

# 候補者へのリプライテキストの感情度を考慮した 国内選挙予測

細川 祐生<sup>1</sup> 沖本 吉生<sup>2</sup> 張 建偉<sup>1</sup>

**概要:** 現在、ウェブ上の情報を用いて選挙結果の予測をする研究が行われている。本研究では、既存手法に対して候補者の Twitter アカウントへのリプライテキストの分析を元にリプライ数と感情度を特徴量として加え、その有用性の評価とより高い汎化性能をもつモデルの構築を試みた。2019 年参議院議員選挙を対象に行った評価実験の結果として、F 値は既存手法と比べて高い数値となり、リプライテキストと感情分析を用いた手法を利用することは選挙予測において有用であることが示された。

## Japanese Election Prediction Considering Sentiment Values of Reply Texts to Candidates

**Abstract:** Studies that predict election results using information on the Internet is being conducted currently. In this paper, we added the number of replies and sentiment values as features to the previous method, based on the analysis of reply texts to Twitter accounts of candidates, and we built models that have higher generalization performance and evaluated those usability. As a result of evaluation experiment conducted for Japanese general election in 2013, F-measure was shown to be higher than the previous method, consequently, it indicated that using our method consisting of reply texts and sentiment analysis is useful for election prediction.

### 1. はじめに

ソーシャルメディアには世界中のユーザ達が自由に発信した大量の情報で溢れている。その情報を元に実世界の観測や予測を行うタスクが多く存在しており [1]、その一つに選挙予測がある。選挙予測とは、ウェブ上の情報を元に選挙結果を予測するものである。このタスクを扱った研究は各国の選挙を対象に行われている [2], [3], [4], [5]。日本国内の選挙では 2013 年参議院議員選挙から選挙運動へのインターネット利用が開始され、当時から情報拡散のために全体の半数以上の候補者たちがソーシャルメディアの一つである Twitter を盛んに利用している [6]。また、実際に Twitter の情報を元に候補者の当選・落選の結果を予測することがある程度可能であることが分かっている [7]。ここで主な関連研究の 2 つを紹介する。

Tungawan 氏らの研究 [4] は 2016 年のアメリカ大統領選挙を対象に候補者の得票数順位の予測を試みた。データ

には Twitter 上の #Election2016 が含まれる英文ツイートを利用し、そのテキストデータの感情に着目している。結果として予測精度は 54.8% であり良い結果ではない。しかし、ツイートの感情に着目することの妥当性を否定する要素はなく、特徴量としての有用性を評価する価値は十分にあると考える。

2013 年参議院議員選挙を対象とした那須野氏らの研究 [5] では、Twitter を利用する候補者アカウントから得られる 2 つの指標（知名度指標と選挙地盤指標）を元に当選者の予測を試みた。結果として選挙予測に対して有用な特徴量が示され、拡張可能性として有権者に焦点を当てた指標を取り入れることへの期待が記されていた。

本研究では、有権者に焦点を当てた指標として Twitter 上の候補者へのリプライテキスト（以降、リプライと記述）に着目する。国内選挙において選挙期間中に Twitter を利用する候補者たちは自身の選挙運動に関する告知や報告などに関するツイートを投稿する。そのツイートに対して複数の Twitter ユーザたちから候補者への応援や意見などが記されたリプライが送信される。ここで候補者のツイート

<sup>1</sup> 岩手大学 理工学部

<sup>2</sup> 株式会社コスモ・コンピューティングシステム

と送信されたリプライの内容を比較すると送信されたリプライの方がポジティブ・ネガティブのような感情極性が現れやすいものとなっていた。また、リプライの多くは日本語で記述されている。日本語は英語よりも一文字あたりの情報量が多いとされており [8]、日本語のテキストから抽出した特徴量は予測タスクにおいて良い結果をもたらすと予想される。

以上の着目点から各候補者へのリプライの件数と感情度を新たな特徴量として既存手法 [5] の 9 次元特徴量から構成されるモデルに追加することで、モデルの汎化性能の向上を目指し、その新規特徴量の選挙予測に対する有用性について検証することを本研究の目的とする。本研究では、2019 年参議院議員選挙を対象に、出馬した 370 名の候補者の内 Twitter を選挙運動に利用していた 262 名のアカウントから得られる情報を用いて当選者の予測を行う。

