

# 立場表現ペアの獲得とその偏りを考慮した皮肉検出

肥合 智史<sup>1</sup> 嶋田 和孝<sup>2</sup>

<sup>1</sup>九州工業大学大学院 情報工学府 <sup>2</sup>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、相手の欠点や弱点を意地悪く遠まわしに非難する表現である<sup>1</sup>。皮肉によって、肯定表現を用いて、逆に否定的な意味が伝えられる。

例 1 彼は部下に仕事を押し付ける素晴らしい上司だ。

例 1 においては、「素晴らしい」という肯定表現が用いられているが、上司への批判が示されている。文章中の極性表現は、主に文章の内容が肯定的か否定的か分類するタスクとして研究されている感情分析における重要な特徴である。しかし、皮肉では、表層と意図が一致しないため、感情分析の誤分類の原因となる。よって、近年、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 [2]。

先行研究において、皮肉検出は、対象文章が皮肉を含むかどうかの二値分類問題として取り組まれている。我々の以前の研究 [11] では、文章中の立場を表す表現のペアを考慮した皮肉の検出手法を提案した。例 1 においては、「部下」から、「上司」に対して皮肉が言われている。「上司と部下」のような立場ペアのリストを用意して、文章中の立場ペアを検出の手がかりとして利用し、その有効性を確認した。しかし、立場ペアリストの規模が小さく、文章中に立場ペアを含むと認識された皮肉データは少なかった (全体の 20% 強)。

本研究では、皮肉検出の精度向上のために、以前の研究より大規模な立場ペアリストの獲得に取り組む。まず、ブートストラップ法を用いて、大規模な立場ペアのリストを自動で獲得する。次に、獲得したリストを利用して、皮肉検出を行い、その精度を確認する。検出モデルの概要を図 1 に示す。Attention [1] 機構付き RNN [10] を利用した検出モデルで、文章中の単語と、文章に出現する立場ペアの関係を表すベクトルを入力とする。この関係を表すベクトルは、橋本ら [3] の手法で構築され、似た関係同士が似た連続値のベクトルで表現される。文章中の立場ペアをリストによ

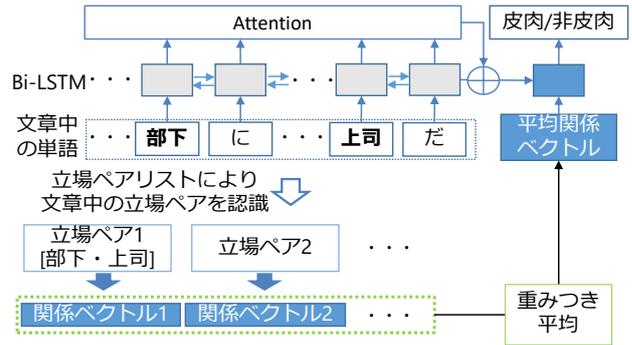


図 1: 皮肉検出モデルの概要

て認識し、その関係を表すベクトルを検出モデルに入力することで、どのような関係の立場ペアが皮肉に出現しやすいかという特徴を学習できる。ただし、ブートストラップ法を用いて獲得した大規模なリストには、多くのノイズが含まれることが考えられる。そこで、立場ペアの関係を表すベクトルに、各立場ペアに対応した重み付けを行い、新たなリストをより効果的に検出に利用することを試みる。

## 2 先行研究

本節では、我々の以前の研究 [11] の立場ペアリストの獲得方法と、リストを利用した検出モデルを説明し、本研究との違いを説明する。

### 2.1 立場ペアリストの獲得

以前の研究 [11] でも、立場ペアのリストを利用して文章中の立場ペアを認識した。以前の研究 [11] での立場ペアリストの獲得方法とその問題点を説明する。

抽出の概要を図 2 に示す。図 2 の上段のように、立場ペアは、Step1: コーパスからの立場表現の抽出と、Step2: 抽出された立場表現中での対となる組の決定という二段階で抽出される。Step1 では、直前の語が立場であることを表す「にとって」という手掛かり表現によって、「上司」などの単体の立場表現を抽出する。次に、Step2 において、「上司と部下」のような並立する立場表現のペアを決定する。コーパス中で並列助詞

<sup>1</sup><https://www.weblio.jp/content/皮肉>

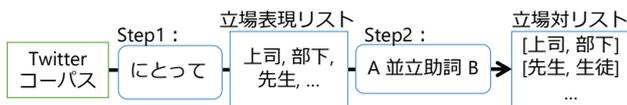


図 2: 立場表現ペアの抽出の概要

の前後に、二つの立場表現が出現したとき、その二つを並列の立場とみなし、立場ペアとして抽出する。

以上のように、立場ペアのリストを用意した。しかし、立場ペアリストの規模が小さく、文章中に立場ペアが含まれていると認識された皮肉データは少なかった。リストの規模が小さくなった原因として、Step1において利用した手掛かり表現が「にとつて」に限られていたことが考えられる。これによって、立場表現の抽出数が少なくなり、その組み合わせによって得られる立場ペアのリストも小規模になったと考えられる。

