

発話者の発話意図推定に影響を及ぼす表情の絵文字の検出

廣瀬将也¹ 森辰則¹

¹ 横浜国立大学大学院環境情報学府

hirose-masaya-km@ynu.jp tmori@ynu.ac.jp

概要

絵文字は文字では伝えにくい意味や感情を表現することができるため、感情表現の簡略化やコミュニケーションの充実のために広く採用されている。絵文字は近年研究においても注目されており、自然言語処理分野では主に感情分析に利用されている。本論文では、絵文字を使った感情分析において、既存手法を利用した感情推定結果に影響を及ぼすと考えられる皮肉や自嘲の 😏 というレトリカルな用法の絵文字を検出する方法を提案する。

1 序論

SNS を始めとするオンラインコミュニケーションでは、表情や仕草といった非言語情報が制限されるため、情報の伝達に支障をきたす可能性がある。こうした状況の中で、非言語的なコミュニケーション機能を有する絵文字が活用されている。

多様な表現ができる絵文字は、感情推定においても重要な要素となりうる。Hussien ら [1] は、“Joy”, “Sadness”, “Anger”, “Disgust” の 4 種類の感情を表すラベル (感情ラベル) を絵文字に割り当て、文と絵文字の感情は一貫しているという仮定の元、割り当てた感情ラベルを絵文字を含む文の感情ラベルとみなして学習データに用いることで、文の感情を推定している。また Felbo ら [2] は、絵文字そのものを文のラベルとして文から絵文字への予測を行い、絵文字予測に使ったモデルを転移学習させることで、絵文字予測タスクから学習した言語表現が感情認識や皮肉検出タスクに有用であることを示した。

しかし、こうした研究の多くは、絵文字の多様性についての考慮が十分ではない。例えば 😊 という絵文字は喜びの感情を表す文と併に使われることが多いが、中には表 1 のように皮肉や自嘲の笑顔として利用される場合もある。

表 1 😊 の利用例

テキスト	笑顔の意図
めちゃくちゃ楽しかった 😊	喜び
まじでうざい 😊	皮肉
転んだの恥ずかしすぎ 😊	自嘲

絵文字に割り当てた感情ラベルを、それを含む文のラベルとして利用する場合、皮肉や自嘲の笑顔のようにレトリックとして使われる絵文字 (以下「レトリカル絵文字」と呼ぶ) は発話者の発話意図推定に影響を及ぼす可能性があり、こうした絵文字を認識して区別することは重要だと考えられる。本研究ではこのような発話者の発話意図推定に影響を及ぼす絵文字を自動的に検出するシステムを提案する。

2 関連研究

近年、絵文字を利用した感情分析手法が注目されているが、絵文字の多様性について考慮した研究は少ない。Felbo ら [2] は、コーパスの予備知識を必要としない絵文字予測を使った事前学習と転移学習によって、感情推定や皮肉検出で成果をあげている。事前学習の絵文字予測に関する考察では、複数の絵文字を使うことで多様な表現を獲得でき、後続のタスクでより良い成果が出ると言及しているが、1 つの絵文字が持つ多様性については詳しく触れられていない。また Chen ら [3] は 1 つの絵文字が正極性にも負極性にも使われることを考慮し (極性の説明は 3 章を参照)、極性ごとに異なる埋め込みを得ることを期待して、絵文字を埋め込む際、正極性用と負極性用に 2 つの埋め込みを割り当てている。しかし、絵文字に対する 2 つの埋め込みが極性の違いを捉えたものとは限らず、評価実験における精度の向上は学習パラメータ数の増加によるものの可能性があるため、より詳細な分析が求められる。本研究の提案手法は、絵文字が持つ多様性をより詳細に分析する際に活用することが期待できる。

3 絵文字の用法に関する分析

本章ではレトリカル絵文字の検出手法を検討するために、絵文字の用法に関する分析について説明する。

3.1 ツイートの収集とアノテーション

Twitter のデータを対象として、ツイートにおける絵文字の用法を分析した。ツイートデータは Twitter API を利用して収集し、2019 年 9 月から 11 月におけるツイートのうち 😊・😞・😡 のいずれかを含むものを合計 1000 件取得した。ツイートは複数の文で構成されているものが多く、1 ツイートに複数の絵文字を含む場合もあるため、収集したツイートから絵文字に対応すると考えられる 1 文（以下ツイート文とする）を手動で抽出した。

絵文字が感情推定に及ぼす影響を調べるために、それぞれのツイート文について「言語情報のみから推定される書き手の感情」および「言語情報と絵文字の組み合わせから推定される書き手の感情」の 2 通りのラベル付けを行った。このラベル付けは第一著者によって行われた。ラベルとして付与した感情は、中村による感情表現辞典 [4] で利用されている 10 種類のラベル“喜”・“好”・“安”・“哀”・“厭”・“怖”・“怒”・“恥”・“昂”・“驚”に 3 種類のラベル“N(ニュートラル)”・“願”・“他”を加えたものを利用した。最終的に絵文字 😊 を有する文 228 件、😞 を有する文 240 件、😡 を有する文 132 件について、各々に前述の 2 通りのラベルを付与したコーパスを作成した。ラベル付けの例を表 2 に示す。