本研究のように選挙予測の精度向上を目指すことで候補者たちの視点と有権者たちの視点の双方向で以下のような利用価値が考えられる。

- 候補者たちと政党は当選に寄与するユーザの特徴や行動を知ることができ、選挙活動と自党議席の獲得を効率的に進めることができる。
- 有権者たちは、Twitter から候補者の支持率が明らかになることで、簡単に投票する候補者を選んだり選挙の動向を知ることができる。

本稿の構成は次の通りである。まず 2 章で本研究の提案手法について説明する。次に 3 章で利用する対象選挙に関するデータと収集したデータについて述べる。次に 4 章で評価実験を通して新規特徴量の有用性について検証する。最後に 5 章で本研究における課題に関する議論を述べ、6 章でまとめとなる。

## 2. 提案手法

### 2.1 ベースラインモデル

本研究では、既存手法 [5] で提示された 2 つの指標を用いてベースラインモデルを構築する。これらの指標は候補者の知名度に関する知名度指標 6 つ、候補者への後援や支持の強固さに関する選挙地盤指標\*1 3 つの特徴量で構成される。2 つの指標の各特徴量を以下に示す。

知名度指標

フォロワ数、フレンド数、被登録リスト数、アカウント存在日数、承認バッジの有無、選挙期間中のツイート数

選挙地盤指標

リーチ: 候補者の 1 ツイートが拡散されるユーザ数の期待値  
バラエティ: 選挙地盤を構成する後援者の多様さ  
ロイヤルティ: 選挙地盤を構成する後援者の忠誠度

\*1 計算手法は那須野氏らの論文 [5] 内の定義に基づいている。

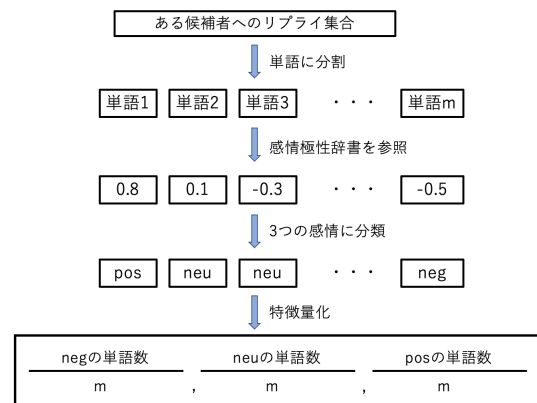


図 1 単語ベース手法

Fig. 1 Word-based method

### 2.2 感情度計算と特徴量抽出

本研究における提案手法では、候補者ごとにリプライの感情度を計算し、その感情度を閾値によってネガティブ (neg)・ニュートラル (neu)・ポジティブ (pos) の 3 つに分類した後、それらの比率を特徴量とする。

#### 2.2.1 リプライの感情度計算

以下 2 つの感情度計算手法では、岩波国語辞典に掲載された 55,125 語に感情度を対応付けた高村氏の感情極性辞書 [9] を利用している。辞書内の感情度と本手法で計算される感情度は全て -1 から +1 までの連続値となる。

##### (1) 単語ベースでの手法

この手法の流れは、まずある候補者へのリプライを集集合として、そこに出現した各単語に対して辞書の値を参照する。そして参照した値を感情度として各単語を 3 つの感情に分類する。単語ベース手法の概略図を図 1 に示す。

##### (2) ツイートベースでの手法

この手法の流れは、ある候補者へのリプライを集集合として、各リプライの感情度を計算する。リプライは一つ以上の文から成り立ち、その感情度計算には三和氏らの手法 [10] を利用する。この手法は文内の単語位置を考慮して文の感情度を計算する手法で、文の最後に出現する単語の感情度の方が比較的重要であるという考えに基づいている。そして計算した感情度を元に各リプライを 3 つの感情に分類する。ツイートベース手法の概略図を図 2 に示す。

#### 2.2.2 感情分類と閾値

計算された感情度がネガティブ・ニュートラル・ポジティブのどれに属するか閾値を設けて判定する。分類のルールは閾値を  $\tau$ 、感情度を  $S$  として表 1 に示す。(1) の手法においては  $\tau = 0.2, 0.3, 0.4, 0.5, 0.6$  の全 5 通り。(2) の手法においては計算中の単語の分類に用いる  $\tau_1$  と計算後のツイートの分類に用いる  $\tau_2$  で 2 つの閾値を設定する。したがって、 $\tau_1, \tau_2$  には 0.2, 0.3, 0.4, 0.5, 0.6 の値をそれぞれ設定

