

# 文体と意見極性に基づくツイートの分類

有馬 直也<sup>1</sup> 湯本 高行<sup>1</sup> 磯川 悌次郎<sup>1</sup> 上浦 尚武<sup>1</sup>

**概要:** 近年, Twitter の利用者数の増加に伴い, Web 上には利用者の興味や関心を示すツイートが多く投稿されている。また, Twitter はリアルタイム性が高いためソーシャルセンサとしても利用されている。しかし, ツイートには娯楽目的で投稿されたものが多く, マイニング対象を選定するには多大な負荷がかかる。そこで, トピックに対するユーザの態度と意見に着目する。ユーザの態度は文体に, 意見は意見極性に表れると考え, ツイートの文体と意見極性の分類を行う。本研究では, ツイートの文体が改まった表現か否か, また意見極性がポジティブかネガティブかに分類する手法を提案する。文体分類では, 新聞記事の文末表現に注目する。また意見極性分類ではサポートベクターマシンを用いるが, 固有名詞や否定語の影響を検討する。

## 1. はじめに

近年, スマートフォンや PC などの普及により, ソーシャルネットワークサービス (SNS) を利用する機会が増加している。このなかでも Twitter<sup>\*1</sup> は「ツイート」と呼ばれる 140 文字以内の短文を投稿し, 情報を共有するサービスである。ツイートは PC からだけでなく, 携帯端末からも投稿することができ, この投稿はインターネット環境があれば誰でも簡単に見ることができる。このため, 2016 年 9 月時点での国内利用者数は 4000 万人を超え, ユーザの周りで起こった出来事や, それに対してユーザが感知したことなど, 利用者の興味や関心を示すツイートが多く投稿されている。

また, Twitter の最大の特徴はそのリアルタイム性の高さであり, タイムラインと呼ばれる画面に, フォローしている人のツイートが時系列順に表示される。たとえば, オリンピックで選手が金メダルを取った瞬間の盛り上がりや, 地震発生時の現場の状況などをタイムライン上で即座に知ることができる。そのため, Twitter は利用者の目線で書かれた情報が物理センサの情報を補完するものとされ, 世論調査や疾病, 災害の把握といったソーシャルセンサとして防災分野に生かす取り組みが行われている。Twitter をソーシャルセンサとして用いる研究には, 6 因子の感情を用いて感情変動と経済動向との関係を分析する Bollen らの研究 [1] や, 災害時のツイートの位置情報から地震の検知や発生箇所の推定を行う Sakaki らの研究 [2] がある。

また, ツイートには娯楽目的の投稿や友人とのコミュニ

ケーションが多く, すべてのツイートが重要な内容を含んでいるとは限らない。しかし, マイニングの際にはそのようなツイートを選別しておらず, また選別には多大な負荷がかかる。

そこで, 本研究ではトピックに対するユーザの態度と意見に着目する。ユーザの態度は文体に, 意見は意見極性に表れると考え, ツイートの文体と意見極性の分類を行う。たとえば, あるトピックに対して「けが人が出なくて本当によかった」という文は肯定的であり, さらに真面目な感想を述べていると考えられる。また, 「この事故マジやばくね w」という文はトピックに対して否定的で, さらに冗談の意味を込めた投稿であると考えられる。本研究では, このようなツイートの分別を行うため, ツイートの文体が改まった表現か否か, また意見極性がポジティブかネガティブかの 2×2 の 4 クラスに分類する手法を提案する。文体分類では, 新聞記事を基にした文語体に対する辞書を構築し, この辞書に含まれる表現が使用されているか否かで分類を行う。意見極性分類では, 分類器にサポートベクターマシン (Support Vector Machine, SVM)[3] を使用するが, さらに素性として最適な語の調査や, 文体や否定語を考慮した学習による分類器を構築する。

ツイートの極性分類の研究は数多く行われている。橋本らの研究 [4] では極性辞書を用いて評価情報の抽出を行っている。一方, 機械学習を使う岡村ら [5] や Go ら [6] の研究では, 特に学習データの収集に注目している。本研究では, SVM を用いて機械学習を行うが, さらに素性として最適な語の調査や, 文体や否定語を考慮した学習による意見極性分類器を構築する。

