

2023年度（令和5年度） 修士論文

機械学習を用いた賛成を得やすい  
文章の種々の分析

令和6年2月13日

鳥取大学大学院 持続性社会創生科学研究科  
工学専攻 情報エレクトロニクスコース

自然言語処理研究室

M22J4050Z 三木 謙志

## 概要

本研究は Yahoo!ニュースのコメント欄と機械学習を用いて、賛成を得やすい文章の種々の分析を目的としている。記事を”政治”，”事件”，”病気”，”スポーツ”，”その他”の分野別に分類した実験と記事に投稿されたコメントを数値評価を用いて分類した実験の2種類の提案手法を用いた。

記事分野別に分類した結果，”政治”134 記事，”事件”43 記事，”病気”40 記事，”スポーツ”36 記事，”その他”57 記事となった。そして，2種類の提案手法と機械学習を利用して推定させた結果，正解率は多くの場合で7割前後となり，多くの場合で正解率が BERT > ME > SVM の順になっていた。同時に素性分析を行った結果，”政治”分野で「賢明」が賛成を得やすいと判断されていた。これは「賢明」が知的な単語であるため，文章全体もまとまっていて読みやすくなっているのではないかと推測することができる。また，”事件”分野と”病気”分野で「本当に」，「絶対」なども賛成を得やすいと判断されていたがこれらは強調の意味を持つ単語であるため，自信があるコメントに使用されることで賛成を得やすくなっているのではないかと考えられる。

数値評価を用いて分類し，機械学習で素性分析を行った。低評価を用いた数値評価1では「速報」などが賛成を得やすいとされた。「速報」は速い情報ほど価値があるため賛成を得やすくなっているのではないかと考えることができる。高評価と低評価の比率を利用した数値評価2では「誕生」などが賛成を得やすいとされた。「誕生」は何かが生まれたことを意味するポジティブな意味を持つ単語であるため低評価を得にくく，賛成を得やすいと判断されたと考えることができる。単位時間あたりの高評価を利用した数値評価3では「おめでとう」などが賛成を得やすいとされた。「おめでとう」などの直感的に良いか悪いかを判断できる素性も賛成が得やすいと判断された。直感的に良いか悪いかを判断できる単語を用いて文章を作成すると短時間である程度の賛成を得ることに期待できる。

# 目次

|            |   |          |
|------------|---|----------|
| <b>第1章</b> | <b>はじめに</b>   | <b>1</b> |
| <b>第2章</b> | <b>先行研究</b>   | <b>3</b> |
| 2.1        | 機械学習を利用した賛成を得やすい文章の分析 . . . . .                       | 3        |
| 2.2        | 絵文字を用いた文章における感情伝達効果に関する研究 . . . . .                   | 4        |
| 2.3        | 感情生起事象文型パターンに基づいた会話文からの感情推定手法 . . . . .               | 4        |
| 2.4        | 文間関係認識に基づく賛成・反対意見の俯瞰 . . . . .                        | 4        |
| 2.5        | 意見文課題における説得のアピールの日西対照研究-日本とスペインの<br>学生の作文比較 . . . . . | 5        |
| 2.6        | 感動を与える文の自動取得と分析 . . . . .                             | 5        |
| <b>第3章</b> | <b>提案手法</b>   | <b>6</b> |
| 3.1        | 提案手法 . . . . .  | 6        |
| 3.1.1      | 記事分野別 . . . . .                                       | 7        |
| 3.1.2      | 数値評価 . . . . .  | 7        |
| 3.2        | 機械学習 . . . . .  | 8        |
| 3.2.1      | 教師あり機械学習 . . . . .                                    | 8        |
| 3.2.2      | ME . . . . .  | 8        |
| 3.2.3      | 正規化 $\alpha$ 値 . . . . .                              | 9        |
| 3.2.4      | SVM . . . . .   | 9        |
| 3.2.5      | BERT . . . . .  | 9        |
| 3.3        | データ作成方法 . . . . .                                     | 10       |
| 3.3.1      | 記事分野別データ作成 . . . . .                                  | 10       |
| 3.3.2      | 数値評価データ作成 . . . . .                                   | 11       |
| 3.3.3      | 数値評価1 . . . . .                                       | 11       |
| 3.3.4      | 数値評価2 . . . . .                                       | 12       |
| 3.3.5      | 数値評価3 . . . . .                                       | 12       |

|            |                      |           |
|------------|----------------------|-----------|
| 3.4        | 推定方法 . . . . .       | 12        |
| 3.4.1      | 文章対の推定方法 . . . . .   | 12        |
| 3.4.2      | 数値評価の推定方法 . . . . .  | 13        |
| 3.5        | 素性 . . . . .         | 13        |
| 3.5.1      | 文章対の素性 . . . . .     | 13        |
| 3.5.2      | 数値評価の素性 . . . . .    | 14        |
| <b>第4章</b> | <b>実験</b>            | <b>15</b> |
| 4.1        | 記事分野別 . . . . .      | 15        |
| 4.1.1      | MEの素性 . . . . .      | 16        |
| 4.1.2      | SVMの素性 . . . . .     | 22        |
| 4.2        | 数値評価 . . . . .       | 27        |
| 4.2.1      | 数値評価1 . . . . .      | 27        |
| 4.2.2      | 数値評価2 . . . . .      | 27        |
| 4.2.3      | 数値評価3 . . . . .      | 28        |
| 4.2.4      | 数値評価の性能 . . . . .    | 28        |
| 4.2.5      | ME素性 . . . . .       | 28        |
| 4.2.6      | SVM素性 . . . . .      | 33        |
| 4.2.7      | BERTの素性 . . . . .    | 37        |
| 4.3        | 素性分析 . . . . .       | 41        |
| 4.3.1      | 記事分野別の素性分析 . . . . . | 41        |
| 4.3.2      | 数値評価の素性分析 . . . . .  | 41        |
| 4.3.3      | 今後の課題 . . . . .      | 42        |
| <b>第5章</b> | <b>考察</b>            | <b>44</b> |
| <b>第6章</b> | <b>おわりに</b>          | <b>46</b> |

# 表 目 次

|      |                     |    |
|------|---------------------|----|
| 4.1  | 記事分野別機械学習の性能評価      | 16 |
| 4.2  | ME の素性 (政治)         | 17 |
| 4.3  | ME の素性 (事件)         | 18 |
| 4.4  | ME の素性 (病気)         | 19 |
| 4.5  | ME の素性 (スポーツ)       | 20 |
| 4.6  | ME の素性 (その他)        | 21 |
| 4.7  | SVM の素性 (政治)        | 22 |
| 4.8  | SVM の素性 (事件)        | 23 |
| 4.9  | SVM の素性 (病気)        | 24 |
| 4.10 | SVM の素性 (スポーツ)      | 25 |
| 4.11 | SVM の素性 (その他)       | 26 |
| 4.12 | 機械学習の性能評価           | 28 |
| 4.13 | ME の素性 (数値評価 1-1)   | 29 |
| 4.14 | ME の素性 (数値評価 1-2)   | 30 |
| 4.15 | ME の素性 (数値評価 2)     | 31 |
| 4.16 | ME の素性 (数値評価 3)     | 32 |
| 4.17 | SVM の素性 (数値評価 1-1)  | 33 |
| 4.18 | SVM の素性 (数値評価 1-2)  | 34 |
| 4.19 | SVM の素性 (数値評価 2)    | 35 |
| 4.20 | SVM の素性 (数値評価 3)    | 36 |
| 4.21 | BERT の素性 (数値評価 1-1) | 37 |
| 4.22 | BERT の素性 (数値評価 1-2) | 38 |
| 4.23 | BERT の素性 (数値評価 2)   | 39 |
| 4.24 | BERT の素性 (数値評価 3)   | 40 |

# 第1章 はじめに

本研究は Yahoo!ニュースのコメント欄の情報と機械学習を用いて、賛成を得やすい文章の種々の分析を目的としている。記事を分野別に分類し、コメント欄の賛成した人数、時刻の情報を使用し、賛成が得やすい文章を機械学習を用いて推定させる。加えて、高評価、低評価、コメント時刻の情報から数値評価を求め、記事に投稿されたコメントを高評価コメント、中評価コメント、低評価コメントの3種類に分類する。その中から高評価コメントと低評価コメントの2種類を機械学習に入力して高評価か低評価を推定させる。同時に機械学習の推定結果の根拠を素性分析の技術を利用して求める。この素性分析を行うことによって、賛成を得やすい素性が明らかになることが期待される。

本研究の主な主張点を以下に整理する。

- 本研究では様々な条件の文章を ME, SVM, BERT の機械学習を利用して実験を行った結果、ほとんどの場合  $BERT > ME > SVM$  の順で性能が高くなった。
- 機械学習での素性分析を行った結果、分野ごとによって賛成を得やすいとされる素性は大きく異なったが、「政治」分野で賛成を得にくい素性とされた「賢明」は単語が知的なため文章がまとまっており読みやすいため賛成が得やすいのではないかと考えることができる。
- 数値評価で分類して素性分析を行った結果、「おめでとう」などの直感的に良いと判断できる単語は少ない時間である程度の高評価を得ることが可能だと分かった。

本論文の構成は以下の通りである。

第2章ではこれまでの関連する研究を説明する。

第3章では本研究における、賛成を得やすい文章の判定方法と判定に利用する技術を説明する。

第4章では賛成を得やすい文章の自動判定とその評価を行う。

第5章では考察を行い、効果的な利用方法を考察する。

第6章ではまとめを行う.

## 第2章 先行研究

本章では、先行研究について記述する。

2.1節では、三木ら [1] [2] が行った機械学習を利用した賛成を得やすい文章の分析の研究について記述する。2.2節では、高橋ら [3] が行った絵文字を用いた文章における感情伝達効果に関する研究について記述する。2.3節では、松本ら [4] が行った会話文からの感情推定手法に関する研究について記述する。2.4節では、水野ら [5] が行った文間関係認識に基づく賛成・反対意見の俯瞰に関する研究について記述する。2.5節では、西條ら [6] が行った意見文課題における説得のアピールの日西対照研究-日本とスペインの学生の作文比較に関する研究について記述する。

### 2.1 機械学習を利用した賛成を得やすい文章の分析

三木ら [1] [2] は Yahoo!ニュースのコメント欄を利用して賛成を得やすい文章の特徴の発見を目的とし、賛成を得やすい文章とは Yahoo!ニュースの同じ記事に対してコメントされた文章を比較し、賛成した人数がより多く、コメント時刻がより最近のコメントを賛成を得やすい文章だと定義している。比較した時の賛成を得やすい文章と賛成を得にくい文章を文章対として機械学習に入力し、どちらが賛成を得やすい文章かを推定させた。推定させた結果、BERT, 最大エントロピー法, SVM の順で正解率が高くなり、一番正解率が高い BERT で 0.75 となった。素性分析の結果、賛成を得やすい文章の素性だと納得できる素性の発見はできていないが、「バカ」や「幼稚」などの賛成を得にくい文章の素性だと納得できる素性はいくつか発見した。今後の課題としては賛成を得やすい文章の特徴の発見が挙げられる。そこで、本論文では記事の分野を分類することで分野別の賛成を得やすい文章の特徴の発見が可能になると考え、実験を行なう。

## 2.2 絵文字を用いた文章における感情伝達効果に関する研究

高橋ら [3] は CMC(コンピュータを媒介したコミュニケーション)における文章中の絵文字の使用頻度と書き手と読み手の感情伝達における絵文字の効果について調べるため実験と分析を行った。その結果、書き手の感情が正しく伝わりやすいのは Smileys と分類された顔の絵文字を使用する文章で、その他の文章は感情の種類によって伝わりやすい場合、伝わりにくい場合が存在することが示唆された。すなわち絵文字が感情を伝達する効果が絵文字のカテゴリや感情の種類によって異なることを示した。また、感情語によって書き手から読み手への伝達の仕方に差が見られた。本実験の結果では「喜び、好き、安心」と「悲しい、怖れ、恥ずかしい、高揚、驚き」という感情語で結果が異なった。

## 2.3 感情生起事象文型パターンに基づいた会話文からの感情推定手法

従来は表面的な感情表現に焦点を当てていたのに対し、松本ら [4] は感情語や慣用句、そしてそれらが出現する文のパターンを分析することでより正確な感情推定を行うアルゴリズムを提案し、評価実験用システムを構築した。評価実験では約 76%の成功率が得られた。しかし、推定結果と人手による解との完全一致率が低いという問題があった。評価実験において感情生起事象文型パターンに一致する文が全体の 5%に満たなかったことから、文型パターンの拡充と共に、文脈を判断するためのアルゴリズムが必要であるとわかった。