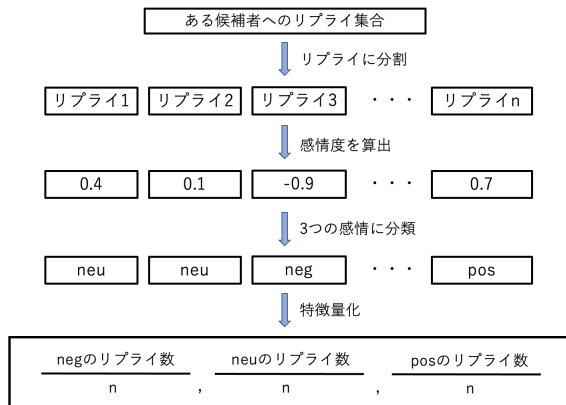


図 2 ツイートベース手法  
Fig. 2 Tweet-based method

表 1 感情度分類ルール

Table 1 Rule of sentiment values classification

ネガティブ	ニュートラル	ポジティブ
$-1 \leq S < -\tau$	$-\tau \leq S \leq +\tau$	$+\tau < S \leq +1$

表 2 当該選挙の概要データ

Table 2 Summary data of the election

総候補者数	当選者数	落選者数
370 人	245 人	125 人

Twitter 利用者数	当選者数	落選者数
262 人	100 人	162 人

し、それらの直積をとった全 25 通りとなる。

### 3. データ

#### 3.1 対象選挙の概要

本研究は 2019 年参議院議員選挙を対象とする。当該選挙の概要データを表 2 に示す。

データによると候補者の Twitter 利用率が 70.8% で、Twitter 利用者内の当選者の割合は 38.2% であった。国内において選挙運動のインターネット利用が開始された 2013 年参議院議員選挙ではそれぞれ 66.3%、26.8% であり、本研究の対象選挙では選挙運動への Twitter 利用の普及が進んだことが分かった。したがって、Twitter 上のデータは選挙予測に対して利用する価値が向上したと考えられる。

#### 3.2 データ収集

データ収集には Twitter REST API を利用した。本研究のデータ収集における課題は知名度指標と選挙地盤指標を選挙当時のデータに可能な限り近づけなければならないことである。そこで、選挙期間を 2019 年 7 月 4 日 00:00 (JST) から 2019 年 7 月 21 日 20:00 (JST) まで (選挙公示日から投票終了時間まで) と定めた。API で収集可能なアカウント・リスト・ツイートの作成時間を利用し、投票

表 3 テキストクレンジングの処理例

Table 3 Example of text cleansing processing

	@candidate1 応援しています Fight!(^^)
処理前	https://example.jp/abc123/ #政党名 #候補者名 #参院選 2019
処理後	応援しています

終了時点のデータに近づける処理を行った\*2。本研究の収集対象となるリプライの定義は、定めた選挙期間中に投稿され、候補者アカウントをメンションしたツイートとした。収集した結果、132,922 件のリプライが得られた。次節で述べる前処理を適用し文字数が 0 となったテキストを除去するとリプライは 128,443 件が残った。

#### 3.3 リプライの前処理

リプライの前処理はテキストデータ内の不要部分を除去するクレンジングと品詞ごとに単語の分割を行う形態素解析である。本研究では、平仮名・片仮名・漢字の日本語の表現のみを分析対象とするので以下のルールに従ってクレンジングを行った。クレンジングの処理例を表 3 に示す。

- 平仮名・片仮名・漢字以外の文字は例外\*3を除いて全て削除する。
- URL・ハッシュタグ (#)・メンション (@)・引用 RT 文は削除する。
- 絵文字・顔文字は削除する。

クレンジング処理後のテキストに対し形態素解析器 JUMAN++[11] を用いて形態素解析を行う。提案手法適用時には単語の原形を用いて、代表表記が登録、されている単語については代表表記を単語の原形として利用する。例外としてインスタグラムのように代表表記が英字となる単語については片仮名に統一して利用する。

### 4. 実験

#### 4.1 実験の設定

本実験では、2019 年参議院議員選挙を対象に当選を 1、落選を 0 とした 2 クラス分類を行う。教師あり学習手法には特徴量の重要度を定量的に測ることができるランダムフォレストを用いる。そのパラメータである決定木の数はグリッドサーチによってチューニングを行う。評価指標は精度 (accuracy)、適合率 (precision)、再現率 (recall)、F 値の 4 つを利用する。モデルの汎化性能の評価には主に F 値に注目する。また、精度の評価には 10 分割交差検証を用いる。本実験では 3 つの実験を行う。各実験で構築する 6 つのモデルを以下に示す。まず構築する全モデルに対するベースラインとして既存手法 [5] の 9 次元特徴量を用い

