

ニュース記事の分類に着目した議員活動可視化手法の提案

Visualizing Activities of the Diet Members by Focusing on Categories of News Articles

竹中 未来* 松下 光範
Miku Takenaka Mitsunori Matsushita

関西大学総合情報学部
Faculty of Informatics, Kansai University

Abstract: Since politics is a part of our lives, paying attention to the activities of the Diet members is important. It, however, is difficult to have a comprehensive understanding of the activities because the activities of the Diet members are so diverse. To solve this problem, we propose a method to visualize the activities of the Diet members by focusing on the categories and attention of news articles in which their activities are described. In the proposed method, the categories of their activities are determined by referring to the classification of the news articles. Also, the degree of attention to activities is determined by considering the pages of the newspaper in which the article appeared. The results suggest that the proposed method helps users to understand the differences in related categories and characteristic activities among the Diet members.

1 はじめに

政治に対する関心の低下は、大きな社会問題である [13]. 特に、若年層の政治に対する興味関心の低さは深刻であり、投票率の低下が問題視されている [12, 14]. この投票率の低下には様々な要因がある. 例えば、全国の 18~20 歳の男女 3000 人を対象として実施された参議院議員通常選挙 (2016 年 7 月) の投票行動調査*¹によると、投票に行かなかった理由として「今住んでいる市区町村で投票することができなかったから」が最も多く、「選挙にあまり関心がなかったから」「投票所に行くのが面倒だったから」「どの政党や候補者に投票すべきかわからなかったから」がそれに次いで挙げられている.

本研究では、政治に関心を持つきっかけとして、政治家の活動内容を分かり易く提示し、身近に感じてもらうことで、この問題の解決の一端を探る. 政治は私たちの生活に関わるため、どのような社会的問題に対してどの議員がどのような活動を行っているのかを知ることは重要である. しかし、議員活動は多岐にわたるため、議員がどのような活動に取り組んでいるのかを包括的に把握

することは多大な労力を必要とする.

この問題解決の一助として、高丸らは、地方議会を対象に議員活動の把握に取り組んでいる [9]. この研究では、地方議会会議録を用いて発言に含まれる単語の傾向から議員ごとに発言内容の差を分析し、各々の議員が取り組んでいる活動の傾向を明らかにしている. しかし、会議録には議員ごとの発言は記載されているものの、その発言は分野別に整理されておらず、発言の注目度を測る指標も示されていない. そのため、議員活動の要点を把握するには政治に関する基礎的知識が必要になり、政治に関心の低い人が端的に理解するには敷居が高い.

この問題を解決するため、本研究では新聞記事で議員活動がどのように報道されているかを端的に理解できるように提示することを試みる. 新聞記事のニュースは、新聞社が定めた紙面の分類基準に従ってカテゴリ別に掲載される. これに加えて、新聞社が重要だと判断する注目度の高いニュースは新聞記事の 1~3 面に掲載される. そこで、このような性質を持つ新聞記事に着目して議員ごとの特徴を可視化することで、議員が取り組む注目度の高い活動や活動分野の把握の容易化を目指す.

* 連絡先：関西大学総合情報学部
〒569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: mat@res.kutc.kansai-u.ac.jp

*¹ 18 歳選挙権に対する意識調査報告書 (総務省), https://www.soumu.go.jp/main_content/000456091.pdf (2021/2/28 確認).

2 関連研究

本章では、新聞記事を用いた政治を分析する研究や国会会議録や地方議会会議録を用いた議員に関する研究について述べる。

田中らは、新聞社による公共事業に対する批判報道の傾向を時系列で分析している [11]。新聞記事の見出しや本文に含まれる公共事業に関連する単語の掲載回数から公共事業に対するポジティブやネガティブと捉えられる報道がされた時期の推移やその報道がされた背景について考察し、公共事業関係費が増加した際に、ネガティブな報道が増加していることを明らかにしている。

