

E-012

## 政治情報サイトの政策表現文抽出・整形による政党間比較 Policy Comparison by Extraction of Manifest Sentences from Political Information Websites

永井 茅希<sup>†</sup>山田 剛一<sup>†</sup>絹川 博之<sup>†</sup>

Kayaki Nagai

Koichi Yamada

Hiroshi Kinukawa

### 1 はじめに

投票を行う際に重要な情報源となる政党の公式サイトは、政策が政策課題ごとに整理されておらず分かりにくいという問題がある。

そのため、各政党の政策課題についての文章を整理し、政党間の比較を行うことで、政策を理解しやすくすることを目的としたシステムを開発する必要がある。その目的を達成するためには、(1) 政策課題の発見、(2) 政策を表す文の抽出、(3) 政策比較に用いる文の絞り込み、(4) 文の縮約が必要である。

本研究では、政策比較に用いる文の絞り込み方式について検討を行う。

### 2 政策比較システム

本研究では、各政党の政策課題についての文章を整理し、政党間の比較を行うシステムを開発する。図1にシステムの全体図を示す。

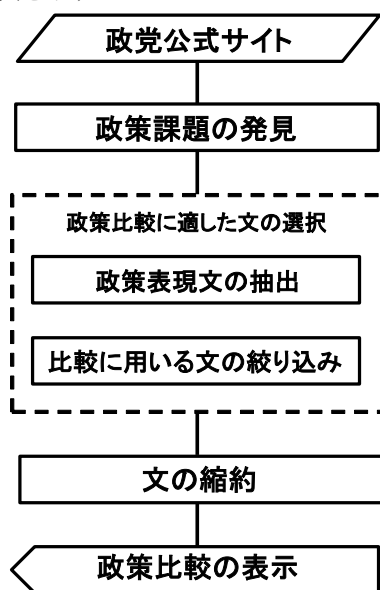


図1 政策比較システムの流れ

#### 2.1 政党公式サイトと政策表現文

政党公式サイトでは、日々の活動や、政党の方針などを有権者に向けて発信している。しかしながら、総選挙・無投票者調査[1]の今後の選挙に希望する改善点において、政策の伝え方が上位になっていることなどから、有権者は現状の政党の政策の伝え方に不満を抱いていることが窺える。

そこで我々は、政党の公式サイトから、政策表現文を抽出する機能の開発を行った[2]。

#### 2.2 政策課題の発見

政党の政策は「集団的自衛権」「待機児童」といった政策課題について記述されている。これらの政策課題ごとに政党間の比較を行うためには、まず政策課題のリストを生成する必要がある。その上で各政策課題に対応する政策表現文を抽出する。

#### 2.3 政策表現文の定義

本研究では、以下の文を政策表現文と定義した。

- (1) 政党の政策
- (2) 政策に付随する意見
- (3) 政策に付随する意見の引用

意見の引用に関しては、党の役員や専門家の発言を引用する形で政策を示すことがあるため、政策を理解するための参考になると考え、定義に加えた。

#### 2.4 政策表現文の抽出結果

政党の公式サイトから、政策表現文を抽出する実験を実施した結果、91.0% というF値を得た[2]。本研究では、政策表現文の中から、さらに政策比較に用いる文の絞り込みを行う。

#### 2.5 政策比較に適した政策表現文の特徴

政策比較に適した文とは、政策課題に対する政党の立場が示されている文である。そのような文の特徴を以下に示す。

- (1) 注目されている政策課題は、与党が賛成、野党が反対という構図になりやすいことから、与党の文にはポジティブな語、野党の文にはネガティブな語が出現しやすい。

なお、語のポジティブ・ネガティブの判別には、日本語評価極性辞書[3]を利用している。

- (2) 政策課題名の係り元や係り先に「アベノミクスが経済再生への最善策である」「格差拡大のアベノミクスストップ」などポジティブ・ネガティブを意味する語が出現する。ただし、「サービスの不足」などポジティブな語がネガティブな語にかかる場合はネガティブ、「問題の解決」などネガティブな語がポジティブな語に係る場合はポジティブとみなす。

- (3) 政策課題が出現する文の述語や述語に係る文節に「推進」「反対」のようにポジティブ・ネガティブを意味する語が出現する。ただし、述語が「～と指摘」や「～と主張」などの場合、前の文節を述語とみなす。

<sup>†</sup> 東京電機大学大学院 未来科学研究科  
Graduate School of Science and Technology for Future Life,  
Tokyo Denki University

## 2.6 日本語評価極性辞書の修正

日本語評価極性辞書において、ポジティブ・ネガティブとされていない語であっても、政策表現文においてポジティブ・ネガティブを意味する語があるため、人手にて修正を行う。修正の例を表1に示す。