表 2 ラベル付けの例

テキスト	感情ラベル
働きたいです!って言ったけどダメだった	哀
働きたいです!って言ったけどダメだった 😡	怒
朝ごはん美味しかった!!!	喜
朝ごはん美味しかった!!! 😊	喜

3.2 分析結果

付与した感情ラベルの割合を絵文字ごとにまとめた結果を表 3, 4 に示す。

表 3 より、😊 は 64.5% が喜、😞 は 57.9% が哀、😡 は 51.5% が怒の感情を表す文と伴に使われていることが分かり、それぞれの絵文字が現れる文の感情ラベルには偏りがあることがわかる。

次に、付与した感情を正極性・負極性の 2 つの極

表 3 言語情報のみから推定される書き手の感情と絵文字の関係(割合)

	N	喜	好	安	哀	恥	厭	怖	怒	昂	驚	願	他
😊	7.5	64.5	15.8	2.2	2.6	0	0.4	0.9	2.6	0	0	0.9	1.8
😞	2.1	5.8	6.7	1.3	57.9	0	3.8	6.7	0.8	7.1	1.7	6.3	0
😡	6.1	1.5	2.3	0	21.2	0.8	12.9	0	51.5	0.8	0.8	2.3	0

表 4 言語情報と絵文字の組み合わせから推定される書き手の感情と絵文字の関係(割合)

	N	喜	好	安	哀	恥	厭	怖	怒	昂	驚	願	他
😊	0	72.8	15.8	2.2	2.6	0	0.4	0.9	2.6	0	0	0.9	1.8
😞	0	10.4	6.7	1.3	57.9	0	3.8	6.7	0	7.1	0	6.3	0
😡	0	0	2.3	0	0	0.8	7.6	0	87.1	0	0	2.3	0

性に分割し、言語情報のみから推定される書き手の感情極性と絵文字の関係を調査した。付与した感情のうち、“喜”・“好”・“安”を正極性、“哀”・“恥”・“厭”・“怖”・“怒”・“昂”を負極性とし、極性の定義できない“N”・“願”・“驚”・“他”は対象から除いた。

このとき、各極性の割合を(正極性, 負極性)とすると、😊 は(82.5%, 6.6%), 😞 は(13.8%, 76.3%), 😡 は(3.8, 87.1%)となり、絵文字が使われる文の極性で考えると、偏りがより顕著に現れることが分かる。

ここで、ある絵文字が付与された文について、言語表現のみから推定された書き手の感情が正極性であるか負極性であるかを集計した結果、大多数を占める感情極性(正極性もしくは負極性)をその絵文字が付与された文集のマジョリティ極性とし、マイノリティの感情極性を文集のマイノリティ極性とする。以下では単にそれぞれ「絵文字のマジョリティ極性(マイノリティ極性)」と呼ぶ。

😊 の各極性の割合より、😊 のマジョリティ極性は正極性、😞 のマイノリティ極性は負極性となるが、表 3 と表 4 を比較すると、絵文字が付与される前に感情ラベルが“N”だった文に 😊 を付与した場合はそのラベルが“喜”に変化しているのに対して、😊 のマイノリティ極性である負極性に 😊 が付与されても極性が変化していないことが分かる。負極性に 😊 が付与される事例についてさらに分析を行ったところ、負極性で利用される 😊 は皮肉や自嘲のニュアンスを持つことが分かった。

3.3 レトリカル絵文字検出のための仮説

上記の結果より、皮肉や自嘲のニュアンスを持つ 😊 は、怒りや悲しみといった負極性を持つ文と伴に使われていることが分かった。そのため、言語情報から怒り・悲しみといった負極性を予測できれば、皮肉・自嘲の 😊 が検出できると考えられる。

また、😡 は怒り、😞 は悲しみを感情ラベルとして

持つ文とともに使われることが多く、文と絵文字の感情ラベルが逆の極性を持つようなレトリカルな絵文字の利用があったとしても少数である。絵文字を感情ラベルの代替物とした場合、一般的な機械学習ではマジョリティの事例に偏る判断をする傾向にあるため、絵文字が利用される文の感情に偏りがあれば、言語情報から絵文字を予測する分類器は、文に感情ラベルを付与する分類器と同等なものとして有効に機能することが期待できる。

以上のことから、次のような仮説が導かれる。

😊 が付与されている文について、絵文字分類器が 😡 もしくは 😞 を予測した場合、文の感情ラベルは怒りや悲しみと対応している可能性が高く、文に付与されている絵文字 😊 は皮肉や自嘲を表すレトリカルな用法である可能性が高い。