細貝らはマスメディアが政治に与える影響力を新聞記事を用いて調査している [15]。新聞記事で報道される内閣に関する内容が内閣支持率の変動に影響を与えていると考え、新聞社説のテキストの文末表現に着目しポジティブ・ネガティブを定量化することで、新聞記事の報道が内閣支持率へ与える影響を分析している。新聞記事で報道される内容がポジティブの場合、内閣支持率が上昇し、新聞記事で報道される内容がネガティブの場合、内閣支持率が下降すると考察している。

内田らは、地方議会会議録より、各議員の発言の文字数や発言に含まれる政治に関する単語の出現頻度を男女別や年齢層別で比較している [4]。その結果、発言に含まれる単語の出現頻度から男女別、年齢層別の発言内容に差異があることを明らかにしている。ただし、この研究では政治に関する単語を著者が選定しており、その客観性や適切性に疑問が残る。そこで、客観的な分析を行うために、内田らは、政治に関する単語の関連度合いを図る指標として、「政治語彙度」を提案している [5]。政治語彙度が 80% 以上のものを政治特徴語と定義づけ、発言内容に含まれる政治特徴語の出現頻度を男女別や年齢層別で比較している。その結果、内田ら [4] が行った、男女別、年齢層別の発言内容の差異の分析結果と傾向が類似していることを明らかにしている。

高丸らは、地方議会を経て国会議員になった議員を対象に、地方議会会議録と国会会議録を用いて、地方議会と国会での発言に含まれる単語を比較している [10]。発言内容から地方議会と国会で同一の政治課題に着目し続けている議員がいることを明らかにしている。

木村らは、地方議会会議録を用いて、議員の発言内容から選挙結果を予測することを試みている [8]。当選者のみの発言内容には「生活支援ロボット」のような単語が含まれ、時代に合った単語を使用していると考察している。しかし、当選者と落選者の発言に含まれる単語の

出現頻度のみの違いからでは選挙結果の予測が難しいことを示唆している。

掛谷らは国会会議録を用いて、問題発言や不祥事によって任期途中で辞任する議員や大臣について発言内容の差を機械学習を用いて分析している [7]。その結果、任期が長く続かない議員は任期が長く続く議員に比べ、発言中に敬語表現が少ないことを明らかにしている。

これらの先行研究を鑑みると、個々の議員の発言傾向や具体的活動の把握には、会議録をリソースとすることが適切であると考えられる。一方で、政治的イシューやそれに対する国民の反応といったマクロな視点からの洞察には、新聞記事をリソースとすることが適切であると考えられる。本研究では、政治への関心が低い若年者をリードユーザと考えているため、Shneiderman の Visualization Mantra [3] に倣い、まず全体を概観することを試みる。

新聞記事のニュースは、新聞社が定めた分類基準に従ってカテゴリ別に掲載される。また、新聞記事の 1~3 面に新聞社が重要だと判断するニュースが掲載される。このように、分類基準や注目度を取得しやすい新聞記事が本研究のリソースとしてより適切な媒体である。そこで本研究では、こうした新聞記事の特性を利用して、議員活動を可視化する。

3 データセット

本稿では、議員活動の分類指標として、毎日新聞社*2 の新聞記事紙面の分類を参考に、国際、経済、家庭、文化、科学、芸能、スポーツ、社会の 8 分野を用いる。新聞記事のデータには、毎日新聞社の記事データベースより 2016 年の 1 年間分 (83,668 件) の新聞記事から、前述した 8 分野に該当する新聞記事 (57,058 件) と、1~3 面の新聞記事 (9,011 件) の計 66,069 件を用いる。

ただし、注目度の高いニュースが掲載されている新聞記事の 1~3 面は前述した 8 分野と各記事が紐付けられていないため、以下に示す手法を用いて、これらの新聞記事に該当する分野を付与する。

3.1 新聞記事の分類モデルの作成

新聞記事の分類には、Support Vector Machine (以下、SVM と記す) を用いる。SVM の f パラメータは、カーネルを rbf 、コストパラメータを 1、 γ を $scale$ とした。SVM の実装には、python (Ver.3.7.4) のライブラリ