よって本研究では、ブートストラップ法を用いて多様な手掛かり表現の獲得に取り組み、より大規模な立場ペアリストを作成する。

## 2.2 皮肉検出モデル

以前の研究 [11] でも、図 1 に示されるものと同様の検出手法をとった。ただし、以前の研究 [11] では、文章中に複数の立場のペアが含まれていた場合、すべてのペアに対応する関係ベクトルの重みづけなしの平均をモデルへの入力とした。

本研究では、ブートストラップ法を用いて獲得された立場ペアリスト中のノイズの存在を考慮し、ノイズの影響を小さくするために、立場ペアに応じた重みつき平均を採用する。

## 3 提案手法

本手法では、多様な手掛かり表現の獲得のために、ブートストラップ法を利用する。ブートストラップ法によって、少量の手掛かり表現を初期値として与えることで、コーパスから新たな手掛かり表現を大量に抽出できる。その手掛かり表現を用いて立場表現を抽出し、ペアとなる立場表現を決定し、立場ペアのリストを獲得する。本節では、ブートストラップ法を利用した大規模な立場ペアリストの獲得と、立場ペアの重みを考慮した皮肉検出手法についてそれぞれ説明する。

### 3.1 ブートストラップ法を利用した大規模な立場ペアリストの獲得

ブートストラップ法によって、立場表現と手掛かり表現の抽出を繰り返すことで、多様な手掛かり表現を自動で抽出できる。図 3 に、処理の流れを示す。

まず、初期値となる手掛かり表現を与え、それを利用して、コーパスから立場表現の候補となる名詞を抽

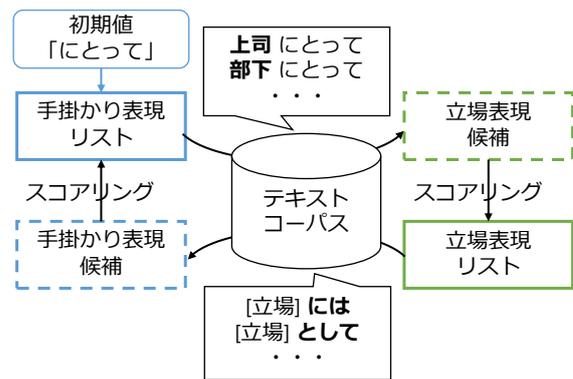


図 3: ブートストラップ法の処理の流れ

出する。本手法では、初期値には、先行研究において利用していた「にとつて」という表現を用いる。抽出された各候補について、より立場表現らしいものを選択するために、スコアをつける。このスコアリングについては後述する。スコアの上位 30 候補を立場表現とみなし、立場表現リストに加える。

次に、抽出された各表現の前後に出現する語句のパターンを新たな手掛かり表現の候補として獲得する。本手法では、前後の 2 単語までからなるすべてのパターンを獲得する。立場表現と同様に、抽出された各候補について、より手掛かり表現らしいものを選択するために、スコアをつける。このスコアリングについても後述する。このスコアの上位 30 候補を新たな手掛かり表現とみなし、手掛かり表現リストに加える。

以上のように、新たな立場表現と手掛かり表現の獲得を介して、手掛かり表現リストが拡張される。さらに、その手掛かり表現リストを用いて同様の処理を繰り返すことで、大量の手掛かり表現のリストを得る。

立場表現候補と、手掛かり表現候補へのスコアリングについて説明する。立場表現候補  $cand$  のスコア  $score_{cand}$  は、次の式で算出する。

$$score_{cand} = \frac{\sum_{key \in keywordList} Co(key, cand)}{Count(cand)}$$

$keywordList$  はその時点での手掛かり表現リストである。 $Count(cand)$  はコーパス中での  $cand$  の総出現回数である。 $Co(key, cand)$  はコーパス中での  $key$  と  $cand$  の共起回数である。抽出された各立場表現候補に対してこのスコアを計算する。手掛かり表現候補についても同様であり、その時点での立場表現リストを  $roleList$  として、スコアは次の式で算出される。

$$score_{cand} = \frac{\sum_{role \in roleList} Co(role, cand)}{Count(cand)}$$

ブートストラップ法によって、獲得された手掛かり表現リストを利用して、立場表現を抽出し、先行研究と同様に、並立助詞の前後に出現するものをペアとみなし、立場ペアリストを獲得する。

