

# 都道府県議会を対象とした 議員の発言と選挙結果の関係性の一考察

A Discussion of the Relation between Assembly Member's Utterance and Next Election Outcome

木村泰知 \*<sup>1</sup>  
Yasutomo KIMURA

川浦昭彦 \*<sup>2</sup>  
Akihiko KAWAURA

内田ゆず \*<sup>3</sup>  
Yuzu UCHIDA

小樽商科大学 \*<sup>1</sup>  
Otaru University of Commerce

同志社大学 \*<sup>2</sup>  
Doshisha University

北海学園大学 \*<sup>3</sup>  
Hokkai-Gakuen University

We created the local assembly minutes corpus, which connected the politicians' profile, the results of elections and politicians' utterances. This corpus includes minutes of plenary assemblies in 47 prefectures of Japan from April 2011 to March 2015. This paper shows that a politicians' utterance in the assembly makes an effect for next election in order to predict next election results by Machine Learning. Therefore we tried to clarify characteristic words between the election winners and election losers in the assembly elections.

## 1. はじめに

近年、政治家は、インターネットを介して、政治理念、政策、政治活動、プロフィールなどを伝えている。このような政治家の情報は、どの程度、次回選挙に影響を与えるのだろうか。選挙結果に影響を与える要因には、当選回数、学歴、政党、年齢、性別、職歴などがある。他にも、政治家がどのようなマニフェストを書いたのか、どのような発言をしたのか、どのような内容の投稿をしたのかのように、発信している情報の内容も影響する。特に、2013年に公職選挙法が改正されたことにより、インターネット選挙運動が解禁され、電子的なテキストを対象とした選挙に関する研究が盛んになっている [2]。インターネット選挙運動では、twitter が注目されており、2013年参議院選挙におけるソーシャルメディアの分析として、つぶやきを対象とした分析の研究がある [3]。また、鳥海らによる 2017 年衆議院選挙における政党公式アカウントフォロワーを分析している研究がある [4]。このように、国政選挙におけるインターネット選挙運動を対象とした研究は進んでいるが、日本の地方議会選挙においては、地方議員の投稿量が少ないことから、国会議員の選挙と同じように分析することが困難である。

そこで、我々は、都道府県議会における発言を対象として、議会の発言と選挙結果との関係に焦点を当てて研究を行う。議会における議員の発言を分析するには、ウェブ上に公開されている議会会議録から、議員の発言を収集し、いつ、どこで、それが、なにを発言したのかを整理するため、膨大な時間とコストがかかる。そのため、大規模なテキスト分析はほとんど行われていなかったが、最近、地方議会会議録コーパスが整備されつつあることから、川浦らが議会における「発言量」と選挙結果の関係について、計量経済学の観点から研究を進めている [1]。しかしながら、地方議会における議員の「発言内容」と選挙の結果についての研究は進められていない。そこで、本研究では、47 都道府県の地方議会会議録コーパスと選挙結果を結びつけることにより、都道府県議会における各議員の発言内容（発言単語）が、次回選挙において当選予測を行う場合に有効な素性となるのかについての調査を行う。我々は、以前から、都道府県議会の平成 23 年 4 月から平成 27 年 3 月までの会議録を対象として、発言者の表記ゆれを人手で対応づけたコー

パスを作成しており、正確な分析を行うことができる。本研究では、議会における発言内容に含まれる単語の出現頻度から、当選者と落選者に特徴的な単語を明らかにする。また、特徴的な単語を用いて、機械学習による選挙予測がどの程度可能なのか試みる。

## 2. データセットの構築

本章では、会議録、議員プロフィール、当選情報に含まれるデータから、発言内容と選挙結果（当落）を結びつけるデータセットの構築方法について述べる。

### 2.1 対象の会議録と選挙結果

対象の会議録は、47 都道府県の地方議会会議録の本会議とする。議会会議録の対象期間は、第 18 回統一地方選挙（2015 年 4 月）を対象とするため、2011 年 4 月から 2015 年 3 月までの 1 任期 4 年とする。第 18 回統一地方選挙は、岩手県、宮城県、福島県、茨城県、東京都、沖縄県を除く 41 道府県議会において、2015 年（平成 27 年）4 月 12 日と 26 日の 2 回に分けて行われた。統一地方選挙は、議会の解散や自然災害の影響により、選挙時期がずれることがある。例えば、岩手県、宮城県、福島県は、東日本大震災の影響により、第 17 回の地方統一選挙が通常よりも、遅れて実施された。そこで、本研究では、統一地方選挙の時期と異なる自治体を除き、41 道府県議会を対象とする。