*2 毎日新聞 <https://mainichi.jp> (2021/2/28 確認)

である, scikit-learn (Ver.0.20.1)^{*3}を用いた. まず, 本稿で用いる新聞記事の本文を形態素に分割し, 名詞のみを抽出した. 形態素の分割には, MeCab (Ver.0.996.2)^{*4}を用いた. SlothLib[6]^{*5}に記載されている名詞に加え, Mecab で未知語 (e.g., マイナンバーカード) と設定されている名詞と 1 文字の名詞をストップワードに設定した. 既に分野が付与されている 1~3 面以外の 8 分野の新聞記事の名詞に word2vec[1] を適用することで単語分散表現を獲得した. word2vec の実装には, python のライブラリである gensim (Ver.3.7.1)^{*6}を用いた. パラメータは, ベクトルの次元数を 100, window サイズを 5, workers を 3 とした. この獲得した単語分散表現を平均化し [2]^{*7}, 各記事をベクトル化した. 1~3 面以外の 8 分野の新聞記事は既に分野毎に分類されており, ラベルが付与されている. そこでこのラベルを分類基準とし, SVM を用いてこの記事ベクトルから新聞記事の分類モデルを作成した. この新聞記事の分類モデルの妥当性を検証するため, 1~3 面以外の 8 分野の新聞記事を学習データと評価データに分割し 5 分割交差検証を行った. 評価データの新聞記事の分類を行った結果の平均は, 正解率は 0.842, F 値は 0.833 であった.

3.2 1~3 面の新聞記事の分類

3.1 節で述べた結果から, 新聞記事の分類モデルの妥当性が検証されたため, このモデルを用いて, 8 分野の分野が付与されていない 1~3 面の新聞記事の分類を行った. 1~3 面以外の 8 分野に該当する新聞記事全てを用いて, 単語分散表現を獲得する. この単語分散表現から, 1~3 面の新聞記事の各記事をベクトル化し, 分野が付与されていない 1~3 面の新聞記事の分類を行った.

4 提案する可視化手法

3 章で述べた 1~3 面以外の 8 分野に該当する新聞記事と, 3.2 節で 8 分野に分類を行った注目度の高い 1~3 面の新聞記事のうち, 新聞記事の本文に議員の名前がフルネームで掲載されている新聞記事を対象に議員活動の可視化を行う. 提案する議員活動可視化システムの実装には p5.js(Ver.1.0.0)^{*8}を用いた.

なお, 本稿では, 第 3 次安倍内閣第 1 次改造内閣^{*9}の留任している国務大臣 8 名を可視化の対象とする.

4.1 偏差値の算出

本稿では, 各議員の議員活動を比較のために基準として偏差値を用いる. 新聞記事の多寡を可視化する際は, 議員ごとの分野別, 同一分野内での注目度別それぞれの偏差値を算出した. 標準偏差 σ は算出式 (1) で求める.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

ここで, n はデータの総数, x は各議員の新聞記事数, \bar{x} は $x_1, x_2, x_3, \dots, x_i$ の平均を示す. また, 偏差値 Z は算出式 (2) を用いて求める.

$$Z = \frac{10(x_i - \bar{x})}{\sigma} + 50 \quad (2)$$

x_i は各議員の新聞記事数, \bar{x} は $x_1, x_2, x_3, \dots, x_i$ の平均を示している. σ は標準偏差を示しており, 分野別の偏差値を算出する際の母集団は分野別の記事を対象とした. 注目度別の偏差値を算出する際の母集団は異なる議員間における同一分野内かつ同一の注目度別の記事を対象とした.