### 3.2 立場ペアの重みを考慮した皮肉検出

モデルへの入力、文章中の単語の分散表現と、文章中の立場ペアの関係を表すベクトルである。単語の分散表現は、word2vec [6] によって作成された 200 次元のベクトルである。立場ペアの関係を表すベクトルは、橋本ら [3] の手法によって、立場ペアごとに作成される。文章中に立場ペアが含まれていれば、それに対応する関係ベクトルを入力する。

前節の方法で獲得される大規模な立場ペアリストを用いることで、多くの皮肉データで新たに立場ペアが認識されることが考えられる。ただし、ブートストラップ法では、抽出結果をもとに処理を繰り返すため、抽出結果にノイズが含まれることが考えられる。そこで、皮肉検出においては、図 1 のように、モデルへの入力となる関係ベクトルに、文章中の立場ペアに対応した重み付けを行う。皮肉・非皮肉データにおける各立場ペアの出現頻度をもとに、各立場ペアの検出への寄与を考慮した重みを算出し、ノイズとなるペアの影響を小さくすることを試みる。

各立場ペアの重みには、特徴選択手法である Strength of Association (SOA) [7] を用いる。SOA は次の式で計算される。

$$\begin{aligned} SOA(x, y) &= PMI(x, y) - PMI(x, \neg y) \\ &= \log_2 \frac{freq(x, y) * freq(\neg y)}{freq(y) * freq(x, \neg y)} \end{aligned}$$

ここで、 $x$  は重みを計算する立場ペアを表し、 $y$  は皮肉クラスもしくは非皮肉クラスを表す。 $freq(x, y)$  は、立場ペア  $x$  のクラス  $y$  での出現頻度、 $freq(x, \neg y)$  は、立場ペア  $x$  のクラス  $y$  以外での出現頻度である。 $freq(y)$  は、データ全体での立場ペアの出現頻度、 $freq(\neg y)$  は、データ全体でのクラス  $y$  以外のデータの数である。SOA は、立場ペア  $x$  のクラス  $y$  に対する関連度 ( $PMI(x, y)$ ) と非関連度 ( $PMI(x, \neg y)$ ) の両方を考慮した、出現頻度の偏りを表す指標である。各立場ペアについて、 $y$  が皮肉クラスの場合と、非皮肉クラスの場合それぞれに SOA の値が算出されるため、その大きい方をペアの重みとする。ただし、低頻度な立場ペアについては、重みづけを行わない。皮肉・非皮肉データ両方で 2 回以上出現していない立場ペアは重みを 1 とし、2 回以上出現したものは重みを  $1+SOA$  スコア

表 1: 抽出結果

	立場表現数	立場ペア数	被覆率
以前 [11] のリスト	3,607	35,688	22%
今回のリスト	10,525	109,296	32%

とする。

## 4 実験

前節の手法により、立場ペアのリストを獲得し、リスト中の立場ペアを含む皮肉データの割合が、以前の研究より、大きくなっているかを確認した。また、新たに獲得したリストを用いて皮肉検出を行い、以前の研究と検出精度を比較した。

### 4.1 立場ペア抽出のためのデータ

立場ペアの抽出のためのコーパスには、Twitter への投稿を、ランダムに約 2,000 万件収集し、利用した。ブートストラップ法による手掛かり表現の抽出には、このうち 100 万件を用いた<sup>2</sup>。抽出された手掛かり表現を用いて、約 2,000 万件全体から立場表現を抽出した。立場ペアの決定にも、約 2,000 万件全体を利用した。

### 4.2 皮肉データ

皮肉の検出の研究には、Twitter<sup>3</sup>への投稿を検出の対象にしたものが多い [4, 8, 9]。本研究でも、Twitter への投稿から皮肉を含んだものを収集し、利用する。皮肉を含んだ投稿の収集のために、以前の研究 [11] では、「(皮肉)」という表現を利用した。日本語の投稿には、次の例のように、「(皮肉)」という表現を用いて皮肉であることを示すものがある。

例 2 うちの会社は役員がミスしてもお咎め無しのもと素敵な(皮肉)会社です。

この「(皮肉)」という表現を含む投稿を収集し、皮肉を含むデータ(皮肉ツイート)とみなしている。本研究でも、同様に皮肉ツイートを収集した。

### 4.3 立場ペアの抽出結果

立場表現、立場ペアの抽出数と、リスト中の立場ペアを含む皮肉データの割合を表 1 に示す。皮肉データには、前節の方法で収集した 5,000 件の皮肉ツイートを利用した。提案手法による抽出の立場表現数、立場ペア数が、以前の研究のリストより多くなった。また、皮肉データでの被覆率も 22% から 32% と大きくなった。