<sup>1</sup> 兵庫県立大学大学院工学研究科

<sup>\*1</sup> <https://twitter.com/>

## 2. 文体分類器の構築

本研究での文体分類では、ツイートの文末表現に着目する。本手法では、新聞記事を基に文末表現辞書を構築し、この辞書を用いてツイートの文体の分類を行う。

### 2.1 文体クラス

本研究では、文体を文章の丁寧さと定義し、ツイートの文体を以下の2クラスに分類する。

- formal クラス
- informal クラス

ユーザがトピックに対しての真剣な意見・感想を発信する際、ツイートの文章は新聞やニュースのような改まった表現が用いられることが多いと考えられる。そこで、formal クラスは「～です」や「～だ」のような改まった表現や「言った」、「高い」のように動詞、形容詞の原型で終わる文章と定義する。また、ユーザがジョークなど娯楽目的で情報を発信する場合、普通の会話で使用している話し言葉を文章で表した口語体で投稿し、さらに文末にはネット特有の感情表現や顔文字を併用することも多いと考えられる。そこで、informal クラスは「～だよね」といった口語体やさらに砕けた表現で書かれている、あるいは顔文字などが含まれている文章と定義する。

利用者がマイニング対象としてのツイートやトピックに対しての真面目な意見・感想を閲覧したい場合は formal な文を提示し、逆に娯楽目的でツイートを閲覧したい場合は informal な文を提示することで知りたい情報のみを探すことができると考えられる。

### 2.2 文末表現の定義

文体分類では、ツイートの文末表現によって formal か informal かの分類を行う。本手法では、JUMAN [7] を用いて文ごとに形態素解析を行い、末尾の形態素からさかのぼり、動詞、形容詞、名詞のいずれかまでを文末表現と定義する。このとき動詞、形容詞であれば活用形も考慮したものを使用する。たとえば、「これはいい記事だ」という文では「だ」の直前が名詞「記事」であるため、「(名詞) + だ」を文末表現として抽出する。また、「これはおいしいですね」という文では「ですね」の直前が形容詞「おいしい」の基本形であるため、「(形容詞の基本形) + ですね」を文末表現として抽出する。

また、文章によっては上記の方法で抽出すると不必要に長い文末表現となってしまうことがある。たとえば、「食べられているそうさ」や「食べられるようになった」という文から文末表現を抽出すると「られているそうさ」、「られるようになった」が得られる。そこで、MeCab[8] で同文を解析した際、以下の条件を満たす場合にその部分を区切りと設定し、文末表現を抽出する。

- JUMAN で接尾辞、MeCab で接続助詞となった場合
- JUMAN で助動詞、MeCab で名詞となった場合

このルールに従って前述の「食べられているそうさ」という文から文末表現を抽出すると、JUMAN では「られて」が接尾辞、MeCab では「て」が接続助詞となるため「いるそうさ」を文末表現として抽出する。同様に、「食べられるようになった」という文から文末表現を抽出すると、JUMAN では「ように」が助動詞、MeCab では「よう」が名詞となるため「になった」を文末表現として抽出する。

### 2.3 文末表現辞書を用いた文体分類

本手法では、ツイートの文体を分類するために、文語体に対する文末表現辞書を構築する。このとき辞書には、新聞記事に対して形態素解析を行って得られた文末表現を使用する。しかし、新聞記事内にも会話文やコラムが存在するため、文語体でない表現が抽出されることがある。これらを除去するため、新聞記事内の「」内の文章から同様に文末表現を抽出し、口語辞書を作成する。この口語辞書を用いて、文末表現辞書に含まれている口語表現を除去する。ただし、口語辞書内にも文語表現が存在するため、口語辞書内で出現頻度が高く、文末表現辞書内で頻度の低い表現を対象として除去を行う。さらに、分類では文末表現辞書内で出現頻度の高い表現のみを使用する。

文体分類ではこの辞書を用いる。まず、ツイートの各文に対して文体分類を行い、それらの結果でツイートの文体の分類を行う。この詳細を以下に示す。

- (1) ツイートの各文から文末表現を抽出。
- (2) 各文に対して文体分類を行う。具体的には、名詞で終わっている文は対象外とし、それ以外の各文の文末表現が以下の条件のいずれかを満たすときその文を formal、そうでなければ informal とする。

- 文末表現辞書に含まれる表現である
- 動詞、形容詞の原型である

- (3) (2) で判定した各文の文体からツイートの文体を決定する。具体的には、informal な文が1つでも存在する、あるいはすべての文の文末が名詞であるときそのツイートを informal、そうでなければ formal とする。