### 2.2 対象の立候補者

本節では、分析対象者について述べる。分析対象者は、2015 年 4 月の統一地方選挙に立候補した議員とする。また、本研究では、議会の発言を利用することから、選挙前に議員であることが前提となり、全ての議員を対象とすることができない。さらに、対象の自治体の場合でも、議長および副議長の議員の発言は、発言内容が質問ではなく議事進行となることから対象外とする。そこで、分析対象の立候補者の選定条件を以下のよう

1. 2015 年 4 月の統一地方選挙に立候補した議員
2. 現職、新人、元職のうち、現職の議員
3. 議長、副議長を務めていない議員

連絡先: 木村泰知, 小樽商科大学, 〒 047-8501 北海道小樽市  
緑 3 丁目 5-2 1, kimura@res.otaru-uc.ac.jp

表 1: 3つのデータ構造と連結キー

| 番号 | 会議録コーパス | 議員リスト | 選挙結果    |
|----|---------|-------|---------|
| 0  | 識別子     | 議員 ID | (異なる)ID |
| 1  | 都道府県名   | 都道府県名 | 都道府県名   |
| 2  | 回       | 自治体名  | 選挙名     |
| 3  | 号       | 姓名    | 選挙区     |
| 4  | 年       | ふりがな  | 当選フラグ   |
| 5  | 月       | 対象行政区 | 得票数     |
| 6  | 日       | 生年    | 姓名      |
| 7  | 開催期間    | 性別    | ふりがな    |
| 8  | 表題      | 職名    | 党派      |
| 9  | 発言者名表層  |       | 新旧      |
| 10 | 議員 ID   |       | 無投票フラグ  |
| 11 | 発言者名    |       | 任期開始    |
| 12 | 発言者の役職  |       | 任期満了    |
| 13 | 発言文     |       | 備考      |
| 14 | 発言以外    |       | 自治体コード  |
| 15 | 原本 URL  |       | 選挙 ID   |
| 16 | —       |       | 選挙区 ID  |

表 2: 結びつけることができた分析対象者

| データ名            | 完全一致  | 不一致   | 合計    |
|-----------------|-------|-------|-------|
| 会議録コーパス・議員リスト   | 2,172 | 775   | 2,947 |
| 第 18 回統一地方選挙の結果 | 2,172 | 1,863 | 4,035 |

表 3: 第 18 回統一地方選挙に立候補した現職議員の発言の有無 (期間: 2011 年 4 月~2015 年 3 月)