4.2 可視化インタフェース

新聞記事の注目度は, 円の内側から新聞記事の 1 面, 2 面, 3 面, 1~3 面以外というように注目度の高い順に並べた (図 1 参照). 新聞記事の多寡を可視化する際は, 4.1 節で述べた偏差値の考え方をを用いて, 議員ごとの分野別, 同一分野内での注目度別それぞれの偏差値を算出し, その段階を色で示すこととした. 分野別, 注目度別ともに, 偏差値を段階に分ける際には, 平均からの差が -0.5σ 未満, -0.5σ 以上 0.5σ 未満, 0.5σ 以上の 3 段階で分ける (ここでの標準偏差 σ は 10 である. ここでの標準偏差は偏差値を段階に分けるためのものであり, 偏差値を算出する過程で算出する標準偏差の値とは異なる). 分野別の記事の多寡は, 分野名の文字の色を使って示す. 分野別の偏差値は, 橙色は平均からの差が -0.5σ 未満, 緑色は -0.5σ 以上 0.5σ 未満, 赤色は 0.5σ 以上を表している. なお, 文字の色が黒色の分野名は, その分野に記事が存在しないことを示す. 同一分野内で

^{*3} <https://scikit-learn.org/stable/> (2021/2/28 確認)

^{*4} <http://taku910.github.io/mecab/> (2021/2/28 確認)

^{*5} <https://ja.osdn.net/projects/slothlib/> (2021/2/28 確認)

^{*6} <https://radimrehurek.com/gensim/> (2021/2/28 確認)

^{*7} <https://github.com/yagays/swem> (2021/2/28 確認)

^{*8} <http://p5js.org> (2021/2/28 確認)

^{*9} <https://www.kantei.go.jp/jp/rekidainaikaku/097.html> (2021/2/28 確認)

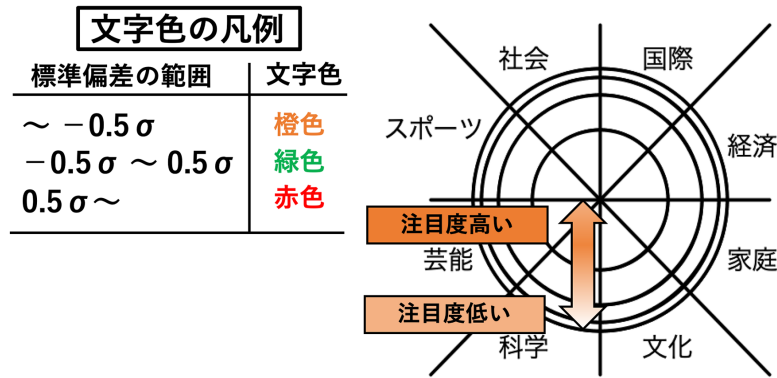


図1 提案する可視化手法

注目度別の記事の多寡は、該当する領域を塗りつぶす色で示す。同一分野内で注目度別の偏差値は、橙色は平均からの差が -0.5σ 未満、緑色は -0.5σ 以上 0.5σ 未満、赤色は 0.5σ 以上を表し、白色は記事が存在しないことを示している。

5 議員活動可視化についての考察

3.2 節で述べた 8 分野に分類を行った注目度の高いニュースが掲載されている 1~3 面の新聞記事と 1~3 面以外の 8 分野に該当する新聞記事をリソースとして、4 章で述べた手法を用いて可視化を行った。第 3 次安倍内閣第 1 次改造内閣の留任している 国務大臣 8 名の議員活動の可視化結果を図 2~9 に示す。この 8 名のうち、岸田文雄氏と塩崎恭久氏を対象にそれぞれの議員活動の可視化結果を分野別、注目度別の観点から比較して述べる。岸田氏は当時、外務大臣を務めており、塩崎氏は当時、厚生労働大臣を務めていた^{*10}。

5.1 分野別の議員活動に関する考察

可視化の結果を、分野別に比較する。岸田氏の可視化の結果を図 4 に示す。岸田氏の可視化結果では、国際および芸術分野が赤く示されている。該当する記事を確認したところ、芸術分野についても外交に関する新聞記事が含まれており、外務大臣としての職務に関連深い活動が可視化できたとと言える。次に、塩崎氏の議員活動可視化の結果を図 5 に示す。塩崎氏の可視化結果では、科学分野が赤く示されている。科学分野の記事には医療や介護に関わる話題など社会保障に関する新聞記事が含まれており、厚生労働大臣としての職務に関連深い活動が可

