**国会会議録に基づく自民系/民主系大臣の特徴分析**

**樋口心 掛谷英紀**

[**S1511175@u.tsukuba.ac.jp**](mailto:S1511175@u.tsukuba.ac.jp)[**kake@iit.tsukuba.ac.jp**](mailto:kake@iit.tsukuba.ac.jp)

**筑波大学**

**概要**本論文ではWeb上で公開されている国会会議録を用いて政治家として正しい議論を行えない政党/人物の特徴を見出すことを目的とする。昨今野党により問題になっている国会答弁内の「ヤジ」について注目し、予算委員会での民主党政権時代の野党と現在の野党の発言の分類を試みる。

具体的にはナイーブベイズ法を用いた機械学習、また両カテゴリの使用頻度の偏りから特徴語の抽出を行う。

使用頻度の偏りから、現在の野党は自分のマニフェストではなく相手の言質を取るような発言が多いという傾向が分かった。また、末尾表現に着目した機械学習の分類は高い精度を誇り、政治テーマに依らずその個人/政党による特徴を抽出することに有効であることが示された

**１はじめに**

**２．研究手法**

国会会議録を対象とした自然言語処理に関する研究はいくつかあるが[]、本研究では国会会議録に収められている予算委員会の答弁を用い、民主党政権時代の野党と現在の野党の言葉遣いを判別し、有意義な議論を行う党を判別するシステムの構築を試みる。

　まず、国会会議録検索システム[]から国会会議録の答弁の文書データを、形態素解析ツールMeCab[]を用いて形態素ごとに分割する。

その後、読点の直前3語または４語を末尾表現と定義し、表出形をそのまま機械学習の素性として使用する。

また、動詞・形容詞・名詞なども追実験の対象とした

このように抽出した素性をもとに、学習データおよびテストデータを作成する。その後、学習データに対して、各カテゴリ間の特徴を機械学習し、テストデータを使用して開発したシステムの精度を算出する。制度を算出する際にはクロスバリデーションを用いる。

以上の手順を図1に示す。

**3．取得データ**

**3.1国務大臣**

民主系野党と自民系野党を比較するために、

民主系の野党は第172回国会から第181回国会

2009/9/16~2012/12/25,

自民系の野党は2015/4/16~2018/7/31

の3年間の言語データを収集した。上記の期間に衆議院予算委員会に参加していた大臣のリストを以下に示す。



また、今回の実験では各カテゴリの人数・データ量が不均衡であると、単語出現比率や機械学習の結果に悪影響を及ぼしてしまう。

以下の大臣は、任期中であったがデータ量が著しく多い、若しくは小さいという理由で取り除いた。



**3.2国会会議録**

本研究では言語データを収集するために、国会会議録検索システム検索用APIを利用して分析対象の大臣の発言を収集する。

先述の通り、民主系および自民系の過去の国会における発言から各カテゴリの特徴分析を試みる。

各大臣の発言の収集対象は以下のように設定する。

両カテゴリの「ヤジ」などの適切でない表現を確認するために衆議院予算委員会の発言のみを収集する。

これは、予算委員会では内閣の実際の政策など話題が偏らず話されているうえに、質疑応答で野党の発言する機会が多いからこのように設定した。

国会会議録は図2のようになっているため、「〇 役職（氏名）」の部分を取り除き、それ以降を　その人物の発言として抽出する。

以上の条件で抽出すると、取得発言数の合計は

旧野党:件

現野党:件

となった。これらのデータをもとに学習データを作成し、機械学習を行う。

**4．実験**

**4.1データセット**

3.2節で収集した発言を表3に示すように

大臣毎に発言をまとめ、それに基づいて機械学習を行う。

機械学習を行う際は各カテゴリの発言件数の差が大きいと発言件数が多いカテゴリに判定される確率が大きくなる可能性がある。

そこで本研究では

また、クロスバリデーションはリーブワンアウト法で行う。

**4.2機械学習**

上記のデータセットに基づき、最大エントロピー法による機械学習を行った結果を表4に示す。

正解率が％となり、末尾表現で、旧野党と現野党を言い当てることができる。つまり両カテゴリの末尾表現の違いがあることがわかる。

**4.3単語出現頻度**

今回の実験では各カテゴリ、大量の言語データを収集した。このような場合、単語の出現頻度も各カテゴリの特徴を得られる重要な情報元となる。

そこで両カテゴリの合算で出現頻度が大きくことなる素性からカテゴリによって出現頻度が大きく異なる素性を抽出する。

一部の大臣が口癖として同じような末尾表現を使っていることによる特徴を抽出してしまわないように収集した大臣56人のうち30人以上が使用している末尾表現のみ抽出した。

また末尾表現だけでなく、形容詞も政治の話題に依らず、各カテゴリの特徴をつかむことができると考え同様に表に示した。

**5．終わりに**

**参考文献**