## 2.4 文間関係認識に基づく賛成・反対意見の俯瞰

水野ら [5] はウェブ上の大量にあるテキスト情報の信憑性判断にはユーザの持っている先入観が正常な判断を妨げてしまう場合を確証バイアスが高い状態であるといい、そのような確証バイアスを軽減するために言論マップ生成課題に取り組んでいる。言論マップとはユーザが与える言明に対して、その言明に対する賛成、反対意見を抽出し、俯瞰的に示すものを指す。本論文では文間関係認識の技術課題に焦点を当て、制約を段階的に緩和するモデルの開発と分析を行った。結果として段階的に制約を緩め

ていくことで関係分類の精度と再現率をコントロールできることがわかり、文間関係認識において最も重要な課題がアライメントということが明らかになった。

## 2.5 意見文課題における説得のアピールの日西対照研究- 日本とスペインの学生の作文比較

西條ら [6] は日本語母語話者による文章表現の特徴とともに、他の言語の特徴的な表現を理解していくことはグローバル化の近年に重要であると考えて実験を行った。結果としては日本語母語話者は言論のアピールと情動のアピールを組み合わせることで言論をもって真なることと読み手に共感を与える要素を提示する傾向があることが明らかになり、スペイン語母語話者は読み手に共感を与えることはあまり重要視せず、言論のアピールと習慣のアピールを組み合わせることで、言論を持って真なることと生活様式などを訴える要素を提示しながら説得することが明らかになった。

## 2.6 感動を与える文の自動取得と分析

端ら [7] は、人は日々感動を求め、感動によって動かされる生き物であるということから「感動」と「文」に重きを置き、感動を与える文か否かの自動判定に関する研究を行うことによって、感動を与える文の作成支援を行った。自動判定は教師有り機械学習である SVM やパターンマッチングによって行っている。また、感動を与える文で多く使われる単語を収集することで、感動を与える文の言語的特徴を明らかにしている。

## 第3章 提案手法

本章では、本研究の提案手法の説明を記述する。

3.1節では、提案手法の大まかな流れを記述している。3.1.1節では、提案手法の1つである記事分野別の説明を記述している。3.1.2節では、提案手法の1つである数値評価の説明を記述している。

3.2節では、機械学習の説明を記述している。3.2.1節では、教師あり機械学習の説明を記述している。3.2.2節では、本研究で使用する機械学習法である最大エントロピー法 (ME) の説明を記述している。3.2.3節では、本研究で使用する機械学習である ME で求まる  $\alpha$  値を正規化した正規化  $\alpha$  値の説明を記述している。3.2.4節では、本研究で使用する機械学習法である SVM の説明を記述している。3.2.5節では、本研究で使用する機械学習法である BERT の説明を記述している。

3.3節では、機械学習に inputs データの作成方法を記述している。3.3.1節では、文章対データの作成方法を記述している。3.3.2節では、数値評価データの作成方法を記述している。3.3.3節では、数値評価1の説明を記述している。3.3.4節では、数値評価2の説明を記述している。3.3.5節では、数値評価3の説明を記述している。

3.4節では、機械学習での推定方法の説明を記述している。3.4.1節では文章対での推定方法の説明を記述している。3.4.2節では数値評価での推定方法の説明を記述している。

3.5節では、機械学習で使用する素性の説明を記述している。3.5.1節では文章対で使用する素性の説明を記述している。3.5.2節では数値評価で使用する素性の説明を記述している。

### 3.1 提案手法

本実験で使用する2種類の提案手法についての説明をそれぞれ記述する。

### 3.1.1 記事分野別

本研究では賛成を得やすい文章と賛成を得にくい文章の2つを利用して文章対を作成し、ME, SVM, BERTの三種類の機械学習に入力し、どちらが賛成を得やすいかを推定させる。同時に機械学習の推定結果の根拠を素性分析の技術を利用して求める。この分析は賛成を得やすい文章の作成につながると考えている。先行研究では記事の分野を分けずに研究を行った。今回は先行研究で集めた記事を分野別に分けて同様の研究を行う。分野別に分けることで特定の分野に賛成を得やすい文章の特徴の発見などが期待できる。具体的には”病気”, ”事件”, ”政治”, ”スポーツ”, ”その他”の5つの分野に記事を分類し、機械学習に賛成を得やすい文章を推定させる。

### 3.1.2 数値評価

先行研究ではコメントと高評価数とコメント時刻の情報を利用して研究を行っていた。今回の研究では記事分野別で使ったコメントと高評価とコメント時刻に低評価も加えて研究を行う。高評価, 低評価, コメント時刻の情報を利用して数値評価によって高評価コメント, 中評価コメント, 低評価コメントの3種類に分類する。そして分類した高評価コメントと低評価コメントを使用してME, SVM, BERTの3種類の機械学習に入力し、高評価コメントか低評価コメントかどうかを推定させる。

数値評価では以下の3種類の数値評価を求める式を用いて分析を行う。

- 数値評価1 =  $\frac{\text{高評価} - \text{低評価}}{\text{コメント時刻}}$
- 数値評価2 =  $\frac{\text{高評価}}{\text{高評価} + \text{低評価}}$
- 数値評価3 =  $\frac{\text{高評価}}{\text{コメント時刻}}$

数値評価1では高評価から低評価を引き、コメント時刻で割ることで数値を求める。これによって低評価を用いて賛成を得やすい文章の分析を行うことが可能になると考える。

数値評価2では高評価から高評価と低評価の和を割ることで数値を求める。これによって高評価と低評価の比を利用して賛成を得やすい文章の分析を行うことが可能になると考える。

数値評価3では高評価をコメント時刻で割ることで数値を求める。これによって一定時間当たりで得た高評価数が多いコメントで実験を行うことが可能になると考える。

それぞれの式から数値を求め、コメントを高評価、中評価、低評価に分類し、高評価と低評価となったコメントを機械学習にそれぞれ入力する。

## 3.2 機械学習

機械学習 (Machine Learning) は、コンピュータが経験やデータから学習し、新しい情報に基づいて問題を解決するための手法やアプローチの総称である。伝統的なプログラミングでは、人間が明示的にルールや手順をコンピュータに教え込んでいたが機械学習では、アルゴリズムがデータから学習し、パターンを見つけ、そのパターンを用いて未知のデータに対する予測や判断を行う。

### 3.2.1 教師あり機械学習

教師あり機械学習とはデータセットに入力とそれに対応する正解するラベルが含まれており、アルゴリズムはこのデータセットから学習して未知のデータに対する予測を行う。本研究で使用する ME, SVM, BERT も教師あり機械学習に分類される。

### 3.2.2 ME

教師あり機械学習である最大エントロピー法とは、あらかじめ設定しておいた素性  $f_j (1 \leq j \leq k)$  集合を  $F$  とするとき、式 (3.1) を満足しながらエントロピーを意味する式 (3.2) を最大にするときの確率分布  $p(a, b)$  を求め、その確率分布にしたがって求まる各分類の確率のうち、もっとも大きい確率値を持つ分類を求める分類とする方法である。

$$\sum_{a \in A, b \in B} p(a, b) g_j(a, b) = \sum_{a \in A, b \in B} \tilde{p}(a, b) g_j(a, b) \quad (3.1)$$

*for*  $\forall f_j (1 \leq j \leq k)$

$$H(p) = - \sum_{a \in A, b \in B} p(a, b) \log(p(a, b)) \quad (3.2)$$

ただし,  $A, B$  は分類と文脈の集合を意味し,  $g_j(a, b)$  は文脈  $b$  に素性  $f_j$  があってなおかつ分類が  $a$  の場合 1 となり, それ以外で 0 となる関数を意味する. また,  $\hat{p}(a, b)$  は, 既知データでの  $(a, b)$  の出現の割合を意味する.

式 (3.1) は, 確率  $p$  と出力と素性の組の出現を意味する関数  $g$  をかけることで出力と素性の組の頻度の期待値を求めることになっており, 右辺の既知データにおける期待値と, 左辺の求める確率分布に基づいて計算される期待値が等しいことを制約として, エントロピー最大化 (確率分布の平滑化) を行って, 出力と文脈の確率分布を求めるものとなっている.

### 3.2.3 正規化 $\alpha$ 値

正規化  $\alpha$  値とは, ME で求まる  $\alpha$  値を全分類先での合計が 1 となるように正規化した値である. また, 素性  $a$  と分類先  $b$  の対によって定まる値であり, 素性  $a$  のみが適用される場合に分類先  $b$  となる確率に相当する. 各素性の分類先ごとに与えられた正規化  $\alpha$  値が高いほど, その分類先であることを推定するのに重要な素性であることを意味する.

### 3.2.4 SVM

教師あり機械学習である SVM は, 空間を超平面で分割することにより 2 つの分類からなるデータを分類する手法である. このとき, 2 つの分類が正例と負例からなるものとする, 学習データにおける正例と負例の間隔 (マージン) が大きいものほどテストデータで誤った分類をする可能性が低いと考えられ, このマージンを最大にする超平面を求めそれを用いて分類を行う.

### 3.2.5 BERT

教師あり機械学習である BERT は, Bidirectional Encoder Representations from Transformers の略で, 「Transformer による双方向のエンコード表現」と訳され, 2018 年 10 月に Google の Jacob Devlin らの論文 [8] で発表された自然言語処理モデルである. 従来の機械学習では, 大量のラベルのついたデータを用意させ, 処理を行うことで課題に取り組む. しかし従来の手法に対し, BERT は事前学習でラベルのないデー

タをはじめに大量に処理を行う。その後、ファインチューニングで少量のラベルのついたデータを使用することで課題に対応させる。

### 3.3 データ作成方法

本研究の提案手法で使用するデータの作成方法についての説明をそれぞれ記述する。

#### 3.3.1 記事分野別データ作成

記事分野別データの作成には Yahoo!ニュースのコメント欄にある、時刻と賛成した人数の情報を利用する。2021年9月～12月の期間に Yahoo!ニュースの1,000コメント以上投稿された記事を対象に、コメント、賛成数、時刻の情報を310記事分収集している。同じ記事に投稿された2つのコメントにおいて、賛成した人数がより多く、コメント時刻がより最近という2点を満たすコメントを賛成を得やすい文章、もう一方を賛成を得にくい文章だと定義する。そのような賛成を得やすい文章と賛成を得にくい文章を文章対として大量に作成する。ただし、一度でも文章対を作成する際に使用されたコメントは他の文章対では使用しないこととする。例を以下に掲載する。

- コメント A, 賛成：4,234, コメント時刻:4時間前
- コメント B, 賛成：3,823, コメント時刻:2時間前
- コメント C, 賛成：6,923, コメント時刻:8時間前
- コメント D, 賛成：2,182, コメント時刻:5時間前

コメント A を対の一方として文章対を作成する場合、定義に当てはめると文章対として使用できるもう一方はコメント D のみである。コメント B を対の一方として利用する場合も同様に文章対として使用できるもう一方はコメント D のみであり、コメント C は賛成した人数が一番多いがコメント時刻が一番古いためどのコメントとも文章対を作成することはできない。

よって、この4つのコメントで定義に当てはまり、文章対として使用できるコメント対はコメント A とコメント D、コメント B とコメント D のみである。この時、賛成を得やすい文章はコメント A、コメント B となり、コメント D は賛成を得にくい文章となっている。しかし、本研究では文章対作成の際に一度使用したコメントは利用し

ないので、文章対として使用できるのはコメント A とコメント D の文章対かコメント B とコメント D の文章対のどちらか一つの文章対となっている。このように定義にあてはまる文章対をコメントデータから大量に作成する。

### 3.3.2 数値評価データ作成

数値評価データを以下の 3 種類の方法で作成した。

- 数値評価 1 =  $\frac{\text{高評価} - \text{低評価}}{\text{コメント時刻}}$
- 数値評価 2 =  $\frac{\text{高評価}}{\text{高評価} + \text{低評価}}$
- 数値評価 3 =  $\frac{\text{高評価}}{\text{コメント時刻}}$

数値評価のデータ作成に使用した高評価と低評価はコメントについての高評価と低評価の数を利用した。コメント時刻は 1 時間未満の投稿を全て 1 時間として計算を行い、1 時間未満の投稿を 1 時間としたため、1 時間以上の投稿は全て +1 時間として計算を行った。数値評価で出てきた数値をそれぞれの定義で高評価、中評価、低評価のコメントに分類して高評価と低評価のコメントのみを機械学習に入力して実験を行った。先行研究 [1] [2] との大きな違いとしては低評価を用いたことと 1 文ごとに機械学習に入力していることが挙げられる。特に 1 文ごとに入力することで BERT での素性分析が可能となった。

### 3.3.3 数値評価 1

先行研究では高評価とコメント時刻を利用して実験を行っていたが、数値評価 1 では高評価から低評価を引き、コメント時刻で割ることで数値を求めた。得られた数値からコメントを高評価、中評価、低評価に分類し、高評価と低評価となったコメントを機械学習に入力した。数値評価 1 には数値評価 1-1, 1-2 が存在するが、両者の違いは高評価とする数値の範囲が異なっている。

### 3.3.4 数値評価 2

数値評価 2 は高評価から高評価と低評価の和を割ることで数値を求めた。得られた数値からコメントを高評価，中評価，低評価に分類し，高評価と低評価となったコメントを機械学習に入力した。