本論文ではこの仮説を元にレトリカル絵文字を検出する。

4 レトリカル絵文字の検出手法

本章では、3.3 節で述べた仮説を元に、絵文字分類器を利用したレトリカル絵文字の検出手法を説明する。なお、本論文では 😊 がレトリカルな用法で使われているものを検出対象として扱う。

まず、5.1.1 項で説明する前処理によって、絵文字を含む文を「絵文字を除いた文」と「文のラベルとなる絵文字(1つ)」に分割する。提案モデルはこの2つを入力として、入力した絵文字がレトリカル絵文字であるかどうかを判定し、その結果を出力する。

レトリカル絵文字の判定に利用する絵文字分類器は、入力文に対する適切な絵文字を2つの絵文字から選択し、選択した絵文字とその予測確率を出力する二値分類器で作成する。二値分類としたのは、絵文字予測タスクでは多値分類よりも二値分類の方が予測性能が高いという先行研究に基づいている [5]。

絵文字分類器の出力候補となる2つの絵文字の組み合わせは、😊・😡、😊・😞、😊・😏 のいずれかとし、各組み合わせごとに分類器の学習を行う。😊・😏 の組み合わせは、マジョリティ極性が類似した絵文字の組み合わせではレトリカルな絵文字を検出しづらくなるという予想を検証するために利用する。

次に、学習済みの絵文字分類器に対して、学習に利用していない 😊 を含む文から絵文字を除いた文を入力する。このとき、入力文のラベルとして 😊 を保持しておく。絵文字分類器が 😊 以外の絵文字を出力した場合、入力文のラベルとした 😊 をレトリカルな

😊 とみなす。この出力(😊 を陽性としたときの偽陰性)は分類器の間違いそのものである場合も考えられるが、提案モデルではこれをレトリカル絵文字とみなす。このとき、偽陰性である絵文字の予測確率を、入力文が 😊 のマイノリティ極性をもつ文である確率と同等のものであると期待し、これを「レトリカル絵文字である確率」とする。以上の操作を経て、提案モデルは、入力した絵文字がレトリカル絵文字であるかどうかの判定結果とその確率を出力する。

5 評価実験

本章では、提案モデルがレトリカル絵文字の検出に有効かどうかを検証するための実験を行う。

5.1 実験設定

5.1.1 データセット

本論文で利用する絵文字付きデータは、Search Tweets API を利用して収集した。学習・検証用のデータセットとして、4 章で示した4つの絵文字についてそれぞれの絵文字を含むツイートを110万件ずつ取得した。ツイートは2021年10月31日から遡って取得したものを利用し、ハッシュタグ・リンク・メディアを持つツイート、リプライやリツイートであるツイートはクエリを利用して除外した。

次に、学習データの大きさによる検出率の変化を調べるために、各絵文字ごとに訓練用に100万件ずつ・検証用に10万件ずつランダムサンプリングした LARGE サブセット(L)と、各絵文字ごとに訓練用に10万件ずつ・検証用に2万件ずつランダムサンプリングした SMALL サブセット(S)を作成した。これらのサブセットを 😊・😡、😊・😞、😊・😏 の3つの絵文字の組み合わせごとに統合することで、各サブセットの総数は LARGE サブセットが訓練用に200万件・検証用に20万件、SMALL サブセットが訓練用に20万件・検証用に4万件となり、最終的に6種類のサブセットを作成した。

また、評価用のデータセットとして、2021年11月1日から2021年11月30日における 😊 を含むツイートを同様の条件で100万件取得した。評価時にデータセットの重複を避けるため、ランダムに並び替えた評価用データセットを5万件ずつのサブセットに分割し、分割したサブセットを各モデルごとの評価に利用する。

最後に、絵文字がツイート毎に1つになるように

収集したツイートに対して次の前処理を行った。まずクエリとして利用した絵文字のうち末尾のものを除いた全ての絵文字を削除する。1つだけ残った絵文字を文のラベルとして保持しておき、残った絵文字を EMOJI という特殊トークンに置き換える。

5.1.2 実装モデル

絵文字分類器は、埋め込み層、LSTM 層、全結合層、ソフトマックス層から構成される DNN である。絵文字を除いた文と、文のラベルとした絵文字を入力として、入力した絵文字がレトリカル絵文字であるかどうかの判定とその確率を出力する。文は形態素解析器 MeCab[6] と、単語分かち書き辞書 mecab-ipadic-NEologd[7] を用いて単語分割を行い、埋め込み層へと入力する。埋め込み層の次元数は 300、LSTM の隠れ層の次元数は 128 を利用した。

