

国会会議録の主張文取り出しおよびその要約

尾崎 正宗 掛谷 英紀

筑波大学 システム情報工学研究科

s1320805@u.tsukuba.ac.jp kake@esys.tsukuba.ac.jp

1 はじめに

近年、日本の国政選挙では、各党が捧げるマニフェスト（政権公約）をマスメディアが取り上げ、それに基づいて投票先を選ぶことが良いという風潮がもてはやされている。日本国内では、有権者の多くは、日常的には政治に気を配らない。このような状況の中で、各党の思考や政策がまとめられているマニフェストを基に投票先を選ぶのは、合理的な判断であると考えられる。

しかし、マニフェストに基づく投票先の判断には問題もある。2009年8月31日の第45回衆議院選挙はマニフェストが最も注目された選挙であったが、結果として選挙で選ばれた政権与党はマニフェストをほとんど守らなかった。

また、マニフェストは政党が掲げているものであるため、必ずしも個々の議員の意見と一致しているとはいえない。また、選挙時は国民の支持をえるために、耳触りのいいことだけを語る議員も多い。

本来、国会議員の普段の言動に基づいて誰に投票するかを決めるのが理想的である。しかし、現実にはそうした情報を短期間で取得するのは難しい。そこで、本研究では、国会議員の普段の言動を手短かに知る方法として、投票の参考になるような形で、国会会議録を自動要約するシステムを実現することを目的とする。

2 手法

2.1 意見文の抽出

国会は、質問者が質問したことに対して答弁者が回答する形式になっている。質問者、答弁者ともに、自分の主張である意見文の裏付けを行うために、多くの事実文が混在する。国会議員の言動を知るためには、まず意見文の抽出が必要である。

日本語の特徴として、意見文を発言する際には、文尾に「思います」「思っております」・・・など、文尾の分節に「思う」が入っている可能性が高い。本研究では、簡易的に意見文を抽出するために、cabocha[1]により、文尾分節を取り出し、原形に「思う」を含んでいる文を意見文として抽出した。

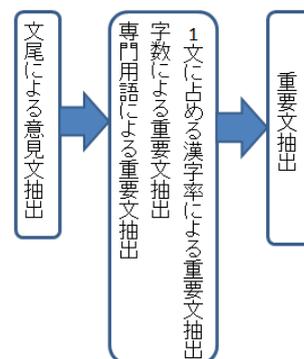


図1 システム概要図

2.2 意見文の要約

文尾情報による意見文抽出により、文章は全体の約20%~25%にすることができる。

意見文を抽出するだけでは、国会会議録を十分に要約できたとは言い難い。そこで、意見文のなかでも、特に重要であるものをさらに抽出する必要がある。

2.3 絶対評価による意見文の要約

以下に示す方法により、専門用語が含まれるか否か、文の長さ、漢字含有率などを基準として、絶対評価で意見文を要約した。

2.3.1 文長による分類

国会会議録の文章は、国会での発言を記録しており、日本語の会話文は1文が長くなりやすいという傾向がある。文長が長いものは、要約文としては使用しにくい。しかし、文長が長いものでも、1部分をみると要約文として成立するものもある。

国会会議録第172回～第181回までの安全保障委員会で文長の平均と標準誤差を算出したところ、平均は約87語、標準誤差は約53語であった。よって、140語より多いものは、特に文長が長いものと考えられる。これは、twitter社のツイートの上限とも偶然一致する。そこで、本研究では、140字より長い文章に対しては、読点で区切った。

2.3.2 専門用語辞書

国会は、本会議と委員会に分かれており、委員会は専門的な分野について議論する場である。意見文の中でも、特に重要度が高い文は、その委員会特有の重要語が含まれている可能性が高い。専門用語を抽出する方法として、東京大学で開発された専門用語自動抽出用モジュールである”TermExtract” [2]を使用し、専門用語が含

まれている文のみを抽出した。

2.3.3 漢字率による分類

重要度の高い文章は、一文に占める漢字が多い文章が高い文章である可能性が高い。国会会議録において、一文に占める漢字率の平均は27.3%であり、この数値より大きいものを抽出することとした。

2.3.4 その他の処理

「聞く」「伺う」などの言葉が入っている文は、意見文でなく、質問文の可能性が高いため、排除した。また、「そういう」などの言葉が入っている文は、前の文章とのつながりが必要であるため、排除した。逆に、「べき」など強い主張をさす言葉が入っている文は要約文に残す処理をした。

2.4 相対評価による意見文の要約

絶対評価による意見文の要約は、議員の主張をつかむことが可能であるが、議員によって文章が残る分量が異なり、国政選挙時に各議員の意見を比較する材料としては不十分である。そこで、絶対評価によって不必要な部分を切り捨てるのではなく、切り捨ててもいい要素が強いものにコストを割り当て、コスト最小のものから順に、400字以内に収まるように要約した。