### 3.3.5 数値評価 3

数値評価 3 は高評価をコメント時刻で割ることで数値を求めた。得られた数値からコメントを高評価，中評価，低評価に分類し，高評価と低評価となったコメントを機械学習に入力した。

## 3.4 推定方法

本節では記事分野別と数値評価の 2 種類で行った推定方法についての説明をそれぞれ記述する。

### 3.4.1 文章対の推定方法

作成した文章対の一方の文章を左側，もう一方を右側として，賛成を得やすい文章が左側の文章なら文章対の先頭に“L”を，賛成を得やすい文章が右側の文章なら文章対の先頭に“R”を付与したものを学習データとする。このような学習データである文章対を入力とし，“L”，“R”を出力とする。入力が与えられるとそれに対する出力を推定できるように 3 種類の機械学習 (ME, SVM, BERT) で学習する。

3.3.1 節の文章対を学習データとした場合の例を以下に示す。例のように作成した文章対の“L”と“R”と文章対の左右を反転させたものも同時に学習データとして機械学習に入力する。

- L, コメント A, コメント D
- R, コメント D, コメント A

例のようなコメント対と先頭の“L” or “R”を学習データとして機械学習に入力し，先頭の “L” or “R” によってどちらのコメントが賛成を得やすいかを学習させる。そして同様の形式で作成したコメント対のみをテストデータとして機械学習に入力し，左側

のコメントが賛成を得やすいと判断した場合には”L”を右側の場合は”R”を出力させることで機械学習の性能を求める。

### 3.4.2 数値評価の推定方法

分類した結果、高評価コメントとなったコメントの先頭には”T”を、低評価コメントとなったコメントの先頭には”F”を付与したコメントを学習データとする。以下に例を示す。

- T, コメント A
- F, コメント D

例のようなデータを学習データとして入力し、そのコメントが賛成を得やすいと判断すると”T”を賛成を得にくいと判断すれば”F”を出力するように3種類の機械学習(ME, SVM, BERT)で学習する。記事分野別の研究とは異なり、1コメント単位で入力している。

## 3.5 素性

本節では、本研究で使用した記事分野別と数値評価で使用した素性についての説明をそれぞれ記述する。

### 3.5.1 文章対の素性

文章対においてME, SVMは左側の文章にある単語は「L: 単語」、右側にある単語は「R: 単語」と「コメントの文字数」を素性として利用する。BERTに入力する際は素性の入力は不要なので文章対のみを入力している。「コメントの文字数」の素性は「L: ~以下」、「L: ~より大きい」、「R: ~以下」、「R: ~より大きい」としており、「~」には「10」、「20」、「50」、「100」、「200」、「500」、「1,000」のコメントの文字数を満たす数字が入る。

MEは正規化 $\alpha$ 値、SVMは分離平面を用いて素性分析を行う。MEでは、正規化 $\alpha$ 値の高いものが重要な素性となる。SVMでは「L: 単語」などの1単語を入力し、分離

平面からの距離が大きいものが重要な素性となる。文章対を用いたBERTの素性分析は困難なため現時点では行えていない。

### 3.5.2 数値評価の素性

数値評価においてME, SVMは「単語」と「コメントの文字数」を素性として利用する。BERTに入力する際は文章対と同様に素性の入力は不要なのでコメントのみを入力している。「コメントの文字数」の素性は「～以下」、「～より大きい」としており、「～」には文章対と同様に「10」、「20」、「50」、「100」、「200」、「500」、「1,000」のコメントの文字数を満たす数字が入る。

MEは正規化 $\alpha$ 値, SVMは分離平面を用いて素性分析を行う。MEでは、正規化 $\alpha$ 値の高いものが重要な素性となる。SVMでは1単語を入力し、分離平面からの距離が大きいものが重要な素性となる。BERTは段落を個々の単語に切り分け、学習済みBERTモデルを用いて1単語を入力する。その確率値を計算し、確率値が大きいものが重要な素性となる。

## 第4章 実験

本章では、本研究で行った実験の説明を記述する。

4.1 節では、実験データと実験結果の説明を記述している。4.1.1 節では、ME の素性分析の結果の説明を記述している。4.1.2 節では、SVM の素性分析の結果の説明を記述している。

4.2 節では、数値評価の説明を記述している。4.2.1 節では、数値評価 1 の説明を記述している。4.2.2 節では、数値評価 2 の説明を記述している。4.2.3 節では、数値評価 3 の説明を記述している。4.2.4 節では、3 種類の数値評価の性能の説明を記述している。4.2.5 節では、ME の素性を記述している。4.2.6 節では、SVM の素性を記述している。4.2.7 節では、BERT の素性を記述している。

4.3 節では、素性分析の説明を記述している。4.3.1 節では、記事分野別の素性分析の説明を記述している。4.3.2 節では、数値評価の素性分析の説明を記述している。4.3.3 節では、今後の課題を記述している。

### 4.1 記事分野別

1,000 コメント以上を持つ記事を対象に 310 記事分のコメントを収集し、同記事内の賛成した人数とコメント時刻の 2 つの条件を満たすコメントから文章対を作成している。対とするコメントの組み合わせは賛成を得やすい文章の定義を満たすコメントからランダムに決定している。

収集した 310 記事を”政治”，”病気”，”事件”，”スポーツ”，”その他”の 5 種類の記事に分類しそれぞれで機械学習で実験した。分類した結果，政治 134 記事，事件 43 記事，病気 40 記事，スポーツ 36 記事，その他 57 記事となった。

表 4.1 は機械学習により “L” or “R” のどちらの文章が賛成を得やすいかを推定したときの正解率を示している。

表 4.1: 記事分野別機械学習の性能評価

| 分類   | 政治     | 事件     | 病気     | スポーツ   | その他    | 平均     |
|------|--------|--------|--------|--------|--------|--------|
| ME   | 0.7280 | 0.7198 | 0.6894 | 0.7163 | 0.7547 | 0.7216 |
| SVM  | 0.6911 | 0.6769 | 0.6602 | 0.7011 | 0.7242 | 0.6907 |
| BERT | 0.7492 | 0.7505 | 0.7070 | 0.7022 | 0.7301 | 0.7278 |

多くの場合で  $BERT > ME > SVM$  の順の性能の高さになっており、平均でも  $BERT > ME > SVM$  の順の性能の高さになっている。平均に関しては BERT と ME が同程度の性能となっているが SVM だけやや低い性能になっている。しかし、全ての結果で約 7 割前後という性能になっている。

#### 4.1.1 ME の素性

記事分野別で行った ME の素性分析の結果で得られた賛成を得やすい素性、賛成を得にくい素性を表 4.2～表 4.6 に上位 30 個ずつ載せている。

表 4.2: ME の素性 (政治)

| 順位 | 賛成を得やすい素性 | $\alpha$ 値 | 賛成を得にくい素性 | $\alpha$ 値 |
|----|-----------|------------|-----------|------------|
| 1  | まし        | 0.8517     | わ         | 0.1483     |
| 2  | 為         | 0.8357     | 思わ        | 0.1642     |
| 3  | やはり       | 0.8085     | よかつ       | 0.1712     |
| 4  | 調査        | 0.7959     | 散々        | 0.1808     |
| 5  | 率         | 0.7743     | M子        | 0.1827     |
| 6  | 代         | 0.7729     | KK        | 0.1884     |
| 7  | 国家        | 0.7672     | 減らし       | 0.2056     |
| 8  | 儀式        | 0.7641     | くれ        | 0.2076     |
| 9  | 費         | 0.7620     | また        | 0.2100     |
| 10 | 欲しい       | 0.7611     | バラマキ      | 0.2136     |
| 11 | 感謝        | 0.7587     | ぞ         | 0.2188     |
| 12 | 何故        | 0.7550     | 詐欺        | 0.2299     |
| 13 | いつも       | 0.7526     | バカ        | 0.2366     |
| 14 | など        | 0.7467     | 投稿        | 0.2435     |
| 15 | 明らか       | 0.7437     | きっと       | 0.2443     |
| 16 | 立憲        | 0.7426     | 頭         | 0.2518     |
| 17 | 車         | 0.7374     | !         | 0.2558     |
| 18 | 技術        | 0.7360     | 神         | 0.2597     |
| 19 | 税金        | 0.7359     | 中国共産党     | 0.2599     |
| 20 | 多く        | 0.7348     | 一生        | 0.2685     |
| 21 | 階         | 0.7347     | 未来        | 0.2725     |
| 22 | 参拝        | 0.7339     | 内容        | 0.2725     |
| 23 | 以上        | 0.7334     | せい        | 0.2768     |
| 24 | 考え        | 0.7287     | 声         | 0.2770     |
| 25 | くる        | 0.7284     | 書         | 0.2772     |
| 26 | 支給        | 0.7267     | 洗脳        | 0.2795     |
| 27 | 立候補       | 0.7265     | 支援        | 0.2807     |
| 28 | 出馬        | 0.7263     | 男         | 0.2819     |
| 29 | 私         | 0.7253     | 当て        | 0.2819     |
| 30 | できる       | 0.7250     | やつ        | 0.2826     |

表 4.3: ME の素性 (事件)

| 順位 | 賛成を得やすい素性 | $\alpha$ 値 | 賛成を得にくい素性 | $\alpha$ 値 |
|----|-----------|------------|-----------|------------|
| 1  | を         | 0.8004     | 何         | 0.2685     |
| 2  | さん        | 0.7931     | 事実        | 0.2706     |
| 3  | トヨタ       | 0.7605     | せ         | 0.2732     |
| 4  | かも        | 0.7421     | やっ        | 0.2740     |
| 5  | 罪         | 0.7289     | 想像        | 0.2770     |
| 6  | ね         | 0.7249     | 起き        | 0.2772     |
| 7  | 車         | 0.7223     | ほしい       | 0.2812     |
| 8  | ,         | 0.7160     | トップ       | 0.2885     |
| 9  | 運転        | 0.7126     | 共犯        | 0.2893     |
| 10 | 思う        | 0.7122     | へ         | 0.2963     |
| 11 | 本当に       | 0.6928     | とか        | 0.3012     |
| 12 | い         | 0.6912     | もん        | 0.3084     |
| 13 | させ        | 0.6779     | くる        | 0.3180     |
| 14 | なっ        | 0.6738     | って        | 0.3274     |
| 15 | 思っ        | 0.6732     | 相談        | 0.3330     |
| 16 | あっ        | 0.6685     | 酷         | 0.3340     |
| 17 | 意味        | 0.6670     | 鬼畜        | 0.3346     |
| 18 | 過ぎる       | 0.6633     | 万         | 0.3383     |
| 19 | まし        | 0.6620     | いれ        | 0.3431     |
| 20 | なり        | 0.6597     | られる       | 0.3440     |
| 21 | 増え        | 0.6589     | いい        | 0.3445     |
| 22 | 鑑定        | 0.6578     | 金         | 0.3478     |
| 23 | もの        | 0.6534     | 見         | 0.3506     |
| 24 | 癖         | 0.6527     | 無い        | 0.3526     |
| 25 | この        | 0.6498     | 性         | 0.3528     |
| 26 | 最近        | 0.6496     | 以上        | 0.3534     |
| 27 | 自民党       | 0.6467     | 殺し        | 0.3541     |
| 28 | .         | 0.6466     | 団体        | 0.3553     |
| 29 | くらい       | 0.6458     | 人間        | 0.3587     |
| 30 | ず         | 0.6428     | 買っ        | 0.3597     |

表 4.4: ME の素性 (病気)

| 順位 | 賛成を得やすい素性 | $\alpha$ 値 | 賛成を得にくい素性 | $\alpha$ 値 |
|----|-----------|------------|-----------|------------|
| 1  | 絶対        | 0.7540     | 日本人       | 0.2431     |
| 2  | ばかり       | 0.7536     | また        | 0.2603     |
| 3  | 札幌        | 0.7533     | 推奨        | 0.2647     |
| 4  | でき        | 0.7476     | たい        | 0.2688     |
| 5  | として       | 0.7460     | 高い        | 0.2696     |
| 6  | 関係        | 0.7443     | より        | 0.2703     |
| 7  | .         | 0.7398     | 思わ        | 0.2770     |
| 8  | そもそも      | 0.7344     | みたい       | 0.2823     |
| 9  | へ         | 0.7295     | 安倍        | 0.2840     |
| 10 | 個人        | 0.7259     | お         | 0.2892     |
| 11 | ば         | 0.7208     | 笑         | 0.2925     |
| 12 | 情報        | 0.7170     | 旅行        | 0.2935     |
| 13 | 持つ        | 0.7155     | 死者        | 0.2939     |
| 14 | なり        | 0.7076     | 強い        | 0.2939     |
| 15 | 言わ        | 0.7075     | 分かっ       | 0.2966     |
| 16 | 従事        | 0.7068     | 出         | 0.2972     |
| 17 | なん        | 0.7032     | 株         | 0.3139     |
| 18 | 組織        | 0.6985     | 日本        | 0.3159     |
| 19 | ほしい       | 0.6969     | ダメ        | 0.3175     |
| 20 | 宣言        | 0.6933     | 基準        | 0.3205     |
| 21 | です        | 0.6914     | 義務        | 0.3244     |
| 22 | 凄い        | 0.6904     | だろ        | 0.3249     |
| 23 | こと        | 0.6897     | とか        | 0.3253     |
| 24 | や         | 0.6890     | 用         | 0.3272     |
| 25 | 開催        | 0.6877     | そう        | 0.3300     |
| 26 | 受け        | 0.6865     | 済ませ       | 0.3322     |
| 27 | ,         | 0.6836     | 数         | 0.3342     |
| 28 | 研究        | 0.6808     | 出来        | 0.3356     |
| 29 | ず         | 0.6802     | もっと       | 0.3383     |
| 30 | はず        | 0.6693     | 大臣        | 0.3398     |