| 番号 | 都道府県 | 議員 ID あり |     |       |            |    |     | 合計    |
|----|------|----------|-----|-------|------------|----|-----|-------|
|    |      | 発言あり     |     |       | 発言なし（議長など） |    |     |       |
|    |      | 当選       | 落選  | 合計    | 当選         | 落選 | 合計  |       |
| 1  | 北海道  | 58       | 10  | 68    | 11         | 0  | 11  | 79    |
| 2  | 青森県  | 40       | 3   | 43    | 0          | 0  | 0   | 43    |
| 3  | 秋田県  | 31       | 4   | 35    | 2          | 0  | 2   | 37    |
| 4  | 山形県  | 33       | 4   | 37    | 4          | 0  | 4   | 41    |
| 5  | 栃木県  | 39       | 7   | 46    | 0          | 0  | 0   | 46    |
| 6  | 群馬県  | 36       | 3   | 39    | 0          | 0  | 0   | 39    |
| 7  | 埼玉県  | 64       | 6   | 70    | 1          | 1  | 2   | 72    |
| 8  | 千葉県  | 69       | 5   | 74    | 2          | 0  | 2   | 76    |
| 9  | 神奈川県 | 59       | 18  | 77    | 4          | 0  | 4   | 81    |
| 10 | 新潟県  | 37       | 5   | 42    | 0          | 0  | 0   | 42    |
| 11 | 富山県  | 32       | 2   | 34    | 1          | 0  | 1   | 35    |
| 12 | 石川県  | 28       | 3   | 31    | 1          | 0  | 1   | 32    |
| 13 | 福井県  | 29       | 0   | 29    | 0          | 0  | 0   | 29    |
| 14 | 山梨県  | 25       | 2   | 27    | 0          | 0  | 0   | 27    |
| 15 | 長野県  | 43       | 8   | 51    | 0          | 0  | 0   | 51    |
| 16 | 岐阜県  | 34       | 3   | 37    | 0          | 0  | 0   | 37    |
| 17 | 静岡県  | 56       | 6   | 62    | 0          | 0  | 0   | 62    |
| 18 | 愛知県  | 64       | 9   | 73    | 0          | 0  | 0   | 73    |
| 19 | 三重県  | 39       | 4   | 43    | 0          | 0  | 0   | 43    |
| 20 | 滋賀県  | 33       | 4   | 37    | 0          | 1  | 1   | 38    |
| 21 | 京都府  | 45       | 2   | 47    | 0          | 0  | 0   | 47    |
| 22 | 大阪府  | 59       | 15  | 74    | 0          | 0  | 0   | 74    |
| 23 | 兵庫県  | 53       | 9   | 62    | 6          | 1  | 7   | 69    |
| 24 | 奈良県  | 33       | 5   | 38    | 0          | 0  | 0   | 38    |
| 25 | 和歌山県 | 33       | 2   | 35    | 0          | 0  | 0   | 35    |
| 26 | 鳥取県  | 25       | 3   | 28    | 0          | 0  | 0   | 28    |
| 27 | 島根県  | 26       | 5   | 31    | 0          | 0  | 0   | 31    |
| 28 | 岡山県  | 46       | 4   | 50    | 0          | 0  | 0   | 50    |
| 29 | 山口県  | 34       | 6   | 40    | 2          | 0  | 2   | 42    |
| 30 | 広島県  | 49       | 2   | 51    | 2          | 0  | 2   | 53    |
| 31 | 徳島県  | 28       | 3   | 31    | 0          | 0  | 0   | 31    |
| 32 | 愛媛県  | 37       | 1   | 38    | 1          | 0  | 1   | 39    |
| 33 | 香川県  | 35       | 2   | 37    | 1          | 0  | 1   | 38    |
| 34 | 高知県  | 23       | 4   | 27    | 0          | 0  | 0   | 27    |
| 35 | 福岡県  | 0        | 0   | 0     | 65         | 4  | 69  | 69    |
| 36 | 佐賀県  | 30       | 4   | 34    | 0          | 0  | 0   | 34    |
| 37 | 長崎県  | 33       | 6   | 39    | 1          | 0  | 1   | 40    |
| 38 | 熊本県  | 31       | 8   | 39    | 1          | 0  | 1   | 40    |
| 39 | 大分県  | 32       | 3   | 35    | 0          | 0  | 0   | 35    |
| 40 | 宮崎県  | 29       | 2   | 31    | 1          | 1  | 2   | 33    |
| 41 | 鹿児島県 | 35       | 6   | 41    | 0          | 0  | 0   | 41    |
| 合計 | -    | 1,565    | 198 | 1,763 | 106        | 8  | 114 | 1,877 |