5.1.3 評価方法

5.1.1 項で作成した評価用データに対して、モデルを用いてレトリカル絵文字の判定を行い、以下の式でレトリカル絵文字 🤔 の検出率を評価する。

$$\text{検出率} = \frac{\text{実際にレトリカル絵文字 🤔 であった事例数}}{\text{モデルがレトリカル絵文字 🤔 と判定した事例数}}$$

実際にレトリカル絵文字であるかどうかは人手で判定し、学習データの異なるモデルごとにそれぞれ 300 件ずつ行った。このとき、🤔 全体におけるレトリカル絵文字の割合を調べるため、評価用のサブセットからランダムに 300 件抽出したのものについても同様に判定を行った。

また、レトリカル絵文字の検出率とモデルの出力における「レトリカル絵文字である確率」の関係性を調べるため、各モデルには評価用サブセット 5 万件を入力し、その出力を確率順に並び替え、上位 (top) と下位 (worst) からそれぞれ 100 件ずつ抽出したもの、それらを除いたものからランダム (random) に 100 件抽出したものを評価した。

5.2 実験結果

6 種類のサブセットを利用して絵文字分類器の学習を行い、5.1.3 項の方法によって評価を行った。評価結果を表 5 に示す。

なお、表 5 における all は、top・worst・random の合計 300 件におけるレトリカル絵文字の割合を表す。また、評価用のサブセットからランダムに 300 件抽出したのものにおけるレトリカル絵文字 🤔 の割合は

3.3%であった。

表 5 レトリカル絵文字の割合

学習データ		検出率			
絵文字	サイズ	top	worst	random	all
🤔・😏	L	41%	18%	34%	31%
	S	27%	4%	14%	15%
🤔・😞	L	58%	12%	14%	28%
	S	50%	7%	19%	25.3%
🤔・😓	L	7%	6%	7%	6.7%
	S	3%	4%	4%	3.7%

5.3 考察

まず、各モデルにおける all の値に注目すると、ランダムに抽出した 🤔 におけるレトリカル絵文字の割合 (3.3%) と比較して、🤔・😏、🤔・😞 の組み合わせではレトリカル絵文字の検出率が大きく上回っている。これに対して、🤔・😓 の組み合わせでは、3 章の分析で 😓 が負極性で使われていた割合が 6.6% だったことも考慮すると、ランダムに 🤔 を取得した場合と殆ど変わらないといえる。この結果より、極性の異なる絵文字の組み合わせを学習に利用することが重要であると考えられる。

また、どの絵文字の組み合わせも学習用のデータサイズを大きくすることで、レトリカル絵文字の検出率が向上していることが分かる。これは、学習用のデータサイズを大きくすることで、それぞれの絵文字のマジョリティ極性に対する識別性能が向上し、結果としてマイノリティ極性を持つレトリカル絵文字の検出率が高まっているからだと考えられる。

さらに、特に 🤔・😏、🤔・😞 の組み合わせについて、レトリカル絵文字の確率が top のものと worst のもので大きく差が生じており、提案モデルの予測確率は有効に機能していると考えられる。

6 まとめ

本論文では、絵文字を使った感情分析において、既存手法の感情推定結果に影響を及ぼすと考えられる皮肉や自嘲の 🤔 といったレトリカルな絵文字を検出する方法を提案した。提案手法では、極性の異なる 2 種類の絵文字を学習データとして絵文字の二値分類器を作成し、分類器の出力における偽陰性を用いてレトリカルな絵文字を検出する。実験の結果、提案手法がレトリカルな絵文字の検出に有効であることが示された。

参考文献

- [1] Wegdan A Hussien, Yahya M Tashtoush, Mahmoud Al-Ayyoub, and Mohammed N Al-Kabi. Are emoticons good enough to train emotion classifiers of arabic tweets? In **2016 7th International Conference on Computer Science and Information Technology (CSIT)**, pp. 1–6, 2016.
- [2] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 1615–1625, 2017.
- [3] Yuxiao Chen, Jianbo Yuan, Quanzeng You, and Jiebo Luo. Twitter sentiment analysis via bi-sense emoji embedding and attention-based lstm. In **Proceedings of the 26th ACM international conference on Multimedia**, pp. 117–125, 2018.
- [4] 中村明. 感情表現辞典. 東京堂出版, 1993.
- [5] Weicheng Ma, Ruibo Liu, Lili Wang, and Soroush Vosoughi. Multi-resolution annotations for emoji prediction. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 6684–6694, 2020.
- [6] Taku Kudo. Mecab: Yet another part-of-speech and morphological analyzer. <http://mecab.sourceforge.jp>, 2006.
- [7] 佐藤敏紀, 橋本泰一, 奥村学. 単語分かち書き辞書 mecab-ipadic-neologd の実装と情報検索における効果的な使用方法の検討. 言語処理学会第 23 回年次大会発表論文集, pp. 875–878, 2017.