表 4.5: ME の素性 (スポーツ)

| 順位 | 賛成を得やすい素性 | $\alpha$ 値 | 賛成を得にくい素性 | $\alpha$ 値 |
|----|-----------|------------|-----------|------------|
| 1  | を         | 0.7851     | ?         | 0.2343     |
| 2  | かも        | 0.7656     | わ         | 0.2359     |
| 3  | 目         | 0.7495     | 最後        | 0.2495     |
| 4  | です        | 0.7417     | 明日        | 0.2689     |
| 5  | しれ        | 0.7249     | !         | 0.2751     |
| 6  | として       | 0.7154     | れる        | 0.2926     |
| 7  | せ         | 0.7124     | すぎ        | 0.2928     |
| 8  | 責任        | 0.7081     | 友達        | 0.2959     |
| 9  | あと        | 0.7032     | 欲しい       | 0.2960     |
| 10 | 言っ        | 0.6973     | 先         | 0.3023     |
| 11 | 見苦しい      | 0.6907     | 陣         | 0.3075     |
| 12 | 記事        | 0.6888     | ん         | 0.3079     |
| 13 | 必要        | 0.6858     | 退任        | 0.3132     |
| 14 | その        | 0.6823     | 終わっ       | 0.3161     |
| 15 | と         | 0.6751     | たら        | 0.3189     |
| 16 | スポーツ      | 0.6704     | 引退        | 0.3248     |
| 17 | 1         | 0.6682     | 0         | 0.3253     |
| 18 | 力         | 0.6672     | そう        | 0.3258     |
| 19 | 降板        | 0.6670     | 連戦        | 0.3260     |
| 20 | な         | 0.6660     | ない        | 0.3270     |
| 21 | よう        | 0.6640     | 中日        | 0.3288     |
| 22 | 怪我        | 0.6629     | わから       | 0.3304     |
| 23 | 良い        | 0.6625     | 思っ        | 0.3329     |
| 24 | が         | 0.6615     | 上げ        | 0.3353     |
| 25 | ファン       | 0.6589     | 前         | 0.3361     |
| 26 | られ        | 0.6515     | やる        | 0.3388     |
| 27 | オリックス     | 0.6494     | よ         | 0.3456     |
| 28 | 思う        | 0.6488     | 黒         | 0.3467     |
| 29 | ば         | 0.6481     | だろ        | 0.3474     |
| 30 | じゃ        | 0.6455     | ねえ        | 0.3504     |

表 4.6: ME の素性 (その他)

| 順位 | 賛成を得やすい素性 | $\alpha$ 値 | 賛成を得にくい素性 | $\alpha$ 値 |
|----|-----------|------------|-----------|------------|
| 1  | 懲り        | 0.8662     | また        | 0.2159     |
| 2  | だっ        | 0.7818     | 女         | 0.2375     |
| 3  | 私         | 0.7628     | 思い        | 0.2516     |
| 4  | いい        | 0.7609     | 0         | 0.2619     |
| 5  | 確か        | 0.7327     | ねえ        | 0.2693     |
| 6  | 仕事        | 0.7321     | 笑         | 0.2699     |
| 7  | 中         | 0.7317     | まま        | 0.2748     |
| 8  | 事務所       | 0.7301     | 捨て        | 0.2869     |
| 9  | 男性        | 0.7300     | か         | 0.2879     |
| 10 | について      | 0.7231     | あり        | 0.2960     |
| 11 | たら        | 0.7198     | ?         | 0.3023     |
| 12 | 思う        | 0.7126     | なん        | 0.3026     |
| 13 | 杏         | 0.7090     | だ         | 0.3029     |
| 14 | 費         | 0.7038     | テレ朝       | 0.3088     |
| 15 | ため        | 0.7004     | 昭和        | 0.3162     |
| 16 | わかる       | 0.6946     | やつ        | 0.3172     |
| 17 | 奥さん       | 0.6940     | 道         | 0.3182     |
| 18 | 高橋        | 0.6932     | 回         | 0.3186     |
| 19 | 留美子       | 0.6912     | 欲しい       | 0.3188     |
| 20 | こういう      | 0.6911     | みんな       | 0.3216     |
| 21 | プライベート    | 0.6909     | 配信        | 0.3234     |
| 22 | なら        | 0.6908     | ラム        | 0.3295     |
| 23 | ご         | 0.6905     | 良かっ       | 0.3306     |
| 24 | 元         | 0.6893     | 人間        | 0.3317     |
| 25 | どう        | 0.6802     | 仕方        | 0.3347     |
| 26 | ～         | 0.6801     | ほう        | 0.3355     |
| 27 | なく        | 0.6800     | 今後        | 0.3367     |
| 28 | 出         | 0.6792     | 口         | 0.3373     |
| 29 | 8         | 0.6789     | 頭         | 0.3376     |
| 30 | まし        | 0.6785     | バカ        | 0.3397     |

## 4.1.2 SVMの素性

記事分野別で行ったSVMの素性分析の結果で得られた賛成を得やすい素性、賛成を得にくい素性を表4.7～表4.11に上位30個ずつ載せている。

表 4.7: SVMの素性(政治)

| 順位 | 賛成を得やすい素性 | マージン   | 賛成を得にくい素性 | マージン    |
|----|-----------|--------|-----------|---------|
| 1  | すまん       | 0.8656 | 案         | -0.8302 |
| 2  | まし        | 0.8598 | M子        | -0.7911 |
| 3  | TBS       | 0.8459 | KK        | -0.7640 |
| 4  | いつも       | 0.8459 | 詐欺        | -0.7629 |
| 5  | 欲しく       | 0.8322 | 散々        | -0.7589 |
| 6  | やはり       | 0.8298 | 果実        | -0.7278 |
| 7  | 代         | 0.8264 | 減らし       | -0.7089 |
| 8  | 賢明        | 0.8260 | 酷い        | -0.7075 |
| 9  | 車         | 0.8220 | 近づい       | -0.6984 |
| 10 | 明らか       | 0.8146 | 思わ        | -0.6952 |
| 11 | 資産        | 0.8088 | きっと       | -0.6863 |
| 12 | 前回        | 0.8087 | よかつ       | -0.6552 |
| 13 | はっきり      | 0.7999 | くれ        | -0.6463 |
| 14 | 決断        | 0.7982 | .         | -0.6368 |
| 15 | 僕         | 0.7949 | また        | -0.6179 |
| 16 | 答え        | 0.7931 | わ         | -0.6120 |
| 17 | 行政        | 0.7900 | 大人しく      | -0.5843 |
| 18 | お疲れ様      | 0.7838 | 行わ        | -0.5781 |
| 19 | 札幌        | 0.7827 | 集         | -0.5775 |
| 20 | 感謝        | 0.7818 | 困っ        | -0.5765 |
| 21 | おお        | 0.7798 | 投稿        | -0.5711 |
| 22 | くる        | 0.7758 | さっさと      | -0.5632 |
| 23 | 為         | 0.7718 | 失脚        | -0.5569 |
| 24 | 上         | 0.7717 | 特例        | -0.5564 |
| 25 | 免除        | 0.7715 | 余地        | -0.5533 |
| 26 | 耳         | 0.7691 | 公共        | -0.5516 |
| 27 | 国家        | 0.7690 | 馬鹿        | -0.5509 |
| 28 | 指示        | 0.7685 | ゆっくり      | -0.5447 |
| 29 | 学芸        | 0.7685 | 早く        | -0.5402 |
| 30 | 各社        | 0.7666 | 場         | -0.5302 |

表 4.8: SVM の素性 (事件)

| 順位 | 賛成を得やすい素性 | マージン   | 賛成を得にくい素性 | マージン    |
|----|-----------|--------|-----------|---------|
| 1  | さん        | 0.4454 | 事実        | -0.4753 |
| 2  | を         | 0.4154 | 酷         | -0.3629 |
| 3  | 運転        | 0.3990 | もん        | -0.3576 |
| 4  | 車         | 0.3865 | よる        | -0.3565 |
| 5  | トヨタ       | 0.3663 | ほしい       | -0.3515 |
| 6  | 罪         | 0.3631 | 想像        | -0.3499 |
| 7  | 本当に       | 0.3302 | 共犯        | -0.3469 |
| 8  | かも        | 0.3199 | 極刑        | -0.3452 |
| 9  | 胸糞        | 0.3186 | 安かつ       | -0.3171 |
| 10 | ね         | 0.3142 | とか        | -0.3168 |
| 11 | 最近        | 0.3025 | やろ        | -0.3161 |
| 12 | 鑑定        | 0.2886 | 起き        | -0.3159 |
| 13 | なっ        | 0.2847 | トップ       | -0.3156 |
| 14 | 思う        | 0.2781 | へ         | -0.3039 |
| 15 | しれ        | 0.2735 | やっ        | -0.2998 |
| 16 | 癖         | 0.2728 | せ         | -0.2987 |
| 17 | 思っ        | 0.2723 | tktube    | -0.2905 |
| 18 | 過ぎる       | 0.2710 | くる        | -0.2870 |
| 19 | もの        | 0.2674 | 何         | -0.2863 |
| 20 | モザイク      | 0.2634 | 新幹線       | -0.2765 |
| 21 | やめ        | 0.2607 | 変わら       | -0.2712 |
| 22 | い         | 0.2581 | スタッフ      | -0.2700 |
| 23 | 増え        | 0.2577 | 無い        | -0.2690 |
| 24 | させ        | 0.2512 | 相談        | -0.2682 |
| 25 | 内輪        | 0.2511 | って        | -0.2668 |
| 26 | 重い        | 0.2489 | 買い物       | -0.2625 |
| 27 | 作成        | 0.2453 | 書く        | -0.2545 |
| 28 | むご        | 0.2344 | きちんと      | -0.2545 |
| 29 | 意味        | 0.2342 | 見         | -0.2524 |
| 30 | なかつ       | 0.2337 | ・         | -0.2493 |

表 4.9: SVM の素性 (病気)

| 順位 | 賛成を得やすい素性 | マージン   | 賛成を得にくい素性 | マージン    |
|----|-----------|--------|-----------|---------|
| 1  | 絶対        | 0.4227 | 日本人       | -0.4691 |
| 2  | ばかり       | 0.4220 | 分かっ       | -0.4471 |
| 3  | 関係        | 0.3929 | 推奨        | -0.4182 |
| 4  | 札幌        | 0.3818 | また        | -0.3998 |
| 5  | そもそも      | 0.3795 | 安部        | -0.3970 |
| 6  | 投稿        | 0.3742 | 株         | -0.3968 |
| 7  | 試算        | 0.3670 | たい        | -0.3859 |
| 8  | 持っ        | 0.3668 | 笑         | -0.3716 |
| 9  | 良く        | 0.3613 | 基準        | -0.3674 |
| 10 | として       | 0.3597 | 死者        | -0.3535 |
| 11 | 今更        | 0.3572 | ほとんど      | -0.3532 |
| 12 | でき        | 0.3548 | お         | -0.3480 |
| 13 | へ         | 0.3536 | 高い        | -0.3443 |
| 14 | ほしい       | 0.3408 | 強い        | -0.3391 |
| 15 | しっかり      | 0.3397 | より        | -0.3323 |
| 16 | あり        | 0.3376 | 旅行        | -0.3279 |
| 17 | なり        | 0.3294 | 思わ        | -0.3233 |
| 18 | 組織        | 0.3281 | 拡大        | -0.3214 |
| 19 | 受け        | 0.3232 | みたい       | -0.3203 |
| 20 | 従事        | 0.3215 | そう        | -0.3172 |
| 21 | はっきり      | 0.3213 | ダメ        | -0.3158 |
| 22 | 総裁        | 0.3186 | 昭和        | -0.3041 |
| 23 | なん        | 0.3136 | ぶり        | -0.3038 |
| 24 | 個人        | 0.3127 | 武漢        | -0.3038 |
| 25 | .         | 0.3124 | とか        | -0.2977 |
| 26 | 変わる       | 0.2952 | 用         | -0.2885 |
| 27 | によって      | 0.2950 | やる        | -0.2883 |
| 28 | 情報        | 0.2945 | かわいそう     | -0.2861 |
| 29 | 招く        | 0.2922 | 続き        | -0.2857 |
| 30 | 正気        | 0.2861 | 報告        | -0.2853 |