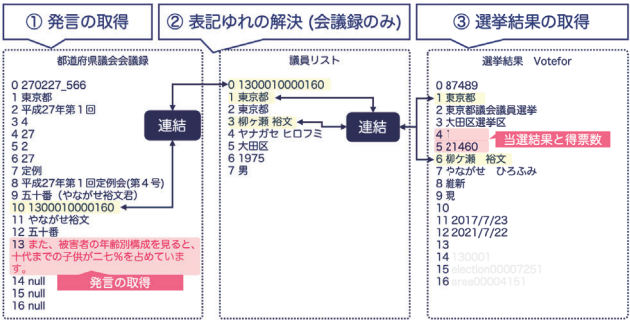


図 1: 会議録と議員リストと選挙情報を結びつける例

2.3 分析に用いる会議録および選挙情報

本節では、発言内容と選挙結果を結びつける方法について説明する。選挙結果は、政治山<sup>\*1</sup>を運営している Votefor から入手した。選挙結果のデータには 4,035 人の (落選者を含む) 立候補者全員の情報が含まれている。一方、47 都道府県の地方議会会議録コーパスには、2,947 人の議員が含まれている<sup>\*2</sup>。

データセット構築では、上記の2つのデータに加えて、地方議会会議録コーパスで利用している「議員リスト」を用いて「都道府県名」「姓名」が一致する場合に、データを結びつける。表 1 に 3 つのデータ構造と連結キー (下線が引いてあるフィールド名) を示す。また、図 1 に会議録と議員リストと選挙情報を結びつける例を示す。

表 2 に会議録コーパスと Votefor で管理している選挙結果情報に含まれる「都道府県名」および「姓名」の字面が完全に一致した立候補者の人数を示す。字面が完全に一致した立候補者を自動で連結した結果、2,172 人に対して「議会の発言」と「選挙結果」を結びつけることができた。

\*1 <https://seijiyama.jp/>

\*2 地方議会会議録コーパスに含まれる議員数は議員定数より多くなっている。その理由は、選挙時期がずれている自治体があり 2 期分の議員を含んでいるため、そして、補欠選挙により当選した議員が含まれているためである。

2.4 分析対象データ

データセットは、第 18 回統一地方選挙の選挙結果を予測するために、2015 年 4 月に議会選挙が実施された 41 都道府県を対象とした。分析対象データは、現職議員を対象としており、選挙結果を結びつけることができた議員のみが含まれる。表 3 は第 18 回統一地方選挙 (2015 年 4 月 12 日) の立候補者における現職議員を対象として、発言の有無を示している。本稿では、福岡県のデータを自動で結びつけることができなかったため、対象外とした<sup>\*3</sup>。

3. 発言内容による分析

本章では、議会における発言内容 (発言に含まれる単語) が次回選挙に与える影響について調査する。

\*3 本稿では、自動で連結したデータを用いているが、今後は、人手により確認したデータを用いる予定である。

3.1 分析方法

本研究では当選者および落選者の言語表現の特徴を明らかにするために、当選者と落選者の発言に含まれる単語の出現頻度の違いに着目する。

本研究では対数尤度比を用いて、当選者と落選者に特徴的な単語をみつける。対数尤度比は、次の式で計算する。

$$LLR = a \log \frac{an}{((a+b)(a+c))} + b \log \frac{bn}{((a+b)(b+d))} + c \log \frac{cn}{((c+b)(a+c))} + d \log \frac{dn}{((c+d)(b+d))}$$

表 4 に対数尤度比を求めるために必要となる単語の頻度を示す。また、n は全単語数であり、 $n = a + b + c + d$  である。対数尤度比を計算するために、当選者と落選者の発言を 2 つのファイルに分けて、NEolgd を用いた MeCab 形態素解析結果を用いて単語分割した後に、名詞だけを対象として、単語頻度を求めている。

表 4: 対数尤度比算出時に使用するクロス集計表

|              | 落選者             | 当選者            |
|--------------|-----------------|----------------|
| 単語 X の出現回数   | a<br>Lose(X)    | b<br>Win(X)    |
| 単語 X 以外の出現回数 | c<br>Lose(NotX) | d<br>Win(NotX) |