また文末表現を抽出する際、砕けた文章であると JUMAN で正しく解析できないことがある。そこで、文末が動詞、形容詞、名詞であるとき MeCab 辞書を併用し、その解析結果が同じである場合に動詞、形容詞、名詞と判断する。

## 3. 意見極性分類器の構築

本手法では、ツイートの意見極性分類に SVM を使用する。SVM の特徴ベクトルの素性には、ツイートに対して形態素解析を行って得られた形態素の原型を使用する。また、分類器の構築方法として学習データをそのまま用いて構築するだけでなく、文章に否定語が存在する場合にその

意見極性ラベルを反転して学習させた分類器や、文体ごとにデータを分けて学習させた文体ごとの意見極性分類器を構築する。

### 3.1 意見極性クラス

本研究では、ツイートの意見極性を以下の2クラスに分類する。

- Positive クラス
- Negative クラス

ユーザがトピックに対して「楽しい」や「面白い」などの反応を示す内容は Positive と考えられる。そこで、Positive クラスはツイートが肯定的、楽観的な内容を表すものと定義する。また、ユーザが「疲れた」や「不便だ」などの反応を示す内容は Negative と考えられる。そこで、Negative クラスはツイートが否定的、悲観的な内容を表すものと定義する。

ユーザがある出来事の良い点だけを知りたい場合、Positive の情報だけを提示し、逆に悪い点だけを知りたい場合、Negative の情報だけを提示することで必要な情報を探することができると考えられる。

### 3.2 特徴ベクトルに使用する素性

SVM の特徴ベクトルの素性には、ツイートに対して JUMAN を用いた形態素解析を行って得られた形態素の原型を使用する。岡村らの手法 [5] では「名詞」、「形容詞」、「顔文字/www」を素性として用いている。本手法でも「名詞」、「形容詞」は素性として使用するが、使用したデータセット内のツイートに「顔文字」があまり見られなかったことから、これを素性として除外する。

また本研究では、これらの素性から以下の2点に着目し、最適な素性を調査する。

- 「名詞」において普通名詞や人名/地名、組織名を含めるか否か
- 素性として「動詞」を加えるが、極性を持たない語も含めるか否か

「名詞」の中で、普通名詞や人名/地名、組織名といった語は Positive, Negative 両方のツイートでみられ、さらに人に対して極性を割り当ててしまうのは不適切である。このため、素性としてこれらの語を取り除いた場合の精度と比較し、素性に用いるか否かを決定する。また、「動詞」では「買う」、「行く」といった極性を持たない語も存在し、これらは印象の判断根拠にならないと考えられる。よって、「動詞」の中でも極性を持つ語のみを使用した場合の精度と比較し、素性として使用するか否かを決定する。

本研究では、「動詞」の極性の有無については単語感情極性対応表<sup>\*2</sup>[9][10] を使用する。抽出した「動詞」がこの辞

<sup>\*2</sup> [http://www.lr.pi.titech.ac.jp/~takamura/pndic\\_ja.html](http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html)

書に含まれていれば極性を持つとみなし、素性に用いる。

### 3.3 否定語を考慮した分類器の構築

ツイートの中には「体調が良くない」や「悪くないと思う」といったように「ない」などの否定語がみられることもある。これらの内容は意見極性として、前者は Negative, 後者は Positive であると考えられるが、これらのツイートに対して形態素解析を行うと、素性として（体調、良い）や（悪い、思う）といった語が抽出される。このとき（良い）は Positive, （悪い）は Negative の要因となる語であるため誤学習を起こしてしまう。また、否定語を素性として考慮する場合には否定語を含むツイートを多数集めて学習させる必要がある。

そこで、ツイート内に奇数個の否定語が存在する場合、意見極性が反転することは自明だと考え、分類器の素性には組み込まず、ルールとしてその意見極性ラベルを反転して学習させる。その後、分類器が出力したラベルを再度反転する。ただし、文章が複数存在するツイートでラベルを反転させると、他の文章の素性にも影響してしまうため、単一文のツイートのみを対象としてラベルの反転を行う。