表 4.10: SVM の素性 (スポーツ)

| 順位 | 賛成を得やすい素性 | $\alpha$ 値 | 賛成を得にくい素性 | $\alpha$ 値 |
|----|-----------|------------|-----------|------------|
| 1  | うーん       | 0.4337     | 友達        | -0.4700    |
| 2  | 目         | 0.4127     | 明日        | -0.4673    |
| 3  | を         | 0.4023     | 最後        | -0.4135    |
| 4  | かも        | 0.3962     | ?         | -0.4084    |
| 5  | として       | 0.3904     | わ         | -0.4047    |
| 6  | 記事        | 0.3565     | 先         | -0.3631    |
| 7  | お疲れさま     | 0.3528     | 退任        | -0.3610    |
| 8  | です        | 0.3480     | 遅         | -0.3252    |
| 9  | しれ        | 0.3454     | 引退        | -0.3220    |
| 10 | 降板        | 0.3384     | 0         | -0.3215    |
| 11 | 氏         | 0.3382     | ねえ        | -0.3159    |
| 12 | 必要        | 0.3378     | 思っ        | -0.3124    |
| 13 | 見苦しい      | 0.3361     | !         | -0.3107    |
| 14 | 定義        | 0.3252     | 黒         | -0.3034    |
| 15 | せ         | 0.3241     | 欲しい       | -0.3027    |
| 16 | スポーツ      | 0.3216     | すぎ        | -0.3021    |
| 17 | しか        | 0.3195     | 現役        | -0.3007    |
| 18 | あと        | 0.3010     | わから       | -0.2998    |
| 19 | 責任        | 0.2988     | 終わっ       | -0.2967    |
| 20 | 我慢        | 0.2828     | 馬場        | -0.2909    |
| 21 | 言っ        | 0.2766     | 問題        | -0.2899    |
| 22 | そして       | 0.2742     | 陣         | -0.2789    |
| 23 | 打撃        | 0.2734     | コイツ       | -0.2786    |
| 24 | なんと       | 0.2601     | れる        | -0.2774    |
| 25 | 直近        | 0.2601     | お疲れ様      | -0.2709    |
| 26 | 確信        | 0.2599     | ん         | -0.2704    |
| 27 | 相撲        | 0.2539     | 訳         | -0.2694    |
| 28 | 人材        | 0.2529     | 内閣        | -0.2620    |
| 29 | ファン       | 0.2524     | 千賀        | -0.2612    |
| 30 | 力         | 0.2434     | なん        | -0.2607    |

表 4.11: SVM の素性 (その他)

| 順位 | 賛成を得やすい素性 | マージン   | 賛成を得にくい素性 | マージン    |
|----|-----------|--------|-----------|---------|
| 1  | ご         | 0.8024 | ブブー       | -0.5573 |
| 2  | 懲り        | 0.7093 | 奴         | -0.4979 |
| 3  | 確か        | 0.6353 | ねえ        | -0.4863 |
| 4  | ...       | 0.5547 | また        | -0.4686 |
| 5  | 多分        | 0.4425 | どん        | -0.4667 |
| 6  | 男性        | 0.4421 | 女         | -0.4629 |
| 7  | 事務所       | 0.4272 | 思い        | -0.4556 |
| 8  | 仕事        | 0.4233 | 笑う        | -0.4374 |
| 9  | いや        | 0.4036 | 良かつ       | -0.4220 |
| 10 | だっ        | 0.3892 | バカ        | -0.4183 |
| 11 | どう        | 0.3866 | 笑         | -0.3855 |
| 12 | すげ        | 0.3865 | でっ        | -0.3824 |
| 13 | 早い        | 0.3797 | 0         | -0.3747 |
| 14 | 養育        | 0.3643 | とけ        | -0.3614 |
| 15 | 中         | 0.3642 | 下         | -0.3530 |
| 16 | について      | 0.3615 | みんな       | -0.3389 |
| 17 | いい        | 0.3576 | かわいそう     | -0.3378 |
| 18 | 演技        | 0.3558 | 病気        | -0.3375 |
| 19 | うーん       | 0.3491 | 結婚        | -0.3358 |
| 20 | 出来る       | 0.3469 | あり        | -0.3305 |
| 21 | 杏         | 0.3373 | スキャンダル    | -0.3301 |
| 22 | モデル       | 0.3352 | 純粹        | -0.3286 |
| 23 | 思う        | 0.3347 | 捨て        | -0.3249 |
| 24 | 留美子       | 0.3335 | まま        | -0.3249 |
| 25 | 私         | 0.3320 | 見         | -0.3227 |
| 26 | 奥さん       | 0.3297 | なん        | -0.3199 |
| 27 | わかる       | 0.3288 | 自己        | -0.3129 |
| 28 | 8         | 0.3211 | 弁護        | -0.3129 |
| 29 | こういう      | 0.3192 | だ         | -0.3070 |
| 30 | ～         | 0.3185 | 蜜         | -0.3041 |

## 4.2 数値評価

本節では使用した3種類の数値評価についての説明と結果を記述する。

### 4.2.1 数値評価1

数値評価1では高評価から低評価をひいて、コメント時刻で割ることで数値を求めた。数値評価1は高評価100以上のコメントを対象に数値評価を求めた。数値評価0未満を低評価、0以上5未満を中評価、5以上を高評価としたものを数値評価1-1とし、数値評価0未満を低評価、0以上10未満を中評価、10以上を高評価としたものを数値評価1-2とした。上記の定義に基づいて分類した結果、数値評価1-1は高評価、低評価のコメントをそれぞれランダムに22,540文ずつ抜き出した。そこから高評価のコメントと低評価のコメントを11,270文ずつ抜きだした。そして抜き出したコメントを足すことで高評価と低評価のコメントの数が同じとなるデータが2つ得られた。片方を学習データ、もう片方をテストデータとして機械学習に入力して実験を行った。同様に数値評価1-2も高評価、低評価のコメントをそれぞれランダムに14,216文ずつ抜き出した。そこから高評価のコメントと低評価のコメントを7,108文ずつ抜きだした。そして抜き出したコメントを足すことで高評価と低評価のコメントの数が同じとなるデータが2つ得られた。片方を学習データ、もう片方をテストデータとして機械学習に入力して実験を行った。

### 4.2.2 数値評価2

数値評価2では高評価を高評価と低評価の和で割ることで数値評価を求めた。数値評価2は高評価+低評価が10以上のコメントを対象に数値評価を求めた。数値評価0.2未満を低評価、0.2以上0.9未満を中評価、0.9以上を高評価と定義した。上記の定義に基づいて分類した結果、高評価が29,685文、中評価が33,633文、低評価が1,336文となった。そこから高評価と低評価をランダムに1,300文ずつ抜き出した。そこから高評価のコメントと低評価のコメントを650文ずつ抜き出した。そして抜き出したコメントを足すことで高評価と低評価のコメントの数が同じとなるデータが2つ得られた。片方を学習データ、もう片方をテストデータとして機械学習に入力して実験を行った。

### 4.2.3 数値評価3

数値評価3では高評価をコメント時刻で割ることで数値評価を求めた。値評価3は高評価が10以上のコメントを対象に数値評価を求めた。数日前に投稿されたコメントは日数×24、数時間前のコメントはその時間+1、数十分前のコメントは1時間前のコメントとすることでコメントの投稿時間の単位を同じにした。数値評価1未満のコメントを低評価、1以上10未満のコメントを中評価、10以上のコメントを高評価と定義した。上記の定義に基づいて分類した結果、高評価17,657文、中評価26,691文、低評価6,389文となった。そこから高評価と低評価をランダムに6,000文ずつ抜き出した。そこから高評価のコメントと低評価のコメントを3,000文ずつ抜き出した。そして抜き出したコメントを足すことで高評価と低評価のコメントの数が同じとなるデータが2つ得られた。片方を学習データ、もう片方をテストデータとして機械学習に入力して実験を行った

### 4.2.4 数値評価の性能

表 4.12: 機械学習の性能評価

| 分類   | 数値評価 1-1 | 数値評価 1-2 | 数値評価 2 | 数値評価 3 |
|------|----------|----------|--------|--------|
| ME   | 0.6679   | 0.8224   | 0.6277 | 0.6332 |
| SVM  | 0.5735   | 0.7247   | 0.6123 | 0.6275 |
| BERT | 0.6879   | 0.7936   | 0.6215 | 0.6418 |

表 4.12 より数値評価 1-1 は高評価の範囲を 5 以上と設定し、数値評価 1-2 は高評価の範囲を 10 以上とした結果、数値評価 1-2 は数値評価 1-1 と比較して ME, SVM は約 15% , BERT は約 10% 性能が向上している。数値評価 2, 数値評価 3 は全ての機械学習で性能が 6 割強となった。

### 4.2.5 ME 素性

数値評価で行った ME の素性分析の結果で得られた賛成を得やすい素性、賛成を得にくい素性を表 4.13～表 4.16 に上位 30 個ずつ載せている。

表 4.13: ME の素性 (数値評価 1-1)

| 順位 | 賛成を得やすい素性 | $\alpha$ 値 | 賛成を得にくい素性 | $\alpha$ 値 |
|----|-----------|------------|-----------|------------|
| 1  | 象徴        | 0.9594     | もり        | 0.0769     |
| 2  | 録音        | 0.9392     | 外人        | 0.0787     |
| 3  | 面会        | 0.9262     | 関心        | 0.0905     |
| 4  | ポジション     | 0.9033     | お幸せ       | 0.0914     |
| 5  | ゾンビ       | 0.8996     | ボク        | 0.0940     |
| 6  | 松葉杖       | 0.8971     | 直す        | 0.0944     |
| 7  | 果たして      | 0.8940     | おじさん      | 0.0967     |
| 8  | 地震        | 0.8909     | 緩め        | 0.0983     |
| 9  | 正常        | 0.8899     | 不能        | 0.1011     |
| 10 | 田嶋        | 0.8890     | ほっとい      | 0.1017     |
| 11 | 宮家        | 0.8869     | リセット      | 0.1040     |
| 12 | 林         | 0.8830     | 姫         | 0.1073     |
| 13 | 自費        | 0.8828     | セクハラ      | 0.1086     |
| 14 | ワクパス      | 0.8823     | 懐かしい      | 0.1137     |
| 15 | 未熟        | 0.8808     | 倍増        | 0.1138     |
| 16 | 実証        | 0.8777     | 悪意        | 0.1145     |
| 17 | ロイヤル      | 0.8776     | どこぞ       | 0.1171     |
| 18 | 山下        | 0.8761     | 矛先        | 0.1171     |
| 19 | 隣         | 0.8733     | 素行        | 0.1211     |
| 20 | 同感        | 0.8725     | そっと       | 0.1214     |
| 21 | 正気        | 0.8706     | 解釈        | 0.1226     |
| 22 | バックアップ    | 0.8695     | ありがたい     | 0.1240     |
| 23 | 震災        | 0.8674     | 口出し       | 0.1270     |
| 24 | 綺麗事       | 0.8639     | おっさん      | 0.1294     |
| 25 | 累進        | 0.8636     | 規律        | 0.1294     |
| 26 | ツイート      | 0.8616     | EU        | 0.1294     |
| 27 | 知人        | 0.8602     | 使い道       | 0.1316     |
| 28 | お陰        | 0.8586     | 許容        | 0.1335     |
| 29 | うっ        | 0.8552     | 手元        | 0.1391     |
| 30 | 湯水        | 0.8544     | イイ        | 0.1393     |