<sup>2</sup>ブートストラップ法を大規模なコーパスに適用すると、処理時間が長くなったため、一部のデータに適用した。

<sup>3</sup><https://twitter.com>

表 2: 分類結果

	適合率	再現率	F 値
以前 [11] のリスト	0.799	0.797	0.797
今回のリスト (重みなし)	0.794	0.794	0.794
今回のリスト (重みあり)	0.776	0.776	0.776

立場対が今回のリストによって新たに認識された皮肉ツイートを観察したところ、約 35%には一つのツイート中に複数の対が認識されていた。例えば、次のような例が存在した。

**例 3** まああそこの事務所はタレントに何か言う前に事務所の方に先に文句言いたくなるからタレント守ってるとも言えるけど

この例においては、二つの対「事務所とタレント」、「何と前」が新たに認識された。しかし、立場ペア「何と前」における、「何」は抽象的な代名詞で「前」も時期を表す表現であり、例中で立場を表しておらず、皮肉の評価には関係しないと考えられる。この「何と前」のようなペアは、ブートストラップ法によって抽出されたノイズであり、分類の際にもノイズとなることが考えられる。

#### 4.4 皮肉検出の精度への影響

新たに獲得したリストを皮肉検出に利用し、検出精度を確認した。双方向 LSTM のパラメータは次のように設定した。入力層の次元数 200, 中間層の次元数 150, 最適化手法には Adam [5] を利用し、学習率は 0.001 に設定した。開発データ, 訓練データ, 評価用データに、前節の方法で皮肉ツイートをそれぞれ、5,000 件, 15,000 件, 1,000 件収集し、正例とした。また、「(皮肉)」を含まないツイートも正例と同数ランダムで収集し、負例とした。3.2 節の重みの算出には、訓練データを利用した。我々の先行研究のリストを利用した検出と、3.1 節で得られた今回のリストを用いた検出を行った。今回のリストを用いた検出では、3.2 節の重みづけを適用しなかった場合と、した場合をそれぞれ確認した。検出結果は、適合率, 再現率, F 値で評価した。

結果を表 2 に示す。今回のリストを利用し 3.2 節の重みづけを適用しなかった場合の精度は、F 値 0.794 であり、我々の先行研究 [11] の精度である F 値 0.797 を下回った。この原因には、ノイズの影響があると考えられる。しかし、今回のリストを利用し重みづけを適用した場合の精度も、F 値 0.776 であり、先行研究の精度を下回った。3.2 節で説明した通り、重みづけ

には、各立場ペアの皮肉・非皮肉データでの出現頻度の偏りをもとにした手法をとり、皮肉・非皮肉データの両方で 2 回以上出現した立場ペアを重みづけの対象とした。しかし、多くの立場ペアがデータ中で 1 回しか出現しておらず、重みづけを適用できた立場ペア (2 回以上出現) は、109,296 ペア中の約 1,300 ペア程度であった。また、重みづけを適用できたものについても、信頼性の低い値となっていることが考えられ、今回の精度低下の原因と考えられる。

## 5 おわりに

本研究では、皮肉検出の精度向上を目的として、立場ペアリストの獲得に取り組んだ。以前の研究より大規模な立場ペアリストを自動で獲得し、皮肉データでの被覆率は上昇した。しかし、ノイズ等の影響により、皮肉検出の精度は低下した。よって今後は、検出に有用な立場ペアの選定方法を検討する必要がある。

## 参考文献

- [1] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. In *Proceedings of ICLR*, 2015.
- [2] 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 of SemEval2015, Co-located with NAACL*, pp. 470–478, 2015.
- [3] K. Hashimoto, P. Stenetorp, M. Miwa, and Y. Tsuruoka. Task-oriented learning of word embeddings for semantic relation classification. In *Proceedings of CoNLL 2015*, pp. 268–278, July 2015.
- [4] A. Joshi, V. Sharma, and P. Bhattacharyya. Harnessing context incongruity for sarcasm detection. In *Proceedings of ACL-IJCNLP 2015*, pp. 757–762, 2015.
- [5] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *The International Conference on Learning Representations (ICLR)*, 2015.
- [6] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In *The International Conference on Learning Representations: Workshops Track*, 2013.
- [7] S. M. Mohammad and S. Kiritchenko. Using hashtags to capture fine emotion categories from tweets. *Comput. Intell.*, Vol. 31, No. 2, pp. 301–326, May 2015.
- [8] 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.
- [9] 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.
- [10] M. Schuster and K. K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, Vol. 45, pp. 2673–2681, 1997.
- [11] 肥合智史, 嶋田和孝. 関係ベクトルを利用した皮肉の検出. 言語処理学会 第 24 回年次大会 発表論文集, pp. 829–832, 2018.