3.1.1 対数尤度比の結果

表 3 からわかるように、落選者が 10 名以上いる自治体が、北海道、神奈川県、大阪府である。表 5 に、落選者が多い 3 つの自治体において対数尤度比上位 5 単語を示す。上位の 5 単語には「お尋ね」「伺い」「見解」「所見」などの議員が質問するときに利用する表現が見られた。また、「日本海」「カジノ」「太陽光パネル」「泉南市」のように具体的な単語もみられた。しかしながら、これらの単語が発言においてどのように利用されているか、わからない。そこで、これらの具体的な単語が、当選者と落選者で異なる使い方をされているのかについて調べることとした。大阪府議会において「太陽光パネル」を含む発言した議員のなかから、落選した議員の発言と当選した議員の発言の例を下記に示す。この 2 つの発言をみると、太陽光パネルの設置を促進する同じような意見を述べている。このような場合には、発言の内容だけでは当落の判断が困難であることがわかる。

落選した議員の発言

「つまり、大阪府は、太陽光パネルの設置検討、これは確かにしたんですけども、残念ながら、結果としてほとんど太陽光パネル、今年度設置いたしません。」

当選した議員の発言

「府は、二十五年度に、太陽光パネルなどの省エネ・省 CO2 設備設置の融資利息の軽減のための預託制度を実施するとのことですが、わずかな利子軽減では、太陽光パネル設置促進にはほど遠いと言わざるを得ません。」

上記 2 名の議員の場合、発言量（発言文字数）を比較すると、当選した議員が落選した議員の 2 倍の発言量であった。選

挙結果を予測するには、発言に含まれる意見に加えて、発言量など他の要素も考慮する必要がある。

次節では、対数尤度比では計算できない、当選者のみが発言した単語について調査する。

表 5: 北海道、神奈川県、大阪府の対数尤度比の結果

| 北海道    |         |            |        |           |       |
|--------|---------|------------|--------|-----------|-------|
| 単語 X   | Lose(X) | Lose(NotX) | Win(X) | Win(NotX) | LLR   |
| 試案     | 17      | 61,134     | 1      | 332,648   | 27.97 |
| お尋ね    | 55      | 61,096     | 82     | 332,567   | 24.00 |
| 伺い     | 304     | 60,847     | 1,037  | 331,612   | 23.51 |
| 値上げ    | 58      | 61,093     | 92     | 332,557   | 23.48 |
| 日本海    | 26      | 61,125     | 17     | 332,632   | 22.44 |
| 神奈川県   |         |            |        |           |       |
| 単語 X   | Lose(X) | Lose(NotX) | Win(X) | Win(NotX) | LLR   |
| 海      | 80      | 84,140     | 42     | 333,510   | 59.05 |
| 見解     | 9       | 84,211     | 355    | 333,197   | 52.18 |
| カジノ    | 30      | 84,190     | 2      | 333,550   | 41.02 |
| 所見     | 324     | 83,896     | 682    | 332,870   | 40.35 |
| 文化芸術   | 38      | 84,182     | 17     | 333,535   | 30.68 |
| 大阪府    |         |            |        |           |       |
| 単語 X   | Lose(X) | Lose(NotX) | Win(X) | Win(NotX) | LLR   |
| お尋ね    | 284     | 55,104     | 345    | 202,439   | 87.69 |
| 障がい    | 62      | 55,326     | 34     | 202,750   | 41.27 |
| 太陽光パネル | 44      | 55,344     | 13     | 202,771   | 40.27 |
| わけ     | 56      | 55,332     | 565    | 202,219   | 34.58 |
| 泉南市    | 25      | 55,363     | 1      | 202,783   | 34.49 |

3.1.2 当選者のみが発言する単語

当選者のみ発言する単語とは、落選者の出現頻度が「0」の単語である。また、落選者が発言しない単語に対して、当選者はその単語を含めて、どのような発言をしているのか、考察する。表 6 に、全データの中で、落選者が発言していない単語の一部を示す。掲載した単語は、落選者が 6 名以上いる道府県において、当選者だけが発言する単語の対数尤度比上位 3 単語である。

神奈川県議会において当選議員が「生活支援ロボット」についてどのような発言をしていたのかを確認する。以下に「生活支援ロボット」を含む発言の一部を抜き出した例を示す。

1. さがみロボット産業特区は、生活支援ロボットの実用化、普及を通じて...
2. 神奈川版オープンイノベーションで開発していく 生活支援ロボットのテーマを発信するとともに...
3. こうした未来像を実現するため、生活支援ロボットのさらなる普及に向けて...