表 4.14: ME の素性 (数値評価 1-2)

| 順位 | 賛成を得やすい素性 | α 値    | 賛成を得にくい素性 | α 値    |
|----|-----------|--------|-----------|--------|
| 1  | 速報        | 0.9488 | 比べれ       | 0.0776 |
| 2  | 重み        | 0.9196 | 詳しく       | 0.0976 |
| 3  | 腹立つ       | 0.9157 | 泥棒        | 0.1033 |
| 4  | 自動車       | 0.9019 | 長友        | 0.1069 |
| 5  | ミヤノ       | 0.8991 | 騒い        | 0.1126 |
| 6  | 得意        | 0.8975 | かっこいい     | 0.1142 |
| 7  | 咎め        | 0.8960 | o         | 0.1148 |
| 8  | アキレス腱     | 0.8921 | 関心        | 0.1216 |
| 9  | 帰国        | 0.8919 | 外す        | 0.1217 |
| 10 | 寄付        | 0.8878 | 神戸        | 0.1228 |
| 11 | 録音        | 0.8823 | 間近        | 0.1251 |
| 12 | 世襲        | 0.8740 | 真子        | 0.1305 |
| 13 | 撃っ        | 0.8713 | ゴミ        | 0.1348 |
| 14 | 時給        | 0.8695 | 抑え        | 0.1356 |
| 15 | 立         | 0.8661 | 放っ        | 0.1362 |
| 16 | あまりに      | 0.8658 | お父さん      | 0.1402 |
| 17 | だに        | 0.8654 | 河野        | 0.1442 |
| 18 | 浜         | 0.8629 | 失業        | 0.1457 |
| 19 | 免除        | 0.8629 | ぬるい       | 0.1461 |
| 20 | 逃げる       | 0.8615 | だいぶ       | 0.1473 |
| 21 | 心身        | 0.8601 | 心境        | 0.1473 |
| 22 | 捻出        | 0.8587 | 組閣        | 0.1479 |
| 23 | ロイヤル      | 0.8562 | 横         | 0.1482 |
| 24 | 天気        | 0.8552 | モンゴル      | 0.1484 |
| 25 | 内親王       | 0.8547 | 許す        | 0.1486 |
| 26 | 制定        | 0.8546 | チンパンジー    | 0.1492 |
| 27 | とおり       | 0.8537 | 癖         | 0.1516 |
| 28 | 報告        | 0.8509 | 白い        | 0.1541 |
| 29 | うつさ       | 0.8506 | 森友        | 0.1551 |
| 30 | 柔道        | 0.8486 | 否         | 0.1553 |

表 4.15: ME の素性 (数値評価 2)

| 順位 | 賛成を得やすい素性 | $\alpha$ 値 | 賛成を得にくい素性 | $\alpha$ 値 |
|----|-----------|------------|-----------|------------|
| 1  | 愛子        | 0.8875     | コメント      | 0.1116     |
| 2  | 誕生        | 0.8633     | 反対        | 0.1931     |
| 3  | 感じ        | 0.8359     | そう        | 0.1966     |
| 4  | 後         | 0.8254     | す         | 0.2061     |
| 5  | 結局        | 0.8063     | 未だに       | 0.2143     |
| 6  | き         | 0.7941     | 中傷        | 0.2159     |
| 7  | 皇室        | 0.7913     | ヤフコメ      | 0.2182     |
| 8  | 全て        | 0.7784     | 期待        | 0.2205     |
| 9  | について      | 0.7672     | ファイザー     | 0.2221     |
| 10 | 天皇        | 0.7652     | 批判        | 0.2223     |
| 11 | 娘         | 0.7642     | やつ        | 0.2260     |
| 12 | 時間        | 0.7523     | 意見        | 0.2315     |
| 13 | 最初        | 0.7522     | マスコミ      | 0.2315     |
| 14 | 今         | 0.7497     | 家庭        | 0.2337     |
| 15 | こんな       | 0.7494     | ここ        | 0.2338     |
| 16 | なあ        | 0.7444     | 賛成        | 0.2350     |
| 17 | 一律        | 0.7438     | くれ        | 0.2393     |
| 18 | 違う        | 0.7375     | たら        | 0.2407     |
| 19 | 労働        | 0.7350     | 早く        | 0.2422     |
| 20 | 9         | 0.7345     | 感じる       | 0.2459     |
| 21 | 加減        | 0.7333     | 自ら        | 0.2465     |
| 22 | もちろん      | 0.7296     | 訳         | 0.2471     |
| 23 | 全         | 0.7289     | 思う        | 0.2477     |
| 24 | 儀式        | 0.7260     | として       | 0.2542     |
| 25 | 会見        | 0.7259     | いい        | 0.2611     |
| 26 | 回         | 0.7239     | なに        | 0.2617     |
| 27 | 監督        | 0.7167     | 変え        | 0.2626     |
| 28 | 人権        | 0.7165     | あげ        | 0.2639     |
| 29 | 給付        | 0.7159     | 出         | 0.2655     |
| 30 | 被害        | 0.7156     | 誹謗        | 0.2676     |

表 4.16: ME の素性 (数値評価 3)

| 順位 | 賛成を得やすい素性 | $\alpha$ 値 | 賛成を得にくい素性 | $\alpha$ 値 |
|----|-----------|------------|-----------|------------|
| 1  | 初めて       | 0.8833     | ヤフコメ      | 0.1009     |
| 2  | 儀式        | 0.8821     | トヨタ       | 0.1043     |
| 3  | お子さん      | 0.8798     | 見舞い       | 0.1119     |
| 4  | 年齢        | 0.8796     | 休み        | 0.1163     |
| 5  | 評論        | 0.8735     | 病気        | 0.1372     |
| 6  | 認め        | 0.8705     | 日大        | 0.1431     |
| 7  | 上皇        | 0.8649     | メンバー      | 0.1512     |
| 8  | 皇族        | 0.8510     | とい        | 0.1519     |
| 9  | 機能        | 0.8450     | 関心        | 0.1526     |
| 10 | 操作        | 0.8443     | 殺害        | 0.1573     |
| 11 | 永住        | 0.8427     | 漫画        | 0.1614     |
| 12 | 駄目        | 0.8381     | 人口        | 0.1648     |
| 13 | 無事        | 0.8360     | 車検        | 0.1665     |
| 14 | 勘違い       | 0.8297     | 福岡        | 0.1669     |
| 15 | 前後        | 0.8283     | 紀子        | 0.1683     |
| 16 | 甘く        | 0.8272     | 孫         | 0.1686     |
| 17 | 一時        | 0.8271     | 始め        | 0.1688     |
| 18 | GO TO     | 0.8251     | 見当たらず     | 0.1693     |
| 19 | 全国        | 0.8227     | 自転車       | 0.1728     |
| 20 | おめでとう     | 0.8206     | 受け入れ      | 0.1740     |
| 21 | 弁護士       | 0.8200     | 奴         | 0.1752     |
| 22 | カッコ       | 0.8196     | 別れ        | 0.1754     |
| 23 | 付き合い      | 0.8195     | 至る        | 0.1759     |
| 24 | 数々        | 0.8192     | 行ける       | 0.1785     |
| 25 | まくっ       | 0.8181     | プラス       | 0.1801     |
| 26 | 本当に       | 0.8176     | 長友        | 0.1805     |
| 27 | 時期        | 0.8169     | 極         | 0.1812     |
| 28 | にくい       | 0.8167     | 軽く        | 0.1813     |
| 29 | 反論        | 0.8161     | いかが       | 0.1852     |
| 30 | モデルナ      | 0.8145     | 韓国        | 0.1865     |

## 4.2.6 SVM 素性

数値評価で行った SVM の素性分析の結果で得られた賛成を得やすい素性，賛成を得にくい素性を表 4.17～表 4.20 に上位 30 個ずつ載せている。

表 4.17: SVM の素性 (数値評価 1-1)

| 順位 | 賛成を得やすい素性 | マージン   | 賛成を得にくい素性 | マージン    |
|----|-----------|--------|-----------|---------|
| 1  | メモ        | 1.6523 | おじさん      | -3.5276 |
| 2  | お陰        | 1.5812 | おこなっ      | -2.9175 |
| 3  | ワクパス      | 1.4760 | しかるべき     | -2.8779 |
| 4  | 録音        | 1.4384 | 悪意        | -2.8594 |
| 5  | 消極        | 1.3829 | 解釈        | -2.7682 |
| 6  | 橋         | 1.3521 | 苦しく       | -2.6584 |
| 7  | 理事        | 1.3261 | 緩め        | -2.6344 |
| 8  | 失敬        | 1.3017 | 騒い        | -2.4271 |
| 9  | 在庫        | 1.2798 | 長時間       | -2.4180 |
| 10 | ミリ        | 1.2560 | 関心        | -2.3980 |
| 11 | 様子        | 1.2099 | 壊滅        | -2.3803 |
| 12 | 小便        | 1.2068 | 勇気        | -2.3681 |
| 13 | 黒い        | 1.1871 | 騒ぎ立てる     | -2.3630 |
| 14 | らし        | 1.1698 | 答弁        | -2.3565 |
| 15 | 低かっ       | 1.1221 | 邪魔        | -2.3545 |
| 16 | 同感        | 1.1126 | 不在        | -2.3487 |
| 17 | 土田        | 1.0988 | 不能        | -2.3407 |
| 18 | 葬ら        | 1.0861 | 無能        | -2.3391 |
| 19 | 思い入れ      | 1.0742 | ベラ        | -2.3144 |
| 20 | あほらしく     | 1.0727 | 顔出し       | -2.3057 |
| 21 | 勝た        | 1.0698 | 土俵        | -2.3000 |
| 22 | 累進        | 1.0639 | スキ        | -2.2869 |
| 23 | 地震        | 1.0576 | 望ま        | -2.2839 |
| 24 | 悪化        | 1.0443 | 外人        | -2.2799 |
| 25 | ゾンビ       | 1.0385 | さらに       | -2.2708 |
| 26 | 象徴        | 1.0288 | ありがたい     | -2.2595 |
| 27 | 危害        | 1.0212 | 入り口       | -2.2509 |
| 28 | 悪質        | 1.0178 | ポスト       | -2.2499 |
| 29 | 知人        | 1.0174 | 矛先        | -2.2382 |
| 30 | 人体        | 1.0090 | 垂れ流し      | -2.2378 |

表 4.18: SVM の素性 (数値評価 1-2)

| 順位 | 賛成を得やすい素性 | マージン   | 賛成を得にくい素性 | マージン    |
|----|-----------|--------|-----------|---------|
| 1  | 速報        | 0.9869 | 失業        | -2.2296 |
| 2  | 負傷        | 0.7582 | 見本        | -2.1928 |
| 3  | 録音        | 0.7199 | ライバル      | -2.0978 |
| 4  | 腹立つ       | 0.6899 | 泥棒        | -2.0731 |
| 5  | 控えめ       | 0.6133 | だいぶ       | -1.9853 |
| 6  | 軽井沢       | 0.5956 | 騒い        | -1.9775 |
| 7  | 失敬        | 0.5761 | 蔑ろ        | -1.9599 |
| 8  | ワクパス      | 0.5574 | 比べれ       | -1.9566 |
| 9  | ミヤネ       | 0.5027 | しかた       | -1.9137 |
| 10 | 正に        | 0.4990 | わがまま      | -1.9128 |
| 11 | 心身        | 0.4351 | おそろしい     | -1.9114 |
| 12 | 捻出        | 0.3957 | 待た        | -1.8979 |
| 13 | アキレス腱     | 0.3956 | 付く        | -1.8849 |
| 14 | 旨み        | 0.3842 | 詳しく       | -1.8802 |
| 15 | 得意        | 0.3793 | ポ         | -1.8773 |
| 16 | 消極        | 0.3787 | 関心        | -1.8727 |
| 17 | 始まり       | 0.3732 | 石破        | -1.8403 |
| 18 | 番長        | 0.3673 | カラダ       | -1.8276 |
| 19 | 揉め        | 0.3662 | 栄冠        | -1.8276 |
| 20 | 重み        | 0.3586 | C S       | -1.8271 |
| 21 | 闊歩        | 0.3477 | 担い手       | -1.8260 |
| 22 | 中央        | 0.3438 | 森友        | -1.8135 |
| 23 | 妊婦        | 0.3215 | 徹底        | -1.8134 |
| 24 | 撃っ        | 0.3118 | 神聖        | -1.8112 |
| 25 | とおり       | 0.3067 | 市場        | -1.8090 |
| 26 | 珍しい       | 0.3012 | シングル      | -1.8061 |
| 27 | 恵子        | 0.3011 | p c c     | -1.8024 |
| 28 | 寄付        | 0.2929 | バリ        | -1.8004 |
| 29 | 吸わ        | 0.2884 | かっこいい     | -1.7950 |
| 30 | 撃た        | 0.2763 | 外す        | -1.7925 |

