者ごとに発言を要約することも考えられる。[尾崎 14]は、選挙における投票先の判断を補助するために、重要な意見文だけを抽出することで、議員の発言の要約を作成している。加えて、トピック要約の研究もみられる。[金丸 05]は、会議録をトピック分析にかけ、抽出されたトピックごとに要約を作成することを検討している。要約から離れると、会議録に対する計量テキスト分析が行われており、特定の議題についての議論の分析がされている。[山本 11]は議題に日本語教育を選択し、年代別に頻出名詞を論点として抽出し、手作業によりその論点における典型文を抽出しながら時代背景を考慮した考察を加えている。これらの研究では、賛否の別が検討されていない。そこで本研究では、ディベートの枠組みで賛否の区別を扱い、ユーザから入力された多様な議題に対して、論じる観点を決定し、会議録から理由や根拠の文を自動抽出することで賛否両方の立場から意見を組み立てることをめざす。

### 3. ディベート人工知能による意見生成

本研究では、[柳井 15]に説明しているディベート人工知能による意見生成を行う。競技ディベートにおける立論と同様に、論題についての是非を述べることによって意見生成とする。入力される論題には、「カジノを合法化すべきか」のような是非を問う文を想定している。是非を際に、価値というものを取り扱う。ここで言う価値とは、それ自体が人間あるいは社会において良い存在（悪い存在）と捉えられるものであり、たとえば健康は良い価値、貧困は悪い価値、と考えることができる。意見生成の方針は、賛成意見を述べるためには良い価値を増やしていることの指摘を、反対意見を述べるには悪い価値を減らしていることの指摘をすることである。

したがってディベート人工知能における意見生成とは、議論の対象が価値を増減させることを指摘し、その根拠を示すことである。本章では、ディベート人工知能の主要技術である(1)意見を論じる観点を価値観に基づいて決定する価値決定、および(2)テキスト文が根拠としてもっともらしいかを推定するための文の根拠性認識、という2つの技術の概要を説明する。

連絡先: 佐藤美沙, 株式会社 日立製作所 基礎研究センター,  
〒185-8601 東京都国分寺市東恋ヶ窪一丁目 280 番地,  
042-323-1111, misa.sato.mw@hitachi.com

\*1 [http://kokkai.ndl.go.jp/cgi-bin/KENSAKU/swk\\_logout.cgi](http://kokkai.ndl.go.jp/cgi-bin/KENSAKU/swk_logout.cgi)

表 1. 論題「カジノを合法化すべきか」に対して出力された意見の例。  
下線部は根拠性がある(正解である)と考えられる部分を示す。

賛成側	
毎日新聞*2から作成	国会会議録*1から作成
<p>"収入"の観点で論題に賛成する。"カジノ"は"収入"を促進するからだ。</p> <p>カジノの収益は、客の負け分、すなわち「人の不幸」が大きくなるほど増えていくのです。政府統計では、<u>昨年</u>のカジノを中心とした賭博業収入は前年比18.6%増の約3607億バタカ(約4兆6700億円)と過去最高を更新した。マカオの11年のカジノ収入は前年比約4.2%増の約2兆6000億円と過去最高で、ラスベガスの約5倍に上った。</p>	<p>"活性化"の観点で論題に賛成する。"カジノ"は"活性化"を促進するからだ。</p> <p>したがって、カジノ、そんなたくさん人が来るのかなという個人的な思いがある一方で、いろいろな方の話を聞きますと、やはり<u>カジノがあって総合型リゾートということになると、外国を含めて大変人が来られるので、観光産業を初めとする産業活性化に大変役立つ、こういうことをおっしゃる方も多くいらっしゃいます。統合型リゾートは、ホテル、商業施設、国際会議場、カジノなどを備えた複合施設であり、海外からの観光客の呼び込みと地域経済の活性化に大きな役割を果たすという期待をされております。構造改革特区でもカジノの解禁を地域限定で求める提案は数々あったんですけども、しかし、刑法に絡むものは取り上げないということで全部前払いを受けてきた、こういう経過だったというふうに思います。</u></p>
反対側	
毎日新聞*2から作成	国会会議録*1から作成
<p>"依存症"の観点で論題に反対する。"カジノ"は"依存症"を促進するからだ。</p> <p>カジノ先進国である米国や韓国などの研究では、<u>依存症患者が増え、その対策で社会が負担するコストは、カジノによる税収増分を上回るという。一方で、カジノ解禁によるギャンブル依存症の増加や周辺地域の治安悪化などが懸念されています。シンガポール政府の対策機関がカジノ開業(2010年)をはさむ08年と11年に行った調査では、「依存症」とされる病的なギャンブル依存の割合は、08年が1.2%、11年が1.4%と、ほぼ横ばいだった。</u></p>	<p>"依存症"の観点で論題に反対する。"カジノ"は"依存症"を促進するからだ。</p> <p>これ、韓国の例でございまして、韓国でカジノを解禁して依存症、中毒が大変増えたということでございます。一般論として申し上げれば、<u>カジノ合法化については、ギャンブル依存症や多重債務に陥った人への対策のほか、治安や青少年への影響といった負の側面への対策等を考慮する必要があると考えております。全国各地でカジノ誘致反対あるいは賭博解禁駄目だという声広がっておりますが、資料一枚目御覧になっていただければ、これ八月の厚労省研究班の報告によりますと、日本でギャンブル依存の疑いのある人は男性八・八%、女性一・八%、全体で四・八%、人数にして推計五百三十六万人に上っております。</u></p>