\*2 承認バッジの有無に関しては付与された日時が収集できないため 2020 年 9 月 17 日時点のデータとなっている。

\*3 - (伸ばし棒) は除去すると文意が変わるため除去しない。また、笑の意味で使われる w は全て漢字の笑に変換した。

たモデル(1)を構築する。実験1では知名度指標の一つとして考えられるリプライの件数を特徴量としてモデル(1)に加えモデル(2)を構築する。実験2では2つの提案手法で抽出した単語ベース手法、ツイートベース手法それぞれの3次元特徴量をモデル(1)に加えモデル(3),(4)を構築する。実験3ではモデル(3),(4)からニュートラルの単語またはツイートの比率を除いた2次元特徴量をモデル(1)に加えモデル(5),(6)を構築する。実験3を実施する理由は、実験2で追加する特徴量は合計すると1になるような値であり、閾値の変更がその性質によって補完される可能性を少しでも減らした結果を検証するためである。また、モデル(2)とモデル(3)~(6)の内でF値が最大値を示したモデルでは追加した特徴量の重要度を順位で示し、どれだけ有用な特徴量であったのかを他の特徴量と比較する。

実験1, 2, 3 共通

(1) 那須野氏らの手法 [5] を用いたベースラインモデル  
実験1

(2) (1)にリプライの件数の1次元特徴量を追加したモデル  
実験2

(3) (1)に単語ベース手法の3次元特徴量を追加したモデル  
実験3

(4) (1)にツイートベース手法の3次元特徴量を追加したモデル  
実験3

(5) (1)に単語ベース手法の2次元特徴量を追加したモデル  
実験3

(6) (1)にツイートベース手法の2次元特徴量を追加したモデル  
実験3

## 4.2 実験結果

実験結果として、まず各モデルの評価指標をベースラインと比較する。モデル(3)~(6)では閾値ごとに構築した各モデルのF値の比較を最大値,最小値,中央値,平均,標準偏差の5つの統計量によって示す。

実験1

実験1ではリプライの件数を1次元特徴量としてモデル(1)に追加し評価した(表4)。結果によるとモデル(2)の精度とF値はベースラインを上回り、汎化性能の向上が見られた。よってリプライの件数の特徴量としての有用性が示された。特徴量の重要度順位を表5に示した。リプライの件数は10位中6位でやや中間に位置しており、既存手法と比べてもある程度の有用性が保証されていると考えられる。

実験2

実験2では2つの提案手法によって抽出した3次元特徴量をモデル(1)に追加し評価した。4つの評価指標を表6

表4 実験1の評価実験結果

Table 4 Result of experiment 1

モデル \ 評価指標	精度	適合率	再現率	F 値
(1) ベースライン	0.756	0.705	0.620	0.660
(2) リプライ数	0.767	0.724	0.630	0.685

表5 実験1の特徴量の重要度順位

Table 5 Rank of features importance in experiment 1

順位	モデル(2)特徴量
1位	フレンド数
2位	承認バッジの有無
3位	被登録リスト数
4位	パラエティ
5位	フォロー数
6位	リプライの件数
7位	リーチ
8位	アカウントの存在日数
9位	ロイヤルティ
10位	選挙期間中のツイート数

表6 実験2の評価実験結果

Table 6 Result of experiment 2

モデル \ 評価指標	精度	適合率	再現率	F 値
(1) ベースライン	0.756	0.705	0.620	0.660
(3) 単語ベース	0.770	0.729	0.632	0.677
(4) ツイートベース	0.763	0.725	0.611	0.663

に示し、モデル(3),(4)の閾値ごとのF値を5つの統計量を用いて表7に示す。尚、表6の(3)と(4)の評価値には閾値ごとの平均を示した。

まず(3),(4)では精度とF値ともにベースラインを上回り、汎化性能の向上が見られた。しかし、(3)と(4)を比較すると(4)の方が精度とF値はやや低い結果となった。これよりツイートベース手法の感情分類精度が単語ベース手法よりも劣っている可能性が考えられる。次に(3)と(4)のF値の統計量を見る。標準偏差は小さい値となりデータのばらつきは少ないが、最大値はベースラインを上回り、最小値はベースラインを下回っているという結果から提案手法に対して閾値設定が重要であると考えられる。中央値はベースラインを上回っているため、閾値ごとのモデルの50%以上がベースラインより高い汎化性能を示したといえる。表8に示した特徴量順位では単語ベース手法ではある程度の有用性は保証されているように考えられるが、ツイートベース手法では比較的大きな有用性は保証されていないように見える。