表 4.19: SVM の素性 (数値評価 2)

| 順位 | 賛成を得やすい素性 | マージン   | 賛成を得にくい素性 | マージン    |
|----|-----------|--------|-----------|---------|
| 1  | 誕生        | 1.4349 | 反対        | -1.2551 |
| 2  | 愛子        | 1.1589 | 賛成        | -1.0869 |
| 3  | T B S     | 1.0231 | 一気に       | -1.0845 |
| 4  | P T S D   | 0.9587 | 回忌        | -1.0845 |
| 5  | 都合        | 0.9320 | 脱         | -1.0845 |
| 6  | そ         | 0.9155 | 非国民       | -1.0605 |
| 7  | えー        | 0.9155 | 気の毒       | -1.0359 |
| 8  | えっ        | 0.9155 | ファイザー     | -1.0317 |
| 9  | 最悪        | 0.9155 | よかつ       | -1.0182 |
| 10 | 働か        | 0.9155 | やり        | -1.0152 |
| 11 | 放置        | 0.9155 | イケメンアンチ   | -0.9517 |
| 12 | T o o     | 0.9155 | b i g     | -0.9516 |
| 13 | 選択肢       | 0.9155 | 何でもかんでも   | -0.9516 |
| 14 | 落武者       | 0.9155 | 意見        | -0.9381 |
| 15 | 誇らしく      | 0.9155 | 8         | -0.9362 |
| 16 | について      | 0.8846 | だから       | -0.9316 |
| 17 | 結局        | 0.8712 | フェアプレー    | -0.9272 |
| 18 | 地元        | 0.8542 | ゲレーロ      | -0.9271 |
| 19 | 腰抜け       | 0.8508 | 誇る        | -0.9043 |
| 20 | 違う        | 0.8507 | 家庭        | -0.8928 |
| 21 | やっど       | 0.8393 | コメント      | -0.8791 |
| 22 | .....     | 0.8210 | 反         | -0.8767 |
| 23 | 風邪        | 0.8194 | 立派        | -0.8682 |
| 24 | もちろん      | 0.8188 | 愛         | -0.8648 |
| 25 | 悩ん        | 0.8163 | 未だに       | -0.8528 |
| 26 | 儀式        | 0.7789 | 発表        | -0.8267 |
| 27 | 加減        | 0.7774 | 見解        | -0.8079 |
| 28 | 付き合お      | 0.7694 | くれ        | -0.8049 |
| 29 | 母子        | 0.7690 | 事細か       | -0.7925 |
| 30 | G O       | 0.7661 | 応援        | -0.7880 |

表 4.20: SVM の素性 (数値評価 3)

| 順位 | 賛成を得やすい素性 | マージン   | 賛成を得にくい素性 | マージン    |
|----|-----------|--------|-----------|---------|
| 1  | 永住        | 0.6688 | ヤフコメ      | -2.5939 |
| 2  | 甘く        | 0.4742 | 動員        | -2.3653 |
| 3  | 駄目        | 0.4637 | 休み        | -2.3421 |
| 4  | 認め        | 0.4148 | 殺害        | -2.3407 |
| 5  | 機能        | 0.3946 | 自転車       | -2.2839 |
| 6  | オリックス     | 0.3941 | うわ        | -2.2808 |
| 7  | おめでとう     | 0.3841 | 千賀        | -2.2714 |
| 8  | ついに       | 0.3723 | 見舞い       | -2.2428 |
| 9  | 数々        | 0.3476 | 見当たらず     | -2.1993 |
| 10 | 大河ドラマ     | 0.3147 | トヨタ       | -2.1845 |
| 11 | 〉         | 0.3108 | 至る        | -2.1731 |
| 12 | 11月       | 0.3007 | 軽く        | -2.1612 |
| 13 | カッコ       | 0.2886 | 田舎        | -2.1396 |
| 14 | 上皇        | 0.2882 | すき        | -2.1325 |
| 15 | 年齢        | 0.2609 | 極         | -2.1291 |
| 16 | 評論        | 0.2467 | 大差        | -2.1168 |
| 17 | 自殺        | 0.2392 | 車検        | -2.1149 |
| 18 | 内親王       | 0.2382 | 情けなく      | -2.1059 |
| 19 | 操作        | 0.2378 | 日大        | -2.0978 |
| 20 | 一気に       | 0.2301 | お孫さん      | -2.0974 |
| 21 | 正気        | 0.2181 | きる        | -2.0778 |
| 22 | お子さん      | 0.2037 | 猛々しい      | -2.0734 |
| 23 | 冗談        | 0.2024 | メンバー      | -2.0722 |
| 24 | 愛知        | 0.1981 | 具合        | -2.0701 |
| 25 | 遅し        | 0.1928 | 病気        | -2.0680 |
| 26 | モデルナ      | 0.1919 | 未来        | -2.0658 |
| 27 | 聞か        | 0.1905 | ぬるい       | -2.0631 |
| 28 | 今更        | 0.1745 | 引きずっ      | -2.0614 |
| 29 | 熱         | 0.1725 | 人口        | -2.0593 |
| 30 | 付き合い      | 0.1679 | いかが       | -2.0570 |

## 4.2.7 BERTの素性

数値評価で行ったBERTの素性分析の結果で得られた賛成を得やすい素性，賛成を得にくい素性を表4.21～表4.24に上位30個ずつ載せている。

表 4.21: BERTの素性(数値評価1-1)

| 順位 | 賛成を得やすい素性 | 数値     | 賛成を得にくい素性   | 数値     |
|----|-----------|--------|-------------|--------|
| 1  | 段取り       | 0.9930 | 甘んじる        | 0.0089 |
| 2  | やむを得ず     | 0.9924 | もちい         | 0.0090 |
| 3  | 秋篠        | 0.9920 | アベサポ        | 0.0093 |
| 4  | やむを得      | 0.9919 | アベノマスク      | 0.0093 |
| 5  | やむを得ない    | 0.9918 | 休まん         | 0.0094 |
| 6  | 逆撫で       | 0.9912 | あんぽんたん      | 0.0094 |
| 7  | 生む        | 0.9908 | サッカーガチファン   | 0.0095 |
| 8  | 録音        | 0.9908 | 開き直り        | 0.0095 |
| 9  | 見栄っ張り     | 0.9907 | 休まれる        | 0.0097 |
| 10 | ホロコースト    | 0.9906 | グローバルロックダウン | 0.0097 |
| 11 | やむをえ      | 0.9901 | サッカーサポ      | 0.0097 |
| 12 | 着替え       | 0.9899 | 休まる         | 0.0097 |
| 13 | 田嶋        | 0.9898 | ハマる         | 0.0098 |
| 14 | 転げ落ち      | 0.9897 | 甘える         | 0.0098 |
| 15 | 更正        | 0.9895 | 甘やかし        | 0.0098 |
| 16 | 私的        | 0.9894 | 打てる         | 0.0099 |
| 17 | ひな壇       | 0.9894 | 休まら         | 0.0099 |
| 18 | カーブミラー    | 0.9891 | 休めれ         | 0.0099 |
| 19 | 違約        | 0.9890 | つるし上げ       | 0.0100 |
| 20 | 篠宮        | 0.9889 | おかしかっ       | 0.0100 |
| 21 | 累進        | 0.9888 | とりまき        | 0.0100 |
| 22 | 宮内庁       | 0.9887 | 花開い         | 0.0101 |
| 23 | 略儀        | 0.9885 | 開き直っ        | 0.0101 |
| 24 | ゃいいのにな    | 0.9885 | 甘えん坊        | 0.0101 |
| 25 | 更衣        | 0.9883 | 甘やかす        | 0.0101 |
| 26 | 社会保険庁     | 0.9881 | 噛み付い        | 0.0102 |
| 27 | 検察庁       | 0.9881 | 開き直れ        | 0.0102 |
| 28 | 親権        | 0.9879 | 羨ましかっ       | 0.0102 |
| 29 | 議連        | 0.9879 | うれしかっ       | 0.0102 |
| 30 | 宮家        | 0.9879 | 大きけれ        | 0.0103 |

表 4.22: BERT の素性 (数値評価 1-2)

| 順位 | 賛成を得やすい素性 | 数値     | 賛成を得にくい素性 | 数値     |
|----|-----------|--------|-----------|--------|
| 1  | 本当に       | 0.9918 | 賛成        | 0.0082 |
| 2  | 宮         | 0.9878 | 公明党       | 0.0092 |
| 3  | ようやく      | 0.9873 | ユーロ       | 0.0095 |
| 4  | まさに       | 0.9868 | 魁         | 0.0096 |
| 5  | ヒアリング     | 0.9856 | 春闘        | 0.0098 |
| 6  | 衣         | 0.9848 | 良い        | 0.0098 |
| 7  | 正当        | 0.9828 | 原発        | 0.0099 |
| 8  | 遺影        | 0.9802 | 相撲        | 0.0105 |
| 9  | しっかり      | 0.9801 | おいしい      | 0.0105 |
| 10 | 垣間見       | 0.9775 | 粗悪        | 0.0113 |
| 11 | 起訴        | 0.9771 | 元         | 0.0113 |
| 12 | 気持        | 0.9736 | できれ       | 0.0117 |
| 13 | 奏         | 0.9728 | ありえ       | 0.0117 |
| 14 | 高熱        | 0.9719 | 先行        | 0.0118 |
| 15 | 対外        | 0.9692 | 毎日新聞      | 0.0119 |
| 16 | 寝         | 0.9680 | あり        | 0.0121 |
| 17 | 段取り       | 0.9678 | なし        | 0.0122 |
| 18 | 法廷        | 0.9667 | 持ち込み      | 0.0122 |
| 19 | 生家        | 0.9661 | 東電        | 0.0122 |
| 20 | 聞か        | 0.9657 | 差し込む      | 0.0123 |
| 21 | 顕         | 0.9650 | よい        | 0.0127 |
| 22 | 起こし       | 0.9647 | 支持        | 0.0128 |
| 23 | 傷み        | 0.9647 | あるく       | 0.0129 |
| 24 | 擁立        | 0.9625 | 疲れ        | 0.0129 |
| 25 | 日取り       | 0.9605 | 微増        | 0.0130 |
| 26 | 式典        | 0.9601 | 反対        | 0.0131 |
| 27 | 坂口        | 0.9590 | 民主党       | 0.0132 |
| 28 | 庁         | 0.9581 | 下位        | 0.0133 |
| 29 | それ        | 0.9572 | 低い        | 0.0136 |
| 30 | 摂氏        | 0.9551 | ありがた      | 0.0141 |

表 4.23: BERT の素性 (数値評価 2)

| 順位 | 賛成を得やすい素性 | 数値     | 賛成を得にくい素性   | 数値     |
|----|-----------|--------|-------------|--------|
| 1  | 聞き出す      | 0.9950 | 笑え          | 0.0040 |
| 2  | 起きる       | 0.9944 | あげよ         | 0.0041 |
| 3  | 沙汰        | 0.9944 | もったいない      | 0.0043 |
| 4  | 儀式        | 0.9942 | いくらでも       | 0.0044 |
| 5  | こんな       | 0.9941 | すりゃ         | 0.0048 |
| 6  | 犯す        | 0.9941 | 何でもかんでも     | 0.0048 |
| 7  | 件         | 0.9941 | びしていてすごいなあと | 0.0049 |
| 8  | 発生        | 0.9940 | 憎く          | 0.0050 |
| 9  | 未代        | 0.9939 | 中傷          | 0.0054 |
| 10 | 引越し       | 0.9938 | 頑張れ         | 0.0054 |
| 11 | 政調        | 0.9938 | つまらない       | 0.0055 |
| 12 | 終わっ       | 0.9938 | めちゃくちゃ      | 0.0055 |
| 13 | 伴う        | 0.9937 | 煽ら          | 0.0057 |
| 14 | 罪         | 0.9937 | 愚か者         | 0.0057 |
| 15 | いじっ       | 0.9936 | でないと        | 0.0064 |
| 16 | 統合        | 0.9936 | さされ         | 0.0065 |
| 17 | 変わっ       | 0.9936 | 良かれ         | 0.0067 |
| 18 | 立てる       | 0.9936 | うまく         | 0.0072 |
| 19 | こういう      | 0.9936 | だせ          | 0.0074 |
| 20 | 願っ        | 0.9935 | おめでとう       | 0.0076 |
| 21 | 下さっ       | 0.9935 | すぎる         | 0.0079 |
| 22 | 末         | 0.9935 | のしかかる       | 0.0084 |
| 23 | 仕分け       | 0.9935 | だらし         | 0.0085 |
| 24 | 詰め寄っ      | 0.9935 | ビックリ        | 0.0089 |
| 25 | 時         | 0.9935 | 持てる         | 0.0090 |
| 26 | 事態        | 0.9934 | よろしく        | 0.0090 |
| 27 | 断交        | 0.9934 | 愚か          | 0.0092 |
| 28 | 交代        | 0.9934 | やれやれ        | 0.0093 |
| 29 | 某         | 0.9934 | 早く          | 0.0093 |
| 30 | 事件        | 0.9934 | 怖く          | 0.0094 |