### 3.1 価値決定

価値決定とは、入力された議題について、どの価値に基づいて意見生成するかを決定することである。価値決定は、予め用意された価値の一覧から最大6つの価値を選択する。ディベート人工知能は価値オントロジを備え、価値の一覧はオントロジから提供される[柳井 15]。それぞれの価値はその価値の同義語やインスタンス語の集合により定義される。たとえば病気という価値は“病気”，“病”，“がん”，“依存症”などで構成される。

手法はデータソースにおける共起頻度に基づく。まず論題から議論の対象となる名詞句(カジノなど)が抽出される。次に議論の対象と各価値の共起回数を調べ、共起回数の大きな価値が選択される。このとき、共起文に対してそれぞれ賛否推定を行い、入力された賛否と一致する文のみを数に数える。賛否推定は、文に含まれる述語を中心とした句構造構文木があらかじめ用意した構文パターンに当てはまるか否かなどの情報を元にルールベースで行う[佐藤 16a]。

### 3.2 文の根拠性認識

根拠性認識は、データソースから抽出したテキストが、主張を支持(否定)する根拠としてどの程度もってもらいしかを推定する問題である。現在の実装では、データソースから抽出したテキストが、議論対象が価値を増大(減少)させる関係を支持するか否かを推定する問題であるとしている[佐藤 16b]。たとえば、カジノ合法化の議題に反対する意見生成の際には、カジノが病気を増加させるという関係を支持する文、たとえば「ギャンブルは依存症の問題を引き起こす」という内容の文を根拠とする。根拠性認識は主張を仮説とした場合の含意関係認識とみなすことがで

き、含意関係認識を意見文章生成に必要な問題に限定して解いていると言える。

根拠性認識は、信頼度付き分類問題を教師あり学習として解く。議題語と価値が共起する文を候補文として、賛否の分類結果および根拠か否かの分類結果と信頼度を出力する。議論対象と価値の関係に着目する問題であるため、候補文内のどの語句が議論対象および価値と対応するかに着目したモデル化を行っている。具体的なモデルについては[柳瀬 16]にて説明している。

## 4. 国会会議録を用いた意見生成

本発表では、国会会議録に記載されている国会での発言をデータソースに用いた意見生成システムを展示する。本章では、国会会議録から生成した意見と毎日新聞記事から生成した意見とを比較する主観評価結果を報告する。

### 4.1 国会会議録への適用：質問文の除外

国会会議録に記録されている会議は大部分が質疑の形式で進められる。まず質問者が質問文章を読みあげ、指名された回答者が回答を行う。したがって国会会議録には質問文が多く含まれているが、質問文は意見の根拠にはなりにくいと考えられる。そこで、「お伺いいたします」「お尋ねでございます」といった質問文末表現を手作業で7個収集し、これらの表現を文末にもつものは根拠性認識結果が0であるとして生成文章には含まれないようにした。

\*2 CD-毎日新聞データ集

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

## 4.2 実験

生成された意見の説得力に関する主観評価を行った。比較のため、同一の論題に対して毎日新聞コーパス\*2 を用いて意見を生成した。実験では、国会会議録・毎日新聞共に、2011年1月24日から2016年8月3日までの記録・記事を用いた。前処理として、句点区切りのルールによる文区切りを行った。文区切り後の文の数は、国会会議録が3,300,003文、毎日新聞が9,346,561文であった。

出力は、1論題につき賛否ごとに最大6つの意見が生成され、1意見は最大8文により構成される。

論題には以下の3種類を用いた。

- カジノを合法化すべきか
- 捕鯨を継続すべきか
- 消費税増税に賛成すべきか

## 4.3 結果

表2に、生成された意見文章中で、論題の肯定側/否定側の意見や根拠として正しいと判断された文の割合を示す。国会会議録を用いた場合にも、毎日新聞を用いた場合と分量・精度ともにほぼ同程度となっている。本システムは新聞記事向けに開発したものであるが、データを会議録に変更した場合でも定量的に同程度の精度の意見文章が生成されることが確認された。