神奈川県で当選した議員は、生活支援ロボットという「新しい取り組み」について言及していることがわかる。この例から、時代に合わせた政治課題（となるキーワード）を含む発言が当落の判断材料になり得るのか、検討の余地がある。

4. 選挙結果の予測実験

本章では、議会における発言内容を用いて、選挙結果の予測を行う。本実験の目的は、議会での発言内容を用いて、既存の機械学習でどの程度当選予測が行えるのかを調査することである。本実験では、入力を各議員の発言から作成した特徴ベクトルとして、出力を「当選」あるいは「落選」の 2 値とする。入力データの作成は、議員の地方議会会議録の発言を形態素解析ツール MeCab を用いて分割し、Bag-of-words のベクトルを作成する。素性とする単語は、4 回以上出現する単語とした。また、全議員の発言において 30%以上出現する単語



表 6: 道府県別の当選者だけが発言する単語の例

| 単語 X     | Lose(X) | Lose(NotX) | Win(X) | Win(NotX) |
|----------|---------|------------|--------|-----------|
| 北海道      |         |            |        |           |
| 党        | 0       | 61,151     | 82     | 332,567   |
| 出先機関     | 0       | 61,151     | 60     | 332,589   |
| 移住       | 0       | 61,151     | 58     | 332,591   |
| 栃木県      |         |            |        |           |
| 指定廃棄物    | 0       | 42,515     | 84     | 248,137   |
| 所管       | 0       | 42,515     | 68     | 248,153   |
| 常任委員会    | 0       | 42,515     | 58     | 248,163   |
| 埼玉県      |         |            |        |           |
| 危険ドラッグ   | 0       | 34,030     | 78     | 349,558   |
| 県職員      | 0       | 34,030     | 75     | 349,561   |
| 認知症      | 0       | 34,030     | 72     | 349,564   |
| 神奈川県     |         |            |        |           |
| 生活支援ロボット | 0       | 84,220     | 95     | 333,457   |
| 朝鮮学校     | 0       | 84,220     | 82     | 333,470   |
| がん患者     | 0       | 84,220     | 60     | 333,492   |
| 長野県      |         |            |        |           |
| 消費税      | 0       | 58,493     | 110    | 392,635   |
| 討論       | 0       | 58,493     | 80     | 392,665   |
| 社会保障     | 0       | 58,493     | 72     | 392,673   |
| 静岡県      |         |            |        |           |
| 避難所      | 0       | 32,205     | 95     | 335,423   |
| 自民改革会議   | 0       | 32,205     | 82     | 335,436   |
| 教師       | 0       | 32,205     | 72     | 335,446   |
| 愛知県      |         |            |        |           |
| 意見書案     | 0       | 56,648     | 161    | 406,693   |
| 消防団員     | 0       | 56,648     | 132    | 406,722   |
| 暫時休憩     | 0       | 56,648     | 129    | 406,725   |
| 大阪府      |         |            |        |           |
| 御堂筋      | 0       | 55,388     | 66     | 202,718   |
| 借金       | 0       | 55,388     | 58     | 202,726   |
| 臨時財政対策債  | 0       | 55,388     | 56     | 202,728   |
| 兵庫県      |         |            |        |           |
| 公明党      | 0       | 33,289     | 93     | 236,402   |
| 商店街      | 0       | 33,289     | 81     | 236,414   |
| まち       | 0       | 33,289     | 77     | 236,418   |
| 山口県      |         |            |        |           |
| 岩国錦帯橋空港  | 0       | 77,286     | 81     | 339,782   |
| 開港       | 0       | 77,286     | 50     | 339,813   |
| 表現       | 0       | 77,286     | 48     | 339,815   |
| 長崎県      |         |            |        |           |
| td       | 0       | 55,281     | 405    | 345,529   |
| (株)      | 0       | 55,281     | 245    | 345,689   |
| 農業生産法人   | 0       | 55,281     | 108    | 345,826   |
| 熊本県      |         |            |        |           |
| 山鹿市      | 0       | 79,867     | 75     | 307,779   |
| 北朝鮮      | 0       | 79,867     | 74     | 307,780   |
| 公会       | 0       | 79,867     | 56     | 307,798   |
| 鹿児島県     |         |            |        |           |
| 売却       | 0       | 73,763     | 98     | 474,951   |
| 臨港道路     | 0       | 73,763     | 88     | 474,961   |
| 患者       | 0       | 73,763     | 84     | 474,965   |