表1 日本語評価極性辞書の修正例

	修正前	修正後
撤回	ニュートラル	ネガティブ
推進	ニュートラル	ポジティブ
実行	ニュートラル	ポジティブ

## 2.7 政策比較に適した政策表現文の絞り込み

2.5に示した政策比較に適した文の特徴にもとづいて文の絞り込みを行う。ポジティブ・ネガティブな語の出現頻度、係り受け関係をもとに各文に対して重要度を付与する。重要度が閾値以上のものを政策比較に適した政策表現文とする。その定義式を以下に示す。

$$W_1 = 2.5の(2)に合致する語の数$$

$$W_2 = 2.5の(3)に合致する語の数$$

ただし $W_1$ に含まれる語は除く

$$W_3 = \text{与党はポジティブ, 野党はネガティブな語数}$$

ただし $W_1, W_2$ に含まれる語は除く

$$\text{重要度} = \alpha W_1 + \beta W_2 + \gamma W_3$$

$\alpha, \beta, \gamma$  は重要度係数

## 4 評価実験と考察

政策比較に適した文の絞り込みに関する評価実験に用いたデータおよび実験結果を示す。

### 4.1 評価データ

実験に用いるデータとして、人手で抽出した政策表現文1,535文を用いる。当該データは、(1) TPP, (2) アベノミクス, (3) 憲法改正, (4) 原発, (5) 待機児童, (6) 集团的自衛権, (7) 消費税, 以上人手にて選定した7つの政策課題をもとに、自由民主党, 民進党, 日本共産党の公式サイトから取得および抽出した文である。当該データの取得は2016年5月4日に行い、その際、民進党の公式サイトは設置されて間もなく、十分なデータが取得できないと判断したため、民主党のサイトを利用した。

これらを対象に政策比較に適した文の絞り込み実験を行う。なお、文の選択の際に必要な係り受け解析については CaboCha[4]を用いる。予備実験の結果をもとにして閾値 = 1.2,  $\alpha = 2$ ,  $\beta = 0.5$ ,  $\gamma = 0.1$  に設定する。各政策表現文が政策比較に適した文かどうかは、人手にて判断を行った。

### 4.2 評価指標

精度, 再現率, F値により評価した。定義式は以下の通りである。

$$\text{精度} = \frac{\text{システムが正しく抽出した文数}}{\text{システムが抽出した文数}}$$

$$\text{再現率} = \frac{\text{システムが正しく抽出した文数}}{\text{政策比較に適した文数}}$$

$$F\text{値} = \frac{2 \times \text{精度} \times \text{再現率}}{\text{精度} + \text{再現率}}$$

### 4.3 評価実験結果

評価実験の結果を表2に示す。

表2 政策比較に適した文の絞り込み評価実験結果

	自民党	民進党	共産党	平均
精度	66.6%	62.5%	73.3%	67.5%
再現率	70.6%	65.6%	62.1%	66.1%
F値	68.5%	64.1%	67.2%	66.6%

### 4.4 考察

本研究では争点となる政策課題をシステムに与えたため、与野党でポジティブ・ネガティブが顕著に分かれる例が多かった。しかしながら、そうならない政策課題も存在する。そのため、抽出した文にポジティブな語が多い傾向があれば肯定的なスタンス、ネガティブな語が多ければ否定的なスタンスとみなすなどの対応策を検討する必要がある。

## 5 おわりに

本研究では、複数の政策表現文の中から、政策比較に適した文を選択する方式の検討を行った。その結果、3党平均で 66.6% というF値を得た。今後の課題として、与党が賛成、野党が反対という構図にならない政策課題の対応、選択した文の縮約、政策課題の自動取得などが挙げられる。

**謝辞** 使用させて頂いた日本語評価極性辞書, CaboChaの開発者の方々に深く感謝致します。

### 参考文献

- [1] ㈱ジャパン・マーケティング・エージェンシー, 総選挙・無投票者調査, <http://www.jma-net.com/reports/総選挙・無投票者調査データの公開>, 2012
- [2] 永井茅希, 山田剛一, 絹川博之, “政治情報サイトからの政策表現文の抽出方法とその評価,” 情報処理学会第79回全国大会講演論文集, No.2, pp.523-524, 2016.
- [3] 東山昌彦, 乾健太郎, 松本裕治, “述語の選択選好性に着目した名詞評価極性の獲得,” 言語処理学会第14回年次大会論文集, pp.584-587, 2008.
- [4] 工藤拓, 松本裕治, “チャンキングの段階適用による日本語係り受け解析,” 情報処理学会論文誌, Vol.43, No.6, pp.1834-1842, 2002.