表 4.24: BERT の素性 (数値評価 3)

| 順位 | 賛成を得やすい素性 | 数値     | 賛成を得にくい素性 | 数値     |
|----|-----------|--------|-----------|--------|
| 1  | 本当に       | 0.9964 | 貢い        | 0.0055 |
| 2  | かけ離れ      | 0.9947 | 気苦労       | 0.0059 |
| 3  | こんなにも     | 0.9947 | ヤフコメ      | 0.0060 |
| 4  | 振り回さ      | 0.9941 | ざあああああい   | 0.0063 |
| 5  | 欲張ら       | 0.9937 | 売国奴       | 0.0063 |
| 6  | 凶器        | 0.9936 | 食いっぽぐれ    | 0.0063 |
| 7  | 手のひら      | 0.9934 | イタクラ      | 0.0064 |
| 8  | 左巻き       | 0.9934 | ヤフー       | 0.0065 |
| 9  | P T S D   | 0.9932 | ノルマ       | 0.0068 |
| 10 | 宮内庁       | 0.9931 | 独眼竜       | 0.0069 |
| 11 | 受診        | 0.9926 | 飲み食い      | 0.0069 |
| 12 | 自衛隊       | 0.9926 | アメフト      | 0.0069 |
| 13 | 振込        | 0.9926 | 取り締め      | 0.0072 |
| 14 | 凶行        | 0.9925 | アベノコメント   | 0.0072 |
| 15 | 未必の故意     | 0.9924 | トラウト      | 0.0073 |
| 16 | 歳児        | 0.9922 | これ程       | 0.0074 |
| 17 | とんでも      | 0.9921 | ばらまき      | 0.0074 |
| 18 | 弁護        | 0.9920 | 神々しかっ     | 0.0074 |
| 19 | 若返ら       | 0.9919 | 世界中       | 0.0075 |
| 20 | 正装        | 0.9919 | 愚策        | 0.0075 |
| 21 | 外務省       | 0.9918 | 呪わ        | 0.0075 |
| 22 | 抜擢        | 0.9916 | 随分        | 0.0076 |
| 23 | 支給        | 0.9915 | まずい       | 0.0076 |
| 24 | 顔向け       | 0.9914 | 他山の石      | 0.0076 |
| 25 | 絞ら        | 0.9913 | 売国        | 0.0076 |
| 26 | 語る        | 0.9912 | 悪けれ       | 0.0077 |
| 27 | 寄り添う      | 0.9910 | モリカケ      | 0.0077 |
| 28 | 拘束        | 0.9910 | ワガママ      | 0.0077 |
| 29 | 本省        | 0.9910 | 騙せ        | 0.0078 |
| 30 | 打ち出さ      | 0.9908 | 悪かっ       | 0.0078 |

## 4.3 素性分析

本節では記事分野別と数値評価で行った素性分析の結果についての説明をそれぞれ記述する。

### 4.3.1 記事分野別の素性分析

表 4.2 にあるように政治分野では「賢明」が賛成を得やすいとされている。これは「賢明」が知的な単語であるため、文章全体がまとまっており、理路整然としていて読みやすいのが賛成を得やすくしているのではないかと予想できる。表 4.3 にあるように事件分野では「胸糞」というマイナスな単語が賛成を得やすいとされている。これは事件に対して胸糞悪いなどのコメントをすることで同意の感情から賛成を得ているのではないかと予想できる。表 4.5 にあるようにスポーツ分野では「連戦」が賛成を得にくいと判断されている。これは怪我などのマイナスな出来事が予想される単語が賛成を得にくいと判断されていた。表 4.4 にあるように病気分野では「旅行」が賛成を得にくい素性だと判断されていた。これはコロナウイルスの感染拡大が懸念されていた中で旅行をして感染した人が批判されていると予想できる。また、表 4.3, 表 4.4 には「本当に」、「絶対」などが賛成を得やすいと判断されている。これらの単語は文を強調する単語であるためコメントに自信があり、納得できることが多いのではないかと考えられる。

結果として、三木らの先行研究 [1] [2] と同様に多くの場合で分類された分野で使用されるマイナスイメージを持つ単語が賛成を得にくい素性だと判断された。また、表 4.6, 表 4.11 では「男性」、「奥さん」が賛成を得やすい素性として得られている。しかし同じ表 4.6, 表 4.11 の賛成を得にくい素性としては「女」が得られている。これは「性」をつけていないことや「女」など偉そうな単語が入っていることによって文章全体も攻撃的な文章や差別的な内容を含む文章になっている可能性が高いと推測することができる。

### 4.3.2 数値評価の素性分析

数値評価 1-1, 1-2 の素性分析の結果得られた素性は表 4.13, 表 4.14, 表 4.17, 表 4.18 にあるように「速報」、「録音」などが賛成を得やすい素性とされており、「外人」、「おじさん」などが賛成を得にくい素性とされた。これは速い情報ほど価値があるため、「速

報」を使うコメントは賛成を得やすいのではないかと考えられる。また、「外人」などは4.3.1節の「男」、「女」のように「国」が入っていないため差別用語とみなされ、賛成を得にくくなっているのではないかと考える。この数値評価1-1、数値評価1-2の素性分析は三木らの先行研究 [1] [2] と同じ素性が多く発見された。

数値評価2の素性分析の結果得られた素性は表4.15、表4.19にあるように「誕生」などが賛成を得やすい素性とされており、「中傷」などが賛成を得にくい素性とされた。「誕生」などは何かが生まれたことを意味するポジティブな意味を持つ単語で数値評価2の高評価の比率を求める式では低評価が少なければ高評価のコメントになりやすいのが影響しているのではないかと考える。反対に「中傷」はネガティブな意味を持つ単語であるため低評価が増えやすく賛成を得にくいコメントとなりやすいのが影響しているのではないかと考えることが可能である。

数値評価3の素性分析の結果得られた素性は表4.16、表4.20にあるように「おめでとう」などが賛成を得やすいとされ、「殺害」などが賛成を得にくい素性とされた。数値評価3の単位時間あたりの高評価を求める式では短い時間で多くの高評価を得ると賛成を得やすい文章となるため、「おめでとう」や「殺害」などの直感的に良いか悪いか判断できる単語が入っていると賛成を得やすい、得にくくなるのではないかと考える。

結果として、数値評価の種類別によって「誕生」などが低評価を得にくい素性であるとわかり、「おめでとう」などが短時間である程度の賛成を得ることができるなどが判明した。

### 4.3.3 今後の課題

表4.8では賛成を得やすい素性として「胸糞」が得られている。しかし、この単語が文章に入っていると賛成を得やすいとは言いにくい。これは事件の分野で賛成を得やすい素性なのでマイナスな事件に対しての批判的なコメントに同情している人が多いと考えられる。このことから記事自体の内容をポジティブかネガティブかを分類することでポジティブな記事に対して賛成を得やすい素性、ネガティブな記事に対して賛成を得やすい素性を求めることが可能になると考えられる。また、機械学習の性能が数値評価1-1、数値評価1-2を比較すると向上しているので高評価のコメントと判断する範囲を厳しくすることで性能向上が可能になると考えられる。よって他の提案手法などでも高評価の範囲を厳しくすることで性能が向上し、良い素性が得られる可能

性が高くなることが予想される。

## 第5章 考察

三木らの先行研究 [1] [2] の結果として、賛成を得やすい単語は得られなかった。この理由として、記事の分野が混ざっていることと低評価を用いていないことが挙げられた。そこで記事分野別の実験と低評価を考慮した数値評価を用いた実験を行った。

表 4.2～表 4.11 にあるように分野別に分類して機械学習で実験し、素性分析をするとその分野でよく使用されるであろう単語が賛成を得やすい、賛成を得にくい素性の両方で得られた。特にマイナスなイメージを持つ単語は三木らの先行研究 [1] [2] と同様に賛成を得にくい素性だと判断された。まれにマイナスなイメージを持つ単語でも賛成を得やすい素性と判断されている場合が存在したがこれは記事自体の内容がマイナスな意味を持ち、それを批判する内容のコメントを書くことによってマイナスなイメージの単語でも多くの賛成を得ることができ、賛成を得やすい素性と判断されたのではないかと考える。

数値評価を用いて実験を行った結果、数値評価別の素性を発見できた。三木らの先行研究 [1] [2] と数値評価 1 の違いとしては低評価を先行研究の式に組み込んだ。数値評価 2 の違いとしては高評価と低評価の比率を利用した式を利用した。数値評価 3 の違いとしては高評価をコメント時刻で割ることで単位時間あたりの高評価数を求める式を利用した。

数値評価 1 と先行研究の正解率を比較すると数値評価 1-1 は先行研究を下回り、数値評価 1-2 は先行研究を上回った。これは高評価の定義を高くしたことによってコメントの内容が向上したことが予想される。数値評価 1 の式は高評価、低評価、コメント時刻の 3 種類を利用しているのでこの実験で得られた素性は高評価数が多く、低評価が少なく、コメント時刻が最近という条件を満たした素性が多いと考えられる。よって少ない時間で高評価が多く、低評価が少ないというバランスを満たすコメント作成に役立つと考えられる。数値評価 2 の式では高評価のコメントとなるには低評価を少なくすることが重要となる。つまり、数値評価 2 の賛成を得やすい素性は低評価を得にくい単語と考えることができる。よって、批判を受けにくいコメントを作成するためには数値評価 2 の賛成を得にくい素性を使用すれば良いのではないかと考える。同

様に、数値評価3では単位時間あたりの高評価数が重要なので、少ない時間である程度の高評価を得たい場合には役に立つと考える。しかし、数値評価3の賛成を得やすい素性は別の数値評価では出てこない素性も多くあるため、短時間で多くの高評価を得るために役立つ素性が多いと考える。

## 第6章 おわりに

本研究は Yahoo!ニュースのコメント欄と機械学習を用いて、賛成を得やすい文章の種々の分析を目的としている。Yahoo!ニュース内にある 1,000 コメント以上持つ 310 記事を分野別に分類し、コメント欄の賛成した人数、時刻の情報を使用し、賛成が得やすい文章を機械学習を用いて推定させる。加えて、高評価、低評価、コメント時刻の情報から数値評価を記事に投稿されたコメントを高評価コメント、中評価コメント、低評価コメントの 3 種類に分類する。その中から高評価コメントと低評価コメントの 2 種類を機械学習に入力して高評価か低評価を推定させる。

推定させた結果、記事分野別の正解率は約 7 割前後となった。数値評価の正解率は数値評価 1-2 のみが約 8 割前後となり、それ以外は約 6 割前後となった。また、記事分野別の多くの場合で BERT, ME, SVM の順で正解率が高くなった。しかし数値評価では ME と BERT が同程度の正解率だったが数値評価 1-1,1-2 では SVM だけが正解率が低くなっていた。

素性分析の結果では、分野別ではその分野で使用されることが多い単語でポジティブな意味を持つ素性は賛成を得やすく、ネガティブな意味を持つ素性は賛成を得にくいと判断されやすいとわかった。政治分野では「賢明」などが得られた。これは「賢明」が知的な単語であるためコメントも知的で納得しやすいのではないかと考えられる。また、「本当に」、「絶対」なども賛成を得やすいと判断された。これは強調を意味する単語であるため自信を持ったコメントに使用されることが多く賛成を得ているのではないかと考える。

数値評価 1 では「失敬」、「速報」などが賛成を得やすいと判断されていた。「失敬」は「賢明」と同様の理由で知的な単語であるためコメントも知的で賛成を得ているのではないかと考えられる。「速報」は速い情報に価値があるため賛成を得やすいと判断されていると考えられる。数値評価 2 では「誕生」などが批判を受けにくいとわかり、数値評価 3 では「おめでとう」や「殺害」などのポジティブかネガティブかわかりやすい素性が短時間ではそのまま賛成を得やすい、得にくい素性となることがわかった。

今後は記事自体をポジティブ、ネガティブに分類して実験を行うことでより細分化

して分析を行うことが可能になると考える。

# 謝辞

また、研究の進め方や本論文の書き方など、細部にわたる御指導を頂きました、鳥取大学工学部知能情報工学科自然言語処理研究室の村田真樹教授に心から御礼申し上げます。また、本研究を進めるにあたり、御指導、御助言を頂きました、村上仁一准教授に心から御礼申し上げます。その他様々な場面で御助言を頂いた自然言語処理研究室の皆様に感謝の意を表します。

## 参考文献

- [1] 三木謙志, 村田真樹, 馬青. 賛成を得やすい文章の機械学習を利用した収集と分析. 言語処理学会第 28 回年次大会発表論文集, pp. 1541–1545, 3 2022.
- [2] 三木謙志. 機械学習を利用した賛成を得やすい文章の分析. 鳥取大学工学部電気情報系学科卒業論文, 2021.
- [3] 高橋直己, 上野舞夕, 浜田百合, 庄司裕子. 絵文字を用いた文章における感情伝達効果に関する研究. 日本感性工学会論文誌, Vol. 21, No. 1, pp. 135–142, 2022.
- [4] 松本和幸, 三品賢一, 任福継, 黒岩眞吾. 感情生起事象文型パターンに基づいた会話文からの感情推定手法. 自然言語処理, Vol. 14, No. 3, pp. 239–271, 2007.
- [5] 水野淳太, 渡邊陽太郎, 村上浩司, 乾健太郎, 松本裕治ほか. 文間関係認識に基づく賛成・反対意見の俯瞰. 情報処理学会論文誌, Vol. 52, No. 12, pp. 3408–3422, 2011.
- [6] 西條結人, 田中大輝, 小野由美子. 意見文課題における説得のアピールの日西対照研究-日本とスペインの学生の作文比較. 教育実践学論集, Vol. 16, pp. 95–107, 2015.
- [7] 端大輝, 村田真樹, 徳久雅人. 感動を与える文の自動取得と分析. 言語処理学会第 18 回年次大会, 2012.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.