ここで、本手法では否定語として JUMAN を用いて形態素解析を行い、得られた形態素が形容詞の「無い、駄目だ」、接尾辞の「ない」、助動詞の「ぬ」を用いる。

### 3.4 文体を考慮した分類器の構築

ツイートの formal な文と informal な文を分けると、それぞれの文体ごとで使用されている語に偏りがあると考えられる。たとえば、「かわいい」や「楽しい」といった語は informal のツイートに多くみられる。また、「語る」や「認める」といった語は formal のツイートに多くみられる。

そこで、本手法では文体ごとに意見極性分類器を構築する。formal の意見極性分類器では、formal とラベル付けされたデータから Positive のツイートを正例、Negative のツイートを負例として学習させる。informal の意見極性分類器も同様に、informal とラベル付けされたデータから Positive のツイートを正例、Negative のツイートを負例として学習させる。

また、文体ごとに分けた場合の学習データ数は文体ごとに分けずに学習させた分類器と比べ、おおよそ半分になってしまう。そこで、別クラスのデータも負例として取り入れ、同規模の分類器を構築する。

## 4. 評価実験

分類精度の評価は適合率、再現率の調和平均である F 値によって行う。この F 値が大きいほど、分類精度が良いといえる。これらの評価指標の定義を以下に示す。

**適合率 (Precision)** : 分類器の出力結果のうち、実際に正解のデータである割合

表 1 データセット

	formal	informal
Positive	396	479
Negative	366	478

表 2 最適なパラメータを用いた際の分類精度

	適合率	再現率	F 値
formal	0.910	0.866	0.886
informal	0.897	0.932	0.914

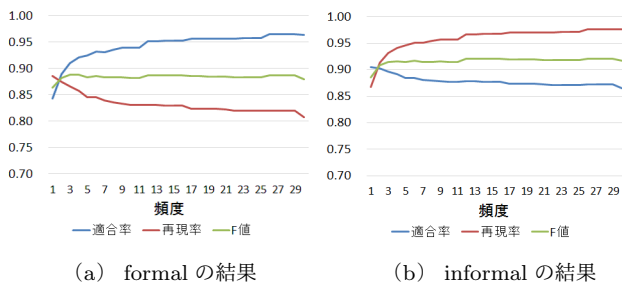


図 1 使用する頻度を変化させたときの分類精度の推移

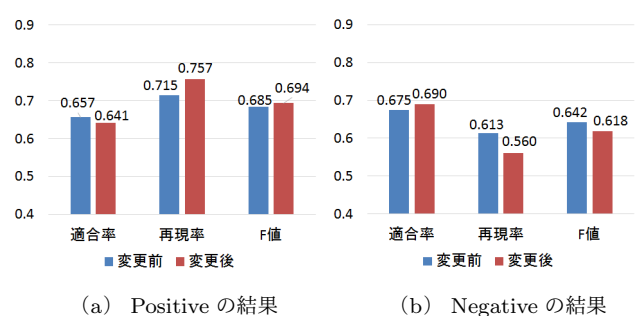


図 2 普通名詞などを除去した結果

**再現率 (Recall)** : 全ての正解データのうち、分類器が正解として出力した割合

**F 値** : 適合率と再現率の調和平均

#### 4.1 データセット

本研究で使用する各クラスのツイートを表 1 に示す。これらのツイートは、著者のうち 1 名が 2015 年 7 月のツイートを 1204 件、2016 年 6 月 9 日～7 月 21 日のツイートを 515 件収集し、ラベルも付与した。

また、あるニュース記事などに対してツイートする際、ツイートの中には URL やそのタイトルが含まれることがある。本研究では、ユーザのコメント部分のみを対象とするため、あらかじめそのリンク先のページに埋め込まれた title タグからタイトルテキストを抽出し、ツイートからタイトルの除去を行った。その際、ツイートの制限文字数の対策によりユーザがタイトルの一部を省略する場合があるため、部分一致も考慮した。

#### 4.2 文体分類の評価

2.3 節で述べた手法により、文末表現辞書を構築する。文末表現辞書の構築に使用する新聞は、2014 年 1 月 1 日～20 日の毎日新聞の記事 4445 件を使用した。これに対して形態素解析を行い、文末表現を抽出した。このとき、口語辞書による除外の対象として口語辞書内で出現頻度が 6 以上 (上位約 10%)、文末表現辞書内で出現頻度 19 以下の表現を除外した。また、口語辞書だけでは取り除くことができなかったものに対する対策として、辞書には出現頻度の高い表現のみを使用した。このとき、使用する表現の頻度を変化させたときの分類精度を図 1 に示す。