視化できたとと言える。

分野別の議員活動の比較から、以下の 2 点がうかがえた。

- (1) 各議員の職務に関連深い活動分野が明確化されている
- (2) 職務に関連深いと予想が難しい活動分野だが赤く示されている活動分野でも、職務に関連深い活動が把握できた

職務に関連深い活動分野の可視化ができており、本可視化手法の有効性が示唆された。

5.2 同一分野内での注目度別の議員活動に関する考察

可視化の結果を、両議員に共通した社会分野について注目度の観点から分析する。岸田氏は、1 面の新聞記事が赤く示されており、塩崎氏は 1~3 面以外の新聞記事が赤く示されている。この結果から、社会分野において、塩崎氏の活動より岸田氏の活動がより世間に注目されていたことがうかがえる。岸田氏の 1 面の新聞記事には、日米地位協定や核兵器禁止条約、G7 に関するニュースが含まれ、塩崎氏の 1~3 面以外の新聞記事には、働き方改革や医療、待機児童の内容が含まれていた。

同一分野内での注目度別の議員活動の比較から、以下の 2 点がうかがえた。

- (1) 同一の活動分野で議員によって世間の注目度が異なる
- (2) 同一分野内だが、議員によって注目される活動は異なる

世間の注目度の違いが明確化されており、本可視化手法の有効性の一端が示唆された。

可視化の結果、5.1 節、5.2 節より、各議員に関連深い

^{*10} 本稿で記載する議員の役職名は第 3 次安倍内閣第 1 次改造内閣時のものとする。

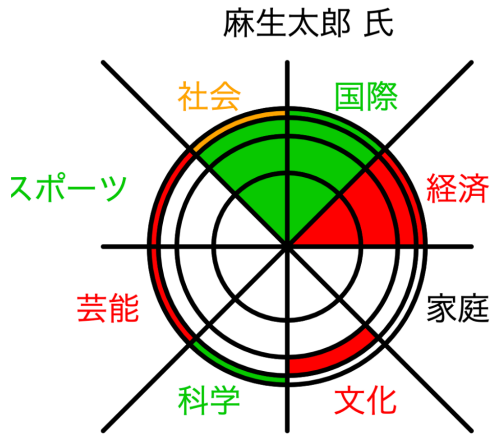


図2 麻生太郎氏の議員活動の可視化結果

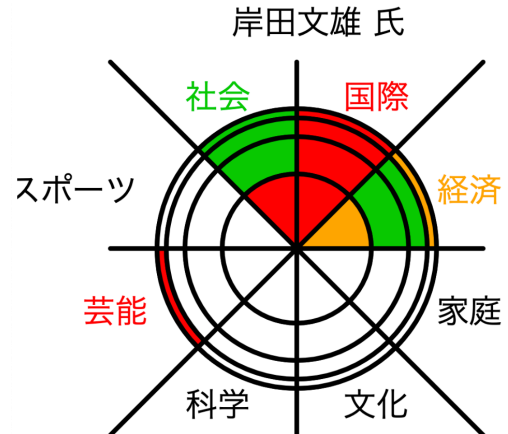


図4 岸田文雄氏の議員活動の可視化結果

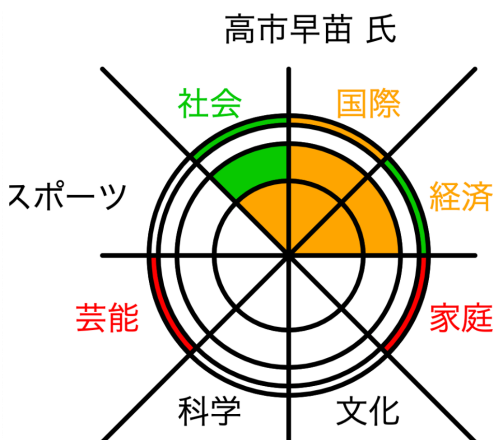


図3 高市早苗氏の議員活動の可視化結果