2.4.1 文長によるコスト生成

先程の手法と同様に、140字以上の文は、読点で区切った。

140字以下の文に対しては、文がある程度短いものが重要文とされる可能性が高い。しかし、逆に短すぎる文は意味のある意見を含まない可能性が高い。よって、あ

る程度字数が少ないもののコストが低くなるような設定が望ましいと考えられる。そこで、文毎に文長の標準偏差(u)を算出し、標準偏差が 40 付近のコスト(C)を少なくするために、

$$C = |40 - u|$$

とした。このように設定することで、字数が極端に少ないものには高コストを付与することができる。

140 字以上のものについては、読点で区切ってるので、区切られた後に生成される文は短い。文長の平均が約 17.7 語、標準偏差が約 12.4 語となっている。そのため、ある程度長めの文のコストを低く設定した方が重要な意見文が抽出できると考えられる。そのため、コストを

$$C = |60 - u|$$

とした。

2.4.2 専門用語辞書

専門用語を含む文章はできるだけ残したいので、負のコストを割り当てる。具体的には、文中に専門用語が 1 つでもある場合、一律に -20 のコストを付与した。

2.4.3 漢字率によるコスト生成

漢字率は、ある程度高いもののほうが、重要な意見文である可能性が高い。そこで、漢字率の標準偏差(u)を算出し、コストを

$$C = |60 - u|$$

と設定した。

3 実験

3.1 使用データ

本研究での使用データは、国会会議録第 172 回～第 181 回までの期間の安全保

障委員会のデータである。この期間は、民主党政権が誕生してから解散するまでの期間である。本研究は、第 46 回衆議院総選挙での投票支援を想定して行ったものである。

3.2 専門用語辞書

今回使用した専門用語は、表 1 の通りである。TermExtract で抽出した用語のうち、安全保障委員会に関係ある用語を 15 語抽出した。

表 1 専門用語辞書

日本	防衛	米国
自衛隊	中国	自衛官
国民	日米	隊員
外務大臣	安全保障	外交
政府	沖縄	米軍

3.3 結果

上記委員会の中で、赤澤議員の要約文を表 2 に示す。要約前の原文は 4888 文字である。

表 2 赤澤議員の発言要約文

(絶対評価による要約 : 1091 文字)

特に、福島県に災害派遣出動中の自衛隊員の皆様は、福島原発事故の影響で被曝の危険とも闘っておられるということで、本当に頭の下がる思いでございます。そういったことも踏まえて、大臣にお尋ねをしたいのは、就任後、自衛隊員との信頼関係を強化し、自衛隊員の士気を高めるためにどのようなことを心がけ、どのようなことを行われてきたのか、お話をいただきたいと思います。おっしゃるとおりで、すぐにでもできることは、単純ではあるけれども、常に隊員のことを気にかけること、そしてできる限り現場に足を運ぶこと、さらには、隊員の置かれている状況をつぶさに見て、頭だけではなくて体で理解すること、さ

らには言葉をかけて激励することではないかと私は思っております。稚内から沖縄まで足を運ばれたということは多とするものでありますし。あわせて、大臣も足を運んでおられるようですが、政務三役も、お忙しいとは思いますが、都合がつくのであれば、本当に、隊員の食事や、さらには寝袋とか装備を気にしていただきたいと思っております。大変お若いようにも思いますが、私自身は、この二人が死亡されたことについて、このたびの任務が隊員の心身に非常に大きな負担をかけている証左ではないかと思うんです。本当に隊員も過酷な状況に置かれていると思うので、ぜひその点、メンタルケアも含めて、万全ということはないという思いで常に気にかけていただくということが本当に大事なことかなと私は思っております。隊員が何人も被曝をするというような事態は極力避けなきゃいけないと思うんですね。自衛隊の装備でいうと。各隊員は。財政上の制約がなければ、リアルタイムで計測できるものを一人一個持たせたいということなのかなとも思いますが、その辺は、この事態を受けて、今のところ隊員の健康は保たれている、重大事は起きていないということですが、将来の備えとして、やはりきちっと検討していただきたいなと思います。それから次に、食事の風景というのを私は見せていただいたんですけども、端的に言って、若い隊員にはとても足りるとは思えない量の食事しか与えられていないと思いました。隊員が少しでも元気で。「事に臨んでは危険を顧みず、身をもつて責務の完遂に務め、もつて国民の負託にこたえる」というサービスの宣誓をされて仕事をされているわけですから、全身全霊を挙げて職務に邁進している隊員に、ぜひ、今申し上げたような観点から配慮を特段お願いしておきたいと思っております。さらには隊員との信頼関係や士気を上げるための努力がないと。隊員がどんなに厳

しい訓練に耐えても。我が国の独立と平和や国民の安全。

(相対評価による要約： 343 文字)

特に、福島県に災害派遣出動中の自衛隊員の皆様は、福島原発事故の影響で被曝の危険とも闘っておられるということで、本当に頭の下がる思いでございます。そういったことも踏まえて、大臣にお尋ねをしたいのは、就任後、自衛隊員との信頼関係を強化し、自衛隊員の士気を高めるためにどのようなことを心がけ、どのようなことを行われてきたのか、お話をいただきたいと思っております。稚内から沖縄まで足を運ばれたということは多とするものでありますし。隊員が何人も被曝をするというような事態は極力避けなきゃいけないと思うんですね。勲章は上げるけれども経済的に報いるというようなものかどうかはよくわからないというのが現時点でのお答えだったと私は理解をしております。さらには隊員との信頼関係や士気を上げるための努力がないと。

4 まとめ

国会会議録の要約を、絶対評価によるものと、相対評価によるものの2手法で行った。絶対評価によるものは、文の繋がりがあある、相対評価によるものは字数制限があるため、手短かに読むことができるというメリットがある。

5 参考文献

- [1] Taku Kudoh, Yuji Matsumoto (2000) Japanese Dependency Analysis Based on Support Vector Machines, EMNLP/VLC 2000
- [2] 東京大学情報基盤センター・中川裕志, 東京大学経済学部・前田朗, 専門用語 (キーワード) 自動抽出用 Perl モジュール"TermExtract"
<http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html>