図 1 において、formal では使用する文末表現の頻度を限定していくと適合率が高くなり、再現率が低くなっている。informal では反対に、使用する頻度を限定すると適合率が下がり、再現率が高くなっている。このとき、最も精度が高かったのは頻度 3 以上の表現を使用した場合であ

た。そのときの分類精度を表 2 に示す。このとき、辞書に使用されている文末表現は 275 表現であった。

表 2 より、formal/informal の両方とも F 値が 0.8 以上の値をとっており、文末表現辞書を用いた分類では高い精度が得られた。しかし、formal の再現率では他の指標より低い値であることがわかる。この理由の多くは、「～だね」のように除去しきれていない口語表現が使用されていることや「～もらいたい」のように辞書に存在しない表現、もしくは辞書に存在するが出現頻度が低いため除外された表現が使用されていることが原因であった。

#### 4.3 意見極性分類の評価

意見極性分類においては、5 分割交差検定を行うことで評価指標を算出する。本研究では、SVM のカーネル関数として RBF カーネルを使用した。また、このときのハイパーパラメータは F 値が最大となるようにグリッドサーチを行い、得られた値を使用した。

##### 4.3.1 使用する素性の決定

特徴ベクトルに使用する素性のうち、3.2 節で述べた「名詞」において普通名詞などを含めるか否か、「動詞」において極性を持たない語を含めるか否かについて検討する。まず、「名詞」のうち素性として数詞以外を使用した場合と普通名詞、人名/地名、組織名を含めなかった場合との比較を図 2 に示す。

図 2 より、普通名詞や人名/地名、組織名を除いても F 値はほとんど変化しなかった。これより、普通名詞などは分類にほとんど影響がないと考えられる。したがって、これ以降の実験では普通名詞、人名/地名、組織名を除いたもの、すなわちサ変名詞と固有名詞のみを使用する。

次に、「動詞」において極性を持つ語のみを使用することで精度が改善されるか実験する。極性を持つか否かの判定には高村らによる単語感情極性対応表を使用した。この辞書では、単語に対して -1～1 までの感情値が割り当てられ

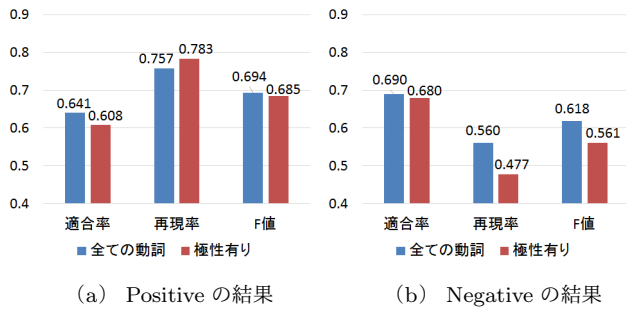


図 3 すべての動詞を使用した場合と極性を持つ動詞のみを使用した場合の精度の比較

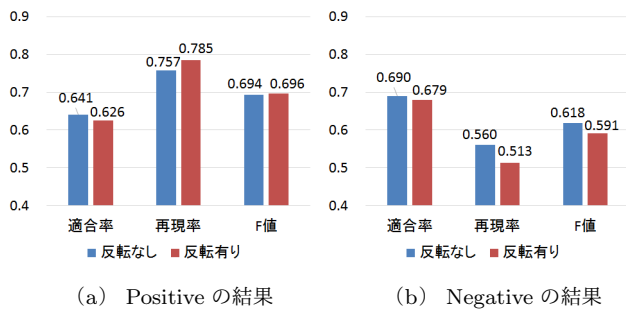


図 4 ラベルを反転して学習した分類器との比較

ている。ただし、この辞書では「書出す」や「受付ける」のように極性を持たない語に対しても感情値を割り当てているため、その感情値の絶対値が小さいものについては極性を持たないものとして扱う。本実験では、感情値の絶対値が 0.8 以上の語を極性を持つとして素性に使用した。このときの精度を図 3 に示す。

図 3 より、Negative については F 値が下がっており、すべての「動詞」を使用した場合のほうが分類器としての精度が高いことがわかる。この理由として、極性を持つ「動詞」のみを使用することで素性の数、すなわち特徴ベクトルの次元数が大きく減少したことが原因と考えられる。