も対象外とした。その後で、ベクトルを LSI(Latent Semantic Indexing) を用いて 100 次元のベクトルに圧縮する。出力データは、当選を「1」、落選を「-1」のいずれかの値である。分類手法は、Support Vector Machine を用いる。実験データは、表 3 に含まれる「発言あり」の議員 1,763 人の発言である。学習データは北海道から広島県までの 1,411 人（当選 1,252 人、落選 159 人）の発言とし、評価データは徳島県から鹿児島県までの 352 人（当選 313 人、落選 39 人）の発言とした。表 7 に、SVM による予測の Precision, Recall, F 値を示す。再現率が 1.00 になっているが、これは全て出力を「当選」としたためである。

この結果から、発言に含まれる単語から適切に学習できていないことがわかる。機械学習を扱う場合に一般的に問題となる「学習データの不足」「不均衡なデータセット」が原因だと考えられる。本研究で用いたデータの場合、自治体によって当選者数にばらつきがあることや、議員数が限られていることも影響している。

また、学習データと評価データの分け方にも工夫の余地がある。例えば、都道府県別にデータを分割し、47 分割の交差検定をすることも考えられる。また、同一自治体を対象として、2011 年の選挙、2015 年の選挙を学習データとして、2019 年

の選挙を予測することも考えられる。

さらに、発言単語だけで選挙結果の予測をできるのか、という根本的な問題もある。この点については「発言量」を用いた川浦らの研究 [1] を参考にしながら、検討を進める予定である。

表 7: Suport Vector Machine による予測結果

| Precision | Recall    | F-measure |
|-----------|-----------|-----------|
| 0.89      | 1.00      | 0.94      |
| (313/352) | (313/313) |           |

5. おわりに

本稿では、会議録、議員プロフィール、選挙結果が含まれるデータから、発言内容と選挙結果（当落）を結びつけて、データセットを構築する方法について述べた。そして、機械学習による選挙予測を行うための予備調査として、構築したデータセット内の各議員の発言内容（発言に含まれる単語）が次回選挙に影響を与えるのかについて、調査を行なった。本調査では、対数尤度比を用いて当選者と落選者の発言に特徴的な単語を明らかにした。また、当選者のみが発言している単語についても考察を行った。その結果、当選者・落選者に特徴的な語はあるものの、単語の出現頻度の情報だけでは当落の予測は困難であり、発言量や発言時期などの要素も考慮する必要があることを確認した。さらに、発言内容に含まれる単語を素性とした当落予測実験を行い、単純な素性では当落の判断が困難であることも明らかになった。

今後は、2019 年 4 月の第 19 回の統一地方選挙の結果を用いて研究を進める予定である。

謝辞

本研究は JSPS 科研費 JP16H02912, JP17K03785 の助成を受けています。

参考文献

[1] Akihiko Kawaura, Yasutomo Kimura, Keiichi Takamaru, and Yuzu Uchida. Elected officials in the local assembly: Analysis of prefectural plenary session transcripts. *Doshisha University Center for the Study of the Creative Economy, Discussion Paper Series*, No.2018-02, 2018.

[2] 湯浅 壱道. インターネット選挙運動と公職選挙法. *選挙研究*, 30(2):75–90, 2014.

[3] 上ノ原 秀晃. 2013 年参議院選挙におけるソーシャルメディア. *選挙研究*, 30(2):116–128, 2014.

[4] 鳥海 不二夫 and 吉田 光男. 2017 年衆議院選挙における政党公式アカウントフォロワーの分析. *人工知能学会全国大会論文集*, JSAI2018:1E305–1E305, 2018.