

# 偏りのある特徴語を考慮した皮肉の検出

肥合 智史

嶋田 和孝

九州工業大学大学院 情報工学府

九州工業大学大学院 情報工学研究院

{s\_hiai, shimada}@pluto.ai.kyutech.ac.jp

## 1 はじめに

皮肉はあえて意図と反対であったり、遠回しな表現を用いて相手を非難することである。例えば、

**例 1** 口先だけで何もしない彼は、政治家の鑑だ。

という文において、書き手は「政治家」に対して、「～の鑑」という肯定表現を用いて、逆に、非難を示している。近年、文章から感情や意見を分析する、感情分析において、より高度な文章理解のための挑戦的課題として、皮肉を検出する研究が進められている [1]。

皮肉の検出は、多くの先行研究 [2, 3, 4] において、ある文章の内容が皮肉を含むかどうかの文章分類タスクとして取り組まれている。そして、分類には機械学習による手法が用いられている。学習の素性としては、文章に含まれる単語やフレーズといった基本的なものだけではなく、例 1 における、「～の鑑」のような、肯定・否定語など、特徴的な語に着目したものも用いられている。また、分類や学習のためのデータには、Twitter への投稿を利用するものが多い [2, 5]。Twitter への投稿には、その内容を表すハッシュタグを投稿者が付与することができる。そこで、「#sarcasm」という皮肉を表すハッシュタグが付いた英語での投稿を収集することにより、皮肉を含む投稿が獲得できる。

本研究では、Twitter への日本語の投稿を対象とした皮肉の検出を行う。機械学習による手法を用いて検出を行うが、その素性には、皮肉に偏って出現する皮肉の特徴語を考慮した素性を採用する。具体的には、皮肉の特徴語として、二種類の語に着目する。一つ目は、「(皮肉)」という表現によって示される語である。Twitter への日本語の投稿には、「(皮肉)」という表現を含むものがある。ここで、この「(皮肉)」という表現は、次の例のように、本文中に現れることがある。

**例 2** 見事 (皮肉) な部屋の散らかりようですね！

この例の「見事」のような、「(皮肉)」に係る特徴語に着目し、皮肉に偏って出現する語を利用する。二つ目は、皮肉としての批判の対象を表す語である。例 1 中の「政治家」という部分は皮肉によって批判される対象を示すが、この対象の存在は皮肉の重要な特徴とさ

れている [6]。本研究では、「政治家」のように、皮肉の対象として投稿中に現れる、立場を表す語に着目する。以上の二種類の特徴語について、それぞれ、「(皮肉)」直前に現れやすい語や、皮肉を言われやすい立場が存在することが考えられ、そのような皮肉に偏って現れる語を利用する。そして、抽出された偏りのある語のリストに基づく素性を新たに提案する。また、機械学習において、素性群を二つの学習器に分けて学習し、その合議を行う手法を取り入れることで、より高精度な分類を目指す。

## 2 関連研究

皮肉に関する研究では、機械学習が用いられることが多い。その基本的な素性としては、Bag-of-Words や、文章の長さが先行研究 [3, 7] で採用されている。それ以外にもさまざまな特徴に着目した素性も取り入れられている。例えば、次の皮肉例、

**例 3** 無視するなんて、君はいい性格してる。

において、肯定語「いい」は皮肉として否定的意味を示している。また、「無視する」のような否定語が伴っている。皮肉に関する素性として、Joshi ら [3] は、このような、文章中の肯定語や否定語の存在に注目し、その頻度や、異なる極性語の連続を素性に利用した。本手法でも同様にこれらの素性を利用する。

また、Karoui ら [4] は皮肉の特徴である批判される対象の存在について、固有名詞や代名詞の有無による素性を取り入れた。本手法では固有名詞や代名詞ではなく、一般名詞を対象とする。データ中の一般名詞に対して、立場を表す表現であるかどうかを判定し、それを利用する。

## 3 偏りのある特徴語の抽出

本節では、皮肉検出のための偏りのある特徴語リストの抽出について説明する。特徴語として、「(皮肉)」の直前の語と立場を表す語に着目し、皮肉に偏って出現する語のリストを抽出する。

### 3.1 「(皮肉)」直前の偏りのある語の抽出

「(皮肉)」の直前の偏りのある語の抽出には、Twitterの投稿を利用する。「(皮肉)」という表現を含む投稿を取得し、その文章中から直前の語を抽出する。

まず、「(皮肉)」という表現を含む投稿を3000件取得し、抽出用データとした。このデータから「(皮肉)」の直前の語のリストを抽出する。ただし、「(皮肉)」という表現が投稿の末尾につけられる場合、投稿中の特定の部分を指すわけではないことが考えられるため、前ページの例2のように末尾以外につけられる場合のみを利用する。「(皮肉)」が末尾以外につけられた投稿は、抽出用データセットの皮肉データ3000件のうち、924件であった。