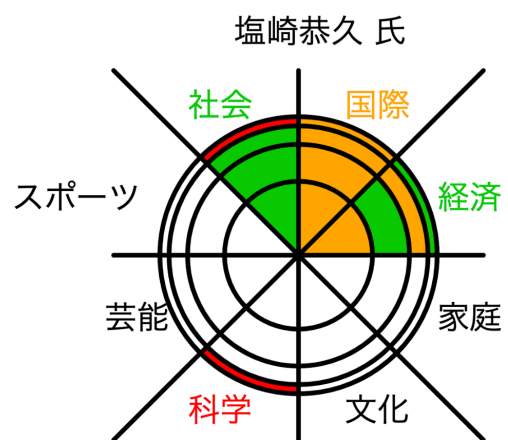


図5 塩崎恭久氏の議員活動の可視化結果

活動分野や世間の注目度の違いが明確化されており、本可視化手法の有効性が示唆された。

6 課題と展望

6.1 提案手法の有効性と適用限界

本研究では、新聞社が判断した注目度の高い議員の発言や世間が関心を持っている情報などを判断し掲載している新聞記事に着目し、各議員の名前が掲載されている新聞記事の分野と注目度に着目した議員活動の可視化手法を提案した。その結果、各議員に関連深い活動分野や世間の注目度の違いが明確化されており、本可視化手法の有効性が示唆された。

しかし新聞記事では、議員の活動の様子や取り組む社会課題への対応といった内容だけでなく、収賄事件などネガティブな文脈で紙面に取り上げられる場合がある。本研究では、現時点でこれらを区別することなく扱っている。そのため、これらを区別し、ユーザに示す技術が

必要になる。そのためには、本来の政治活動とは異なるスキャンダル記事とを区別する技術を提案手法に組み込む必要がある。また、議員活動は必ずしもすべてが新聞記事として記述されるわけではないため、議員活動全体を包括した可視化を行うことができていない。こうした活動を対象に含めるには、新聞記事以外のコーパス（例えば会議録や政治活動報告書など）を併せて処理し、ユーザに提示する必要がある。

6.2 議員活動の可視化の改善

今後議員活動を把握する際に、期待される機能について述べる。まず、会議録に記載されている発言内容が新聞記事のどの分野に該当するかを照らし合わせることで、分野別に分類されていない会議録の発言内容の該当分野を把握可能となり、各議員の分野別の議員活動の把握が期待できると考えられる。次に、可視化の結果から、分野と注目度の観点から議員活動の把握は可能であるが、各分野や注目度の領域に該当する新聞記事の内容