以上より、素性として「名詞」はサ変名詞、固有名詞のみ、「動詞」は極性を持つか否かにかかわらず、すべての語を使用した場合が最も分類器の精度が良かった。これ以降の実験においては、これらの素性を用いる。

#### 4.3.2 否定語を考慮した分類器による分類

3.3 節で述べたように、文章中に奇数個の否定語が存在する場合、そのツイートの意見極性を反転して学習させた分類器を構築した。また、5 分割交差検定による分類結果からラベル反転を行ったツイートのラベルをさらに反転させ、精度を算出した。このときの精度と反転を行わなかったときの精度との比較を図 4 に示す。このとき、ラベルの反転を行ったツイートは 223 件存在し、正例のデータ数は 990 件、負例のデータ数は 729 件となった。

図 4 より、Positive については精度に差はみられないが、Negative については再現率が低下したため、F 値も低下した。したがって、分類器として否定語によりラベルを反転

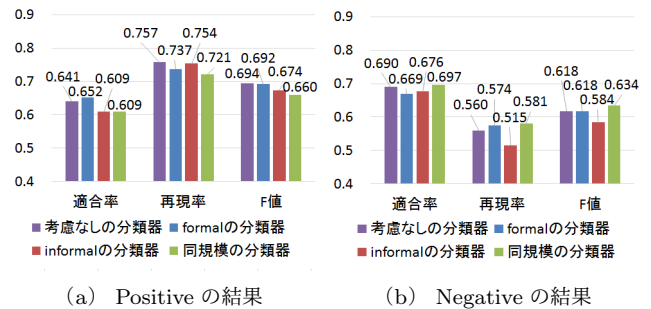


図 5 文体を考慮していない分類器との比較

表 3 2 クラスを負例とした formal の意見極性分類精度

	適合率	再現率	F 値
Positive	0.570	0.348	0.433
Negative	0.741	0.877	0.803

して学習させても精度は改善しないと考えられる。

#### 4.3.3 文体を考慮した分類器による分類

3.4 節に述べたように、formal の意見極性分類器として Positive-formal クラスを正例、Negative-formal クラスを負例として学習させた分類器を構築した。同様に、informal の意見極性分類器として Positive-informal クラスを正例、Negative-informal クラスを負例として学習させた分類器を構築した。そのときの精度を文体ごとに分けずに学習した分類器の精度と合わせて図 5 に示す。このとき、文体を考慮していない分類器は学習データ数が 1719 件であるのに対し、文体ごとに構築した分類器は学習データ数がそれぞれ約 800 件であり、約半分となっている。そこで、精度の違いが学習データ数の違いによるものか否かを判断するため、同規模の分類器としてデータを半分にして学習させた分類器の精度も合わせて示す。

図 5 において、最も精度が良かったのは文体を考慮せず、すべてのデータを学習に用いた分類器であった。また、同規模の分類器と比較すると Positive については精度が改善されているが、Negative については精度が下がっており、大きな変化はみられない。このことから、文体を考慮しても精度は改善しないと考えられる。

次に、文体ごとの分類器として負例のデータ数を増量して分類器を構築した。formal の分類器では、Positive-formal クラスを正例、その他 3 クラスを負例として学習させた分類器を構築した。同様に、informal の分類器として Positive-informal クラスを正例、その他 3 クラスを負例として学習させた分類器を構築した。このとき 5 分割交差検定を行うと、正例と分類された数は数件のみで、ほとんどのツイートが負例に分類されており、適切な分類ができていなかった。そこで負例データとして Negative-formal、Negative-informal の 2 クラスを用いて学習させた分類器の精度を表 3、表 4 に示す。

表 3、表 4 より、どちらの分類器においても約 1200 件

表 4 2 クラスを負例とした informal の意見極性分類精度

	適合率	再現率	F 値
Positive	0.640	0.372	0.470
Negative	0.712	0.882	0.788

表 5 4 クラス分類の精度

	適合率	再現率	F 値
Positive-formal	0.600	0.626	0.613
Positive-informal	0.566	0.733	0.639
Negative-formal	0.628	0.536	0.578
Negative-informal	0.628	0.492	0.552

表 6 4 クラス分類結果のラベルの詳細

PF:Positive-formal, PI:Positive-informal,  
NF:Negative-formal, NI:Negative-informal