このデータから、助詞、助動詞を除いた、「(皮肉)」の直前の単語リストを取得する。そして、この単語リストについて、皮肉に偏って出現する語の絞り込みを行う。そのために、「(皮肉)」を含まないTwitterへの投稿をランダムで取得し、皮肉データと同数の3000件の投稿を非皮肉データとした。そして、リスト中の各語の皮肉データと非皮肉データにおける出現頻度を算出し、非皮肉データにおける出現頻度が皮肉データにおける出現頻度以上のものをリストから除外する。さらに、皮肉データにおける出現回数が1回のもも、低頻度語としてリストから除外する。以上のように、「(皮肉)」の直前の偏りのある語を抽出し、257語のリストを取得した。

### 3.2 立場を表す偏りのある語の抽出

次に、皮肉の対象として、立場を表す偏りのある語の抽出について説明する。例1の皮肉における、「政治家」のような表現は、人の立場を表す表現である。このような表現は皮肉において、対立的な立場の書き手が、相手を批判する際に使われると考えられる。

このような立場表現を抽出するために、直前の語が立場であることを表す「にとっては」という表現と、並列助詞の「と」を利用する。まず、文章中の「にとっては」という表現の直前の名詞を抽出する。次に、この単語リストについて、対立的立場が存在するか確認する。対立的立場は「政治家と一般人」のように、並列助詞「と」の前後に現れることが考えられる。そこで、並列助詞「と」の前後に現れているかということを確認し、並列助詞「と」の前後に一度も現れていない場合、立場表現リストからその表現を除外する。

以上の手順で立場を表す語を抽出するが、この際、「にとっては」という表現や並列助詞「と」などは、Twitterなどの口語的でくだけた表現が多いデータより、書き言葉が集められたデータの方が出現しやすいと考えられる。よって、抽出には日本語の書き言葉が集められたBCCWJコーパス<sup>1</sup>を利用する。BCCWJは現代

<sup>1</sup><http://pj.ninjal.ac.jp/corpus.center/bccwj>

日本語の書き言葉の全体像を把握するために構築されたコーパスであり、多くのジャンルから得られた約1億語の書き言葉サンプルが収録されている。このコーパスから、上記の手順で立場を表す語を抽出する。次に、抽出された各語について、皮肉に偏って出現する語の絞り込みを行う。前節で利用した3000件ずつの皮肉データと非皮肉データそれぞれでの出現頻度を算出し、非皮肉データにおける出現頻度が皮肉データにおける出現頻度以上のものをリストから除外する。さらに、皮肉データにおける出現回数が1回のもも、低頻度語としてリストから除外する。以上のようにして、立場を表す偏りのある語を抽出し、472語のリストを取得した。

## 4 提案手法

提案手法の概要を図1に示す。本節では、まず、二つの学習器による合議制を導入した提案手法の枠組みについて説明する。次に、本手法で採用する素性の詳細を説明する。

### 4.1 提案手法の枠組み

本手法では、先行研究で用いられた素性に、提案素性を加えて分類モデルを構築する。ここで、採用する素性群には、Bag-of-Wordsのような0または1をとる素性と、文章に含まれる単語数のような、より幅広い値域をとる素性が存在する。

Tungthamthiti[8]らは、皮肉の検出において、素性群を二つに分けて、別々に学習を行い、それぞれの学習器の分類結果の合議によって、分類を行う手法を取り入れた。本研究でも、図1のように、先行研究の素性と提案素性を合わせた素性群を、0または1をとる素性と、幅広い値域をとる素性に分けて別々の学習器で学習させる。

二つの学習器のアルゴリズムにはどちらも、線形カーネルのSVMを用いる<sup>2</sup>。分類時に、二つの学習器による結果が一致した場合は、その結果をそのまま採用する。一致しなかった場合は、分離超平面までの距離が大きい方の結果を採用する。以上の方法で、学習と分類を行う。

### 4.2 SVM1の素性

分類器の一つであるSVM1には、1または0をとる素性を採用する。まず、文章全体に関する基本的な素性として、各単語が含まれるかどうかの素性であるBag-of-Wordsを採用する。