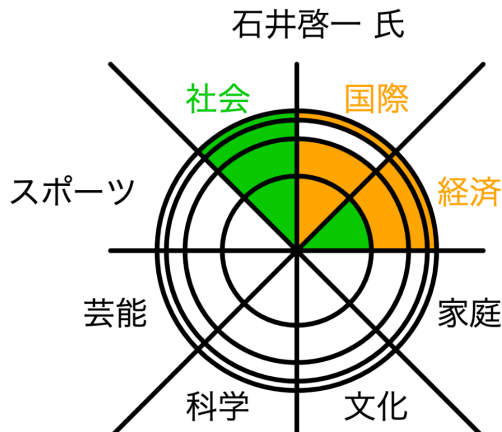


図6 石井啓一氏の議員活動の可視化結果

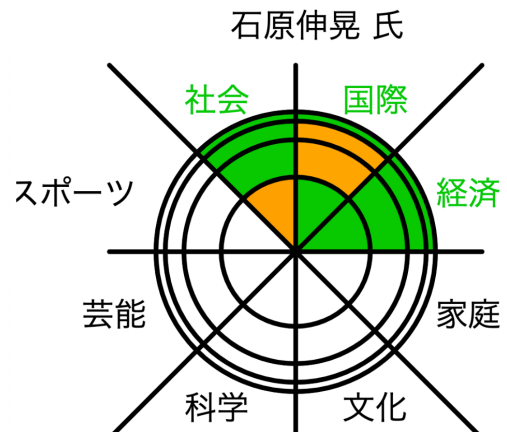


図8 石原伸晃氏の議員活動の可視化結果

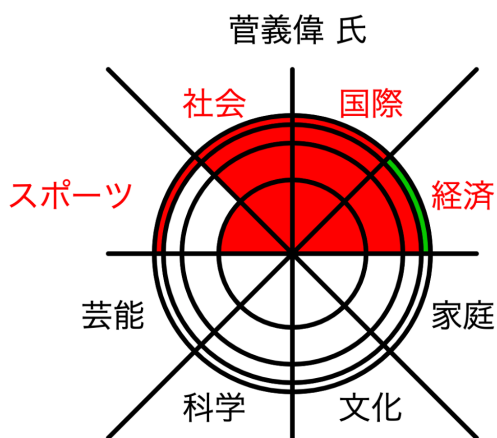


図7 菅義偉氏の議員活動の可視化結果

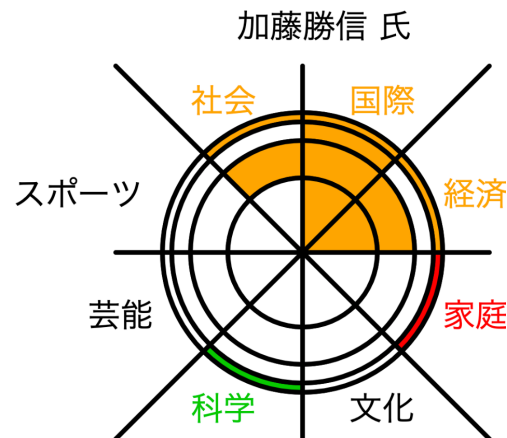


図9 加藤勝信氏の議員活動の可視化結果

の把握ができないため、議員活動の詳細を把握することができない。今後、議員活動の分野と注目度の領域を選択することで、選択した領域に該当する新聞記事を閲覧可能にすることを旨とする。

6.3 議員活動の分類指標の検討

本稿では、議員活動を8分野を分類指標として可視化した。1~3面以外の8分野に分類されている新聞記事のうち、社会分野の新聞記事には国際や経済など他の分野の内容の新聞記事も含まれており、議員ごとに活動分野を比較する際に、社会分野は指標として適していないと考えられる。今後、社会分野の新聞記事も細分化し分類することで、議員活動の分類精度が高くなると予想される。議員活動を8分野に分類し可視化を行ったが、国際、経済、社会分野の新聞記事数が8名の議員全体の記事数の97.9%を占めていた。また、議員別に記事数の割合を比較したところ、8名中7名が各議員の記事数全体の90%以上が国際、経済、社会分野であった。こ

の3分野に新聞記事が偏っており、議員別に議員活動の分野に差異が出ないと考えられる。そのため、他5分野の分類指標の検討や、この3分野の新聞記事を細分化し分類することで議員活動の分類精度が高くなると予想される。

6.4 記事数の多寡による偏差値

本節では、記事数から算出した偏差値の改善点について述べる。本稿では、議員活動を分野別、同一分野内で注目度別に色分けで可視化を行った。分野別、注目度別ともにその色分けをするにあたり、偏差値を平均50からの差が -0.5σ 未満、緑色は -0.5σ 以上 0.5σ 未満、赤色は 0.5σ 以上の3段階に分けた。その結果、5章より、各議員の議員活動の比較が可能であることが示唆された。色分けの基準を検討するために、平均50からの差が -1σ 未満、 -1σ 以上 1σ 未満、 1σ 以上の3段階に分けることも試みた。しかし、平均からの差が -1σ 未満、 -1σ 以上 1σ 未満、 1σ 以上の3段階で色分けを行った