以下では、生成された意見文章を読み比べた際の違いについて述べる。

まず、同じ論題と賛否であっても、論点として選ばれる価値が異なることがある。これは、文書集合によって共起頻度が異なることによって起こる。国会会議録から生成された36意見中、23意見が毎日新聞から生成された意見と同じ価値、13意見が異なる価値であった。たとえば国会会議録では、カジノ合法化に賛成の立場の意見として「活性化」の価値が挙げられた。表1に出力結果を示している。ここでは、地方経済の活性化、東日本大震災の復興や地方自治体の財政再建が賛成の理由として述べられている。この価値は毎日新聞を用いた際には上から12番目と低順位であり、最終出力には含まれなかった。なお、紙面の都合上、表1では実際に出力された意見の前半4文のみを掲載している。

同じ論題に対して同じ賛否の立場から同じ価値に基づいて生成された意見を比較すると、内容は似通っており、同一の事例が挙げられる場合が多くあった。たとえば、「捕鯨」に「賛成」する「伝統」観点からの意見では、捕鯨を通じて伝統的な漁法の継承が行われているという事例が共通して述べられた。一方で、国会会議録利用時のみ発見された事例もわずかではあるが存在した。消費税増税に反対する理由として抽出された「一九九七年の消費税増税を引き金とした大不況で税収が落ち込んだ」という事例は、毎日新聞から抽出した議題語と価値語が共起した候補文には含まれなかった。

表2. 生成された意見における根拠性認識の正答文割合

議論の対象 (平均)	賛否	国会会議録 52% (264文)	毎日新聞 46% (272文)
カジノ	賛成	49%	50%
	反対	73%	68%
捕鯨	賛成	38%	47%
	反対	61%	34%
消費税増税	賛成	27%	52%
	反対	69%	27%

このように、異なるデータソースからの文書を利用することにより、システムに大きな変更を加えることなく、新しい発見を得ることができることが確認された。一方で問題点としては、表1にも表れているように、国会会議録から生成される意見の方が文書量が大きく冗長になる傾向にある。これは元々書き言葉である新聞に比べて、会議での発言の書き起こしである会議録は文が長い傾向にあり、不要な情報を含むことも多いためである。文単位の要約を行うなどにより、より可読性の高い簡潔な意見文が出力できると考えられる。

## 5. おわりに

本稿ではディベート人工知能における意見生成について、その主要技術である価値選択や根拠性認識についての現状を述べ、国会会議録をデータソースとした場合の出力結果について評価を行った。国会会議録に対してディベート人工知能を用いて与えられた論題について意見を生成することにより、国会会議録から意思決定を助ける情報を抽出できることが確認された。国会会議録を元に生成された意見は、過去の答弁の観点別まとめと捉えることもできるため、過去の発言を参照する作業の多い国会答弁作成業務などの実務者の補助にもなると考えている。

今後は、地方議会会議録もデータソースとして利用することで、より幅広い発見が得られることが期待される。また、同じ論題、同じ価値が選ばれた際の意見文章が国会会議録と毎日新聞とでほぼ同じような内容となった点について、データソースの内容が似通っていたのか、あるいは手法により抽出される根拠の偏りによるものであるのかを明らかにすることは今後の課題である。

## 参考文献

- [尾崎 14] 尾崎 正宗, 掛谷 英紀: 国会会議録の主張文取り出しおよびその要約, 言語処理学会第20回年次大会, pp. 504-507, 2014.
- [金丸 05] 金丸 浩司, 西崎 博光, 関口 芳廣: 文間の類似性を用いた国会会議録のトピック別要約の検討, 第67回全国大会講演論文集, pp. 439-440, 情報処理学会, 2005.
- [川端 07] 川端 正法, 山本 和英: 話題の継続に着目した国会会議録要約, 言語処理学会第13回年次大会, pp. 696-699, 2007.
- [佐藤 16a] 佐藤 美沙, 柳井 孝介, 柳瀬 利彦, 是枝 祐太, 丹羽 芳樹: ディベート人工知能における影響関係認識のためのテキスト内の論理構造に関する考察, 人工知能学会全国大会論文集, 第30巻, 4B1-3, pp.1-3, 人工知能学会, 2016.
- [佐藤 16b] 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, 人工知能学会論文集, 第31巻, no. 6, pp. AI30-L\_1-12, 人工知能学会, 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, 3M3-2in, pp. 1-3, 人工知能学会, 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 Attention 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, 第30巻, 1A4-PS-27b-5in1, pp.1-4, 人工知能学会, 2016.
- [山本 05] 山本 和英, 安達 康: 国会会議録を対象とする話し言葉要約, 自然言語処理, 第12巻, No. 1, pp. 51-78, 2005.

---

[山本 11] 山本 冴里:国会における日本語教育関係議論の  
アクターと論点 ―国会会議録の計量テキスト分析からの概観  
―, 日本語教育 第 149 巻, pp. 1-15, 2011.