さらに、次の2種類の提案素性を採用する。

- 「(皮肉)」直前の偏りのある語  
3.1節で得た、「(皮肉)」直前の偏りのある語リス

<sup>2</sup>pythonの機械学習ライブラリ scikit-learn による実装を利用し、パラメータはデフォルトにした。

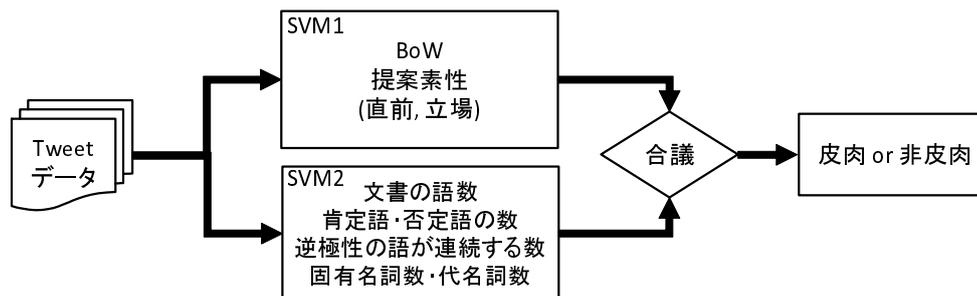


図 1: 手法の概要

トの 257 語それぞれが文章中に含まれているかどうかの素性である。

- 立場を表す偏りのある語  
3.2 節で得た、立場を表す偏りのある語リストの 472 語それぞれが文章中に含まれているかどうかの素性である。

### 4.3 SVM2 の素性

分類器の一つである SVM2 には、1 または 0 より幅広い値域をとる次の素性を採用する。

- 文章に含まれる単語数  
文章全体に関する基本的な素性である。
- 肯定語・否定語、逆極性の語が連続する数  
これらは、先行研究 [3] において採用されている。本手法では、肯定語、否定語の一覧には、日本語評価極性辞書 (用言編 [9]・名詞編 [10]) と Polar Phrase Dictionary[11] を利用する。
- 文章中の固有名詞と代名詞の数  
先行研究 [4] において採用されている。

## 5 実験

提案手法を用いて、Twitter への投稿が皮肉を含むかどうかの二値分類を行い、その精度を確認した。

### 5.1 データ

3.1 節の抽出用データと同様に、「(皮肉)」という表現が含まれる投稿を取得し<sup>3</sup>、それを基に、学習データと評価データを作成した。

- 学習データ  
「(皮肉)」という表現が含まれる投稿 3120 件を皮肉の正例とし、それと同数の「(皮肉)」を含まない投稿を取得し、それを負例として、学習データを構築した。ただし、学習時は「(皮肉)」という表現自体が分類の手掛かりとならないように、「(皮肉)」という表現は本文中から削除した。

<sup>3</sup>先行研究のように、本研究でも、皮肉を含む日本語の投稿を収集するために「#皮肉」というハッシュタグが付いた投稿を収集することを試みたが、投稿数が少なく、十分な量の投稿を収集できなかった。

- 評価データ  
精度評価のためのデータセットを作成した。3 人の作業者によって、「(皮肉)」という表現を含む投稿 1000 件ずつに対して、「(皮肉)」という表現を本文から除去した状態でアノテーションを行い、109 件の皮肉の正例を得た。また、同数の皮肉の負例を、「(皮肉)」を含まない投稿をランダムで抽出して獲得した。この 109 件ずつのデータを評価データとして利用した。

### 5.2 評価

まず、先行研究の素性のみでかつ、一つの学習器のみによる分類をベースラインとして、分類精度の比較対象とした。

ベースラインに採用した素性は、まず、文書全体に関する素性として、Bag-of-Word と文章に含まれる語数である。それらに加えて、評価表現と皮肉の対象に関する素性のみを先行研究から選んで採用した。まず、評価表現に関する素性として、Joshi ら [3] の肯定語・否定語、逆極性の語が連続する数を採用した。また、皮肉の対象に関する素性として、Karoui ら [4] の、固有名詞数・代名詞数を採用した。また、学習には、提案手法のそれぞれの分類器と同じく SVM を用いた。

さらに、ベースラインと提案手法の違いは、提案手法において、学習器を合議制にした点と、「(皮肉)」直前の語、立場表現に関する素性をそれぞれ採用した点である。それぞれの点についてその有効性を確認するため、ベースライン (手法 1) の分類精度に加えて、提案手法について、

- 提案素性を追加しない場合の合議 (手法 2)
- 「(皮肉)」直前の語の素性のみを追加した場合の合議 (手法 3)
- 立場表現の素性のみを追加した場合の合議 (手法 4)
- 「(皮肉)」直前の語と立場表現に両方の素性を追加した場合の合議 (手法 5)

それぞれの場合の評価データでの分類精度を確認した。

表 1: 結果

手法	P	R	F 値
手法 1(ベースライン)	0.74	0.74	0.74
手法 2(合議)	0.82	0.81	0.81 <sup>†</sup>
手法 3(合議+直前)	0.86	0.85	0.85 <sup>†*</sup>
手法 4(合議+立場)	0.85	0.84	0.84 <sup>†*</sup>
手法 5(合議+直前+立場)	0.85	0.84	0.84 <sup>†*</sup>

片側符号検定で有意差検定を行った。†が付いたものは手法 1 と  $p < 0.01$  で有意差があり、\* が付いたものは手法 2 と  $p < 0.05$  で有意差があることを示している。