実験3

実験3では2つの提案手法によって抽出した特徴量からニュートラルの単語またはツイートの比率を除いた2次元特徴量をモデル(1)に追加し評価した。実験2と同様に4つの評価指標を表9に、モデル(5),(6)の閾値ごとのF

表 7 実験 2 の F 値比較 (閾値は  $\tau_1, \tau_2$  の順に記載)

Table 7 Comparison of F-measure in experiment 2  
(Thresholds are ordered by  $\tau_1$  and  $\tau_2$ )

モデル \ 指標	統計量	閾値	F 値
(1) ベースライン	-	-	0.660
	最大値	0.2	0.689
	最小値	0.4	0.656
(3) 単語ベース	中央値	-	0.681
	平均	-	0.677
	標準偏差	-	0.0115
	最大値	0.4,0.4	0.685
(4) ツイートベース	最小値	0.4,0.3	0.634
	中央値	-	0.667
	平均	-	0.663
	標準偏差	-	0.0144

表 8 実験 2 の特徴量の重要度順位

Table 8 Rank of features importance in experiment 2

順位	モデル (3) 特徴量	モデル (4) 特徴量
1 位	フレンド数	フレンド数
2 位	被登録リスト数	被登録リスト数
3 位	承認バッジの有無	承認バッジの有無
4 位	フォロー数	フォロー数
5 位	バラエティ	バラエティ
6 位	ポジティブの比率	アカウント存在日数
7 位	リーチ	リーチ
8 位	アカウント存在日数	ニュートラルの比率
9 位	ネガティブの比率	ネガティブの比率
10 位	ニュートラルの比率	ロイヤルティ
11 位	ロイヤルティ	ポジティブの比率
12 位	選挙期間中のツイート数	選挙期間中のツイート数

表 9 実験 3 の評価実験結果

Table 9 Result of experiment 3

モデル \ 評価指標	精度	適合率	再現率	F 値
(1) ベースライン	0.756	0.705	0.620	0.660
(5) 単語ベース	0.781	0.743	0.652	0.694
(6) ツイートベース	0.769	0.734	0.621	0.673

値の 5 つの統計量を表 10 にそれぞれ示す。

まず (5), (6) では精度と F 値ともにベースラインを上回り, 実験 2 と同様に汎化性能の向上が見られた。ツイートベース手法を用いた (6) は単語ベース手法を用いた (5) より低い値を示した。次に (5) と (6) の F 値の統計量を見る。標準偏差を見るとデータのばらつきは少ないが, (6) では最大値がベースラインを上回り, 最小値がベースラインを下回っていることから閾値設定の重要性が実験 2 と同様に見られた。表 11 で示した特徴量の重要度順位ではツイートベース手法よりも単語ベース手法の方が特徴量の有用性が保証されているように考えられる。

最後に実験 2 と実験 3 の結果を比較する。ここでの比較は主に (3) と (5), (4) と (6) の比較である。まず精度と

表 10 実験 3 の F 値比較 (閾値は  $\tau_1, \tau_2$  の順に記載)

Table 10 Comparison of F-measure in experiment 3  
(Thresholds are ordered by  $\tau_1$  and  $\tau_2$ )

モデル \ 指標	統計量	閾値	F 値
(1) ベースライン	-	-	0.660
	最大値	0.5	0.706
	最小値	0.4	0.681
(5) 単語ベース	中央値	-	0.697
	平均	-	0.694
	標準偏差	-	0.0112
	最大値	0.4,0.4	0.695
(6) ツイートベース	最小値	0.3,0.5	0.644
	中央値	-	0.674
	平均	-	0.673
	標準偏差	-	0.0142

表 11 実験 3 の特徴量の重要度順位

Table 11 Rank of features importance in experiment 3

順位	モデル (5) 特徴量	モデル (6) 特徴量
1 位	フレンド数	フレンド数
2 位	被登録リスト数	被登録リスト数
3 位	承認バッジの有無	承認バッジの有無
4 位	フォロー数	フォロー数
5 位	バラエティ	バラエティ
6 位	ネガティブの比率	アカウント存在日数
7 位	リーチ	リーチ
8 位	アカウント存在日数	ネガティブの比率
9 位	ポジティブの比率	ロイヤルティ
10 位	ロイヤルティ	ポジティブの比率
11 位	選挙期間中のツイート数	選挙期間中のツイート数