		正解ラベル			
		PF	PI	NF	NI
分類器 の出力	PF	248	26	124	15
	PI	37	351	22	210
	NF	92	6	196	18
	NI	19	96	24	235

のうち 1000 件以上が負例と分類されたため偏りが生じ、Positive の精度が低い値となった。このように分類結果に偏りが生じた理由として、学習データにおいて負例のデータ数は正例のデータ数より 2 倍程度多く偏っていることが原因であると考えられる。

以上より、文体を考慮して分類器を構築しても精度は改善しないと考えられる。

#### 4.4 4 クラス分類実験

これまでの実験より、文体分類では文末表現辞書から出現頻度が 3 以上である表現を用いた分類が最も高い精度であった。意見極性分類では「名詞」はサ変名詞と固有名詞のみ、「動詞」ではすべての語が特徴ベクトルの素性として最適であった。また、分類器の構築方法は学習データに手を加えずに学習させた場合が最も高い精度であった。これらの分類器の出力結果を合わせ 4 クラス分類を行った。このときの分類精度とラベルの詳細を表 5、表 6 に示す。

表 5 より、F 値はどのクラスにおいても 0.6 程度という結果となった。また表 6 より、正解は Negative-formal であるが分類器で Positive-formal と分類されてしまったミスが 124 件あるといったように分類ミスは意見極性分類器による誤分類が最も多かった。

## 5. おわりに

本研究では、ツイートの文体が formal か informal か、意見極性が Positive か Negative かの  $2 \times 2$  の 4 クラスに分類する手法を提案した。文体分類の方法は、新聞記事を基にした文語体に対する文末表現辞書を構築し、ツイートから抽出した文末表現がこの辞書に含まれているか否かで分類

を行った。意見極性分類では、素性を変更し、最も高い精度となる素性を決定した。また、文体を考慮した意見極性分類器を構築する方法や、否定語が存在する場合にラベルを反転して学習させた分類器を構築する方法を試し、最も精度が良くなる方法を調査した。

実験では、文体分類の精度は F 値が 0.8 以上と高い精度であった。意見極性分類では、文体や否定語を考慮した学習方法を試したが精度の改善はみられず、単にデータを分けることやラベルを反転させるだけでは精度の改善に影響しないことがわかった。また、4 クラス分類では意見極性分類器による誤分類が多くみられた。

実験では、文体分類のミスは口語表現を除去しきれていないことや口語辞書によって formal な表現が誤って除去されていることが原因であった。したがって、口語辞書の構築方法と口語表現の除去方法の改善が今後の課題となる。また、意見極性分類では文体ごとの分類器構築の際に負例データを増やしたことで学習データに偏りができ、適切な分類ができなかった。そのためアンダーサンプリングを行い、データの偏りをなくした分類器での分類が必要であると考えられる。

謝辞 本研究の一部は、平成 29 年度科研費基盤研究 (C)(17K00429) によるものである。

#### 参考文献

- [1] Johan Bollen, Alberto Pepe, Huina Mao: Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena, Computing Research Repository, vol.abs/0911.1, 2009.
- [2] Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo: Earthquake shakes Twitter users: real-time event detection by social sensors, WWW2010, pp.851-860, 2010.
- [3] V.Vapnik: Statistical Learning Theory, Wiley, 1998.
- [4] 橋本 和幸, 中川 博之, 田原 康之, 大須賀 昭彦: センチメント分析とトピック抽出によるマイクロブログからの評判傾向抽出, 電子情報通信学会論文誌 D, Vol.J94-D, No.11, pp.1762-1772, 2011.
- [5] 岡村 康行, 湯本 高行, 新居 学, 上浦 尚武: ソーシャルタグを用いたツイートの印象の推定, 情報処理学会研究報告データベースシステム, 2015-DBS-162(9), pp.1-8, 2015.
- [6] Alec Go, Richa Bhayani, Lei Huang: Twitter sentiment classification using distant supervision, CS224N Project Report, Stanford, pp.1-12, 2009.
- [7] JUMAN,  
<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>
- [8] MeCab,  
<http://taku910.github.io/mecab/>
- [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] Hiroya Takamura, Takashi Inui, Manabu Okumura: Extracting Semantic Orientations of Words using Spin Model, In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL2005), pp.133-140, 2005.