## 6 結果

分類結果を表 1 に示す。精度は適合率 (P), 再現率 (R), F 値によって評価した。

ベースラインの F 値 0.74 を, その他の場合の F 値が上回った。よって, 合議制を採用したことが有効であったといえる。また, 提案素性を追加せず, 合議のみ採用した手法 2 の F 値 0.81 を, 提案素性のいずれかを追加した 3 つの手法 3~5 の F 値 0.85, 0.84, 0.84 が上回った。よって, それぞれの提案素性が有効であったといえる。ただし, 両方の提案素性を追加した手法 5 の F 値 0.84 は, どちらか一方だけの素性を追加した手法 3, 4 の F 値以下であり, 両方の素性を追加する有効性は確認できなかった。

## 7 考察

まず, 本研究の「(皮肉)」直前の語に関する提案素性では, 本文中の「(皮肉)」という表現を利用し, 抽出用データから偏りのある特徴語を抽出した。これは, 「(皮肉)」の直前の語はより分類に重要だという仮定に基づいている。この仮定が正しいことを確かめるために, 今回の抽出用データから, 「(皮肉)」の直前ではなく, ランダムで単語を抽出し, 同様の絞り込みを行って, 今回の「(皮肉)」直前の偏りのある語リストと同数の 257 語のランダム語リストを取得し, それを素性に用いて分類精度を確認した。その結果, F 値 0.84 で, 「(皮肉)」直前の偏りのある語リストを用いた場合の 0.85 に近い値となった。このことから, 偏りのある語の絞り込みは有用であるが, 「(皮肉)」の直前の語の有用性については, 明らかでなく, さらなる分析が必要であると考えられる。

また, 今回二つの学習器の合議による分類手法を採用したが, 二つの分類器は別々の素性群を採用していることから, それぞれの分類器の分類に及ぼす重要度は異なると考えられる。よって, 今回のような単純な合議ではなく, 何らかの重み付けを行うことで, 精度が向上するのではないかと考えられる。

## 8 おわりに

本研究では Twitter への投稿における皮肉の検出に取り組んだ。「(皮肉)」直前の偏りのある語と立場を表す偏りのある語, それぞれに関する素性を提案し, 二つの学習器の合議による枠組みを取り入れ, 分類を行った。その結果, 二つの提案素性と, 合議による分類, それぞれの有効性を確認した。今後は, 皮肉の分類に重要な表現の分析や, 学習器の重要度に関する分析と重み付けを行い, さらなる精度向上を目指す。

謝辞: 本研究の一部は科研費 26730176 の助成を受けたものです。

## 参考文献

- [1] A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, and A. Reyes. Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. In *Proceedings 9th Int. Workshop on Semantic Evaluation (SemEval2015), Co-located with NAACL*, pp. 470–478, 2015.
- [2] A. Reyes, P. Rosso, and T. Veale. A multidimensional approach for detecting irony in twitter. *Language Resources and Evaluation*, Vol. 47, No. 1, pp. 239–268, 2013.
- [3] A. Joshi, V. Sharma, and P. Bhattacharyya. Harnessing context incongruity for sarcasm detection. In *Proceedings of ACL-IJCNLP*, pp. 757–762, 2015.
- [4] J. Karoui, B. Farah, V. Moriceau, N. Aussenac-Gilles, and L. Hadrich-Belguith. Towards a contextual pragmatic model to detect irony in tweets. In *Proceedings of ACL-IJCNLP*, pp. 644–650, 2015.
- [5] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang. Sarcasm as contrast between a positive sentiment and negative situation. In *Proceedings of EMNLP 2013*, pp. 704–714, 2013.
- [6] J. D. Campbell and A. N. Katz. Are there necessary conditions for inducing a sense of sarcastic irony? *Discourse Processes*, Vol. 49(6), pp. 459–480, 2012.
- [7] Delia Irazu Hernandez Farias, Jose-Miguel Benedi, and Paolo Rosso. Applying basic features from sentiment analysis on automatic irony detection. In *Proc. 7th ibPRIA*, 2015.
- [8] P. Tunghamthiti, K. Shirai, and M. Mohd. Recognition of sarcasm in tweets based on concept level sentiment analysis and supervised learning approaches. In *Proceedings of the 28th PACLIC*, pp. 404–413, 2014.
- [9] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol. 3, pp. 203–222, 2005.
- [10] 東山昌彦, 乾健太郎, 松本裕治. 述語の選択選好性に着目した名詞評価極性の獲得. 言語処理学会第 14 回年次大会論文集, pp. 584–587, 2008.
- [11] 鍛冶伸裕, 喜連川優. HTML 文書からの評価表現辞書の自動構築. 言語処理学会第 13 回年次大会論文集, pp. 420–423, 2007.