F 値は実験 3 の方が高い結果となった。特に (5) は (3) と比べて F 値の最大値が大きく向上した。ニュートラルの比率を除去することによって候補者ごとに生じるリプライ集合内の単語数の差に起因する特徴量としての欠点が変わり, リプライの感情をより適切に表現できたと考えられる。

## 5. 議論

本研究の課題としてツイートベース手法の感情分析精度が単語ベース手法より劣っている可能性があるということが挙げられる。ツイートベース手法では単語の位置を考慮した手法であるが, その他の文構造を考慮して感情度を測る必要があると考えられる。そこで挙げられる改善策は以下の 2 つであり本研究における今後の課題とする。

- 単語同士の係り受けを考慮する。
- 顔文字と絵文字を考慮する。

まず単語同士の係り受けとは, 否定語や強調副詞などのように他の単語と文内で意味的に繋がっていることを指す。係り受けを見ることで単語間の感情度への影響を考慮でき, 例えばポジティブな単語を打ち消してネガティブな文意にしているものを判断することができる。

次に顔文字と絵文字の感情度について述べる．本研究で収集したリプライには顔文字と絵文字が多く含まれていた．文の感情抽出において顔文字と絵文字の影響を考慮することは感情分析精度を向上させるために重要であり，そのようなアプローチをもつ研究が発表されている [12]．

以上の2つが本研究の感情分析精度を向上させるために考えられる改善策である．

## 6. まとめ

本研究では，感情に着目して Twitter のリプライから抽出した特徴量の国内の選挙予測に対する有用性評価を行った．評価実験を通してリプライの件数とリプライの感情度を用いた特徴量の有用性を示した．選挙運動へのインターネット利用が開始されてから約6年が経過した2019年参議院議員選挙に関するデータは Twitter だけでも大量に存在していた．今後，感情度計算手法の改善や有権者を観測する新たな指標の追加などによって，さらなる汎化性能向上が期待できる．

## 参考文献

- [1] 松尾 豊：ウェブからの実世界の観測と予測，電子情報通信学会論文誌 B，Vol. J96-B, No. 12, pp. 1309–1315 (2013).
- [2] Cameron, M. P., Barrett, P. and Stewardson, B.: Can Social Media Predict Election Results? Evidence From New Zealand, *Journal of Political Marketing*, Vol. 15, pp. 416–432 (2016).
- [3] Bermingham, A. and Smeaton, A.: On Using Twitter to Monitor Political Sentiment and Predict Election Results, *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology(SAAIP2011)*, pp. 2–10 (2011).
- [4] Tunggawan, E. and Soelistio, Y. E.: And the Winner is ...: Bayesian Twitter-based Prediction on 2016 U.S. Presidential Election, *IC3INA 2016 Conference* (2016).
- [5] 那須野薫，奥山晶二郎，中西鏡子，松尾 豊：Twitter における候補者の選挙地盤に着目した国政選挙の当選者予測，情報処理学会論文誌，Vol. 56, No. 10, pp. 2044–2053 (2015).
- [6] 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，Vol. 30, No. 2, pp. 116–128 (2014).
- [7] 船木洋晃，佐々木彬，岡崎直観，乾健太郎，深田陽介，竹下隆一郎，田森秀明，野澤 博：インターネット上の当選運動・落選運動の分析，人工知能学会全国大会論文集，Vol. JSAI2014, pp. 1K32–1K32 (2014).
- [8] Neubig, G. and Duh, K.: ツイートの情報量について—情報理論に基づく多言語調査—，自然言語処理学会第20回年次大会 発表論文集，Vol. 20, pp. 1138–1141 (2014).
- [9] 高村大也，乾 孝司，奥村 学：スピンモデルによる単語の感情極性抽出，情報処理学会論文誌ジャーナル，Vol. 47, No. 2, pp. 627–637 (2006).
- [10] 三和未沙希，立間淳士，青野雅樹：単語位置と強弱表現に着目したツイートの感情分析，第13回情報科学技術フォーラム講演論文集，Vol. 13, No. 2, pp. 227–228 (2014).
- [11] 黒橋禎夫，河原大輔：日本語形態素解析システム JUMAN++ version1.01 マニュアル (2016).
- [12] 山本湧輝，若井祐樹，熊本忠彦，瀬本明代：顔文字の役

割に着目したツイートの文の感情抽出手法の提案，第6回データ工学と情報マネジメントに関するフォーラム (DEIM2014) (2014).