ところ、分野別、注目度別両者ともに偏差値による色分けに差異が見られなかった。この要因として、議員ごとの新聞記事数の割合に偏りが大きいことが考えられる。

7 おわりに

本稿では、各議員の議員活動を比較可能にするために、議員の名前が掲載されている新聞記事を用いて、その分野と注目度に基づいて可視化する手法の提案を行った。その結果、各議員に関連深い活動分野や世間の注目度の違いが明確化されており、本可視化手法の有効性が示唆された。今後、疑惑や不祥事などの議員活動と関連が薄い新聞記事が含まれる可能性を考慮し、これらを除いた議員活動のみの内容が含まれた新聞記事での可視化を目指す。

謝辞

本研究では、「毎日新聞記事データ集 2016 版」の新聞記事データを使用した。記して謝意を表す。

参考文献

- [1] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed Representations of Words and Phrases and Their Compositionality, *Advances in Neural Information Processing Systems*, Vol. 26, pp. 3111–3119 (2013).
- [2] Shen, D., Wang, G., Wang, W., Renqiang Min, M., Su, Q., Zhang, Y., Li, C., Henao, R. and Carin, L.: Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms, *Proc. 56th Annual Meeting of the Association for Computational Linguistics*, pp. 440–450 (2018).
- [3] Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, *Proc. IEEE Symposium on Visual Languages*, p. 336 (1996).
- [4] 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知: 都道府県議会会議録コーパスを用いた議員の議会活動の可視化に向けて, 第 32 回人工知能学会全国大会論文集, 1E3-03 (2018).
- [5] 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知: 対数尤度比と政治語彙度を用いた議員の議会活動の可視化, 知能と情報, Vol. 31, No. 2, pp. 662–671 (2019).
- [6] 大島裕明, 中村聡史, 田中克己: Slothlib: Web サーチャ研究のためのプログラミングライブラリ, 日本データベース学会 Letters, Vol. 6, No. 1, pp. 113–116 (2007).
- [7] 掛谷英紀, 大南勝: 国会会議録に基づく短命議員・短命大臣の特徴分析, 知能と情報, Vol. 31, No. 2, pp. 617–625 (2019).
- [8] 木村泰知, 川浦昭彦, 内田ゆず: 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, 第 33 回人工知能学会全国大会論文集, 2I5-J-9-02 (2019).
- [9] 高丸圭一, 内田ゆず, 木村泰知: 地方政治コーパスにおける都道府県議会会議録パネルデータの基礎分析, 宇都宮共和大学シティアイフ学論叢, Vol. 18, pp. 136–155 (2017).
- [10] 高丸圭一, 内田ゆず, 木村泰知, 松田謙次郎: 地方議会と国会における同一議員による発言の比較に向けた検討, 第 35 回ファジィシステムシンポジウム講演論文集, pp. 324–329 (2019).
- [11] 田中皓介, 神田佑亮, 藤井聡: 公共政策に関する大手新聞社報道についての時系列分析, 土木学会論文集 D3 (土木計画学), Vol. 69, No. 5, pp. L373–L379 (2013).
- [12] 土倉莞爾: 18 歳選挙権について考える, 關西大學法學論集, Vol. 65, No. 5, pp. 1453–1496 (2016).
- [13] 内閣府: 特集 1:日本の若者意識の現状～国際比較からみえてくるもの～, 令和元年版子供・若者白書 (全体版) (2019).
- [14] 永戸力: 若者の政治参加を考える, 愛知大学法学部法経論集, No. 207, pp. 21–42 (2016).
- [15] 細貝亮: メディアが内閣支持に与える影響力とその時間的变化: 新聞社説の内容分析を媒介にして, マス・コミュニケーション研究, Vol. 77, pp